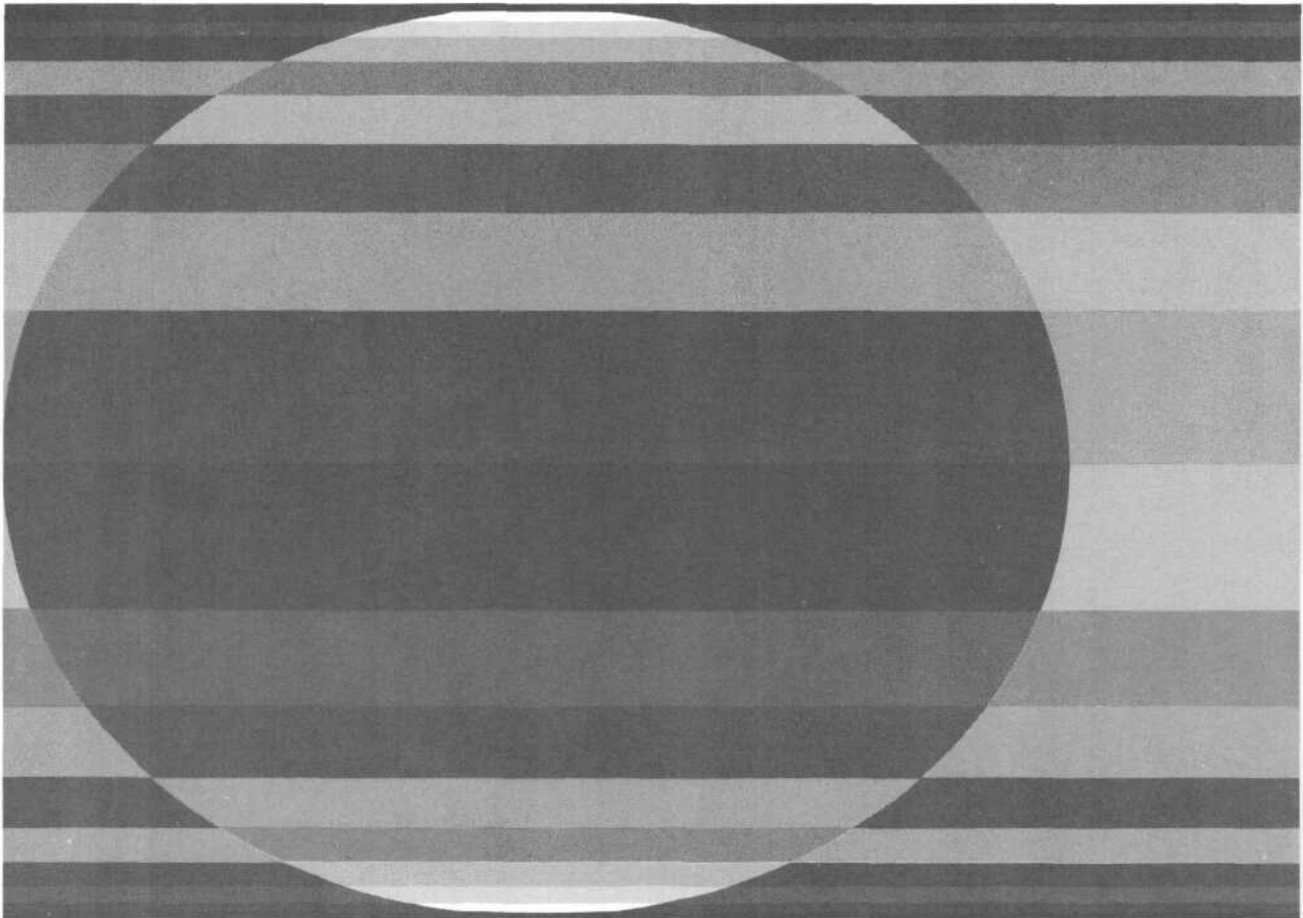


BACKGROUND PAPER

The Effect of PSROs on Health Care Costs: Current Findings and Future Evaluations

June 1979



Congress of the United States
Congressional Budget Office

THE EFFECT OF PSROs ON HEALTH CARE COSTS:
CURRENT FINDINGS AND FUTURE EVALUATIONS

The Congress of the United States
Congressional Budget Office

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

PREFACE

At the request of the Subcommittee on Oversight of the House Committee on Ways and Means, the Congressional Budget Office prepared this background paper analyzing the **effectiveness** of the Professional Standards Review Organizations in curbing the growth of expenditures for health care. The paper also identifies the gaps in what is known about the **effectiveness** of the PSRO program and explains why a more complete and reliable evaluation will depend on the way in which the program is implemented in the future.

The paper was prepared by Paul B. Ginsburg and Daniel M. Koretz of CBO's Human Resources and Community Development Division under the supervision of Robert D. Reischauer and David S. Mundel. The authors are grateful for the help of Allen Dobson, Peter McMenamin, and Paul Eggers of the Health Care Financing Administration. Mitchell Dayton of the University of Maryland helped in many ways, including providing the computer work for CBO's **reanalysis** of the data. Paul Gertman of Boston University contributed valuable suggestions and comments, as did Larry Wilson, Cheryl Smith, and C. Richard Neu of CBO. The manuscript was edited by Johanna Zacharias and Robert L. Faherty, assisted by Brice S. McDaniel. Toni Wright typed the many drafts and **prepared** the final manuscript for **publication**.

In keeping with CBO's mandate to provide objective analysis, this study offers no **recommendations**.

Alice M. Rivlin
Director

June 1979

CONTENTS

	<u>Page</u>
PREFACE	iii
SUMMARY	ix
CHAPTER I. INTRODUCTION--PSROs AND THE REGULATION OF MEDICAL CARE UTILIZATION.	1
Plan of the Paper.	2
The Need for Regulation	3
Professional Standards Review Organizations.	7
CHAPTER II. THE EFFECTIVENESS OF PSROs: FINDINGS AND METHODOLOGICAL LIMITATIONS.	12
Determining PSROs' Effectiveness.	14
Are the Savings the Program Yields Greater Than Its Costs?.	32
The Magnitude of Net Savings Relative to the Size of the Problem	38
Why Aren't PSROs More Effective?.	39
CHAPTER III. PROGRAM AND EVALUATION OPTIONS.	45
Program Options.	45
Options for Evaluation	48
APPENDIX A. ANALYSIS OF THE COLORADO ADMISSIONS PROGRAM AND COMMONWEALTH HEALTH AGENCIES MONITORING PROGRAM	58
CAP.	58
CHAMP.	64
APPENDIX B. A REANALYSIS OF THE HCFA MEDICARE RATE DATA	69
The Regression Model.	69

Estimating Percent Reduction in Utilization71
Ranking PSROs in Effectiveness74
Variations Between Types of PSROs in Effectiveness76

TABLES

TABLE A-1. COLORADO UTILIZATION DATA COMPARED WITH
KANSAS/NEBRASKA DATA: DAYS OF CARE PER
1,000 **MEDICARE** ENROLLEES 61

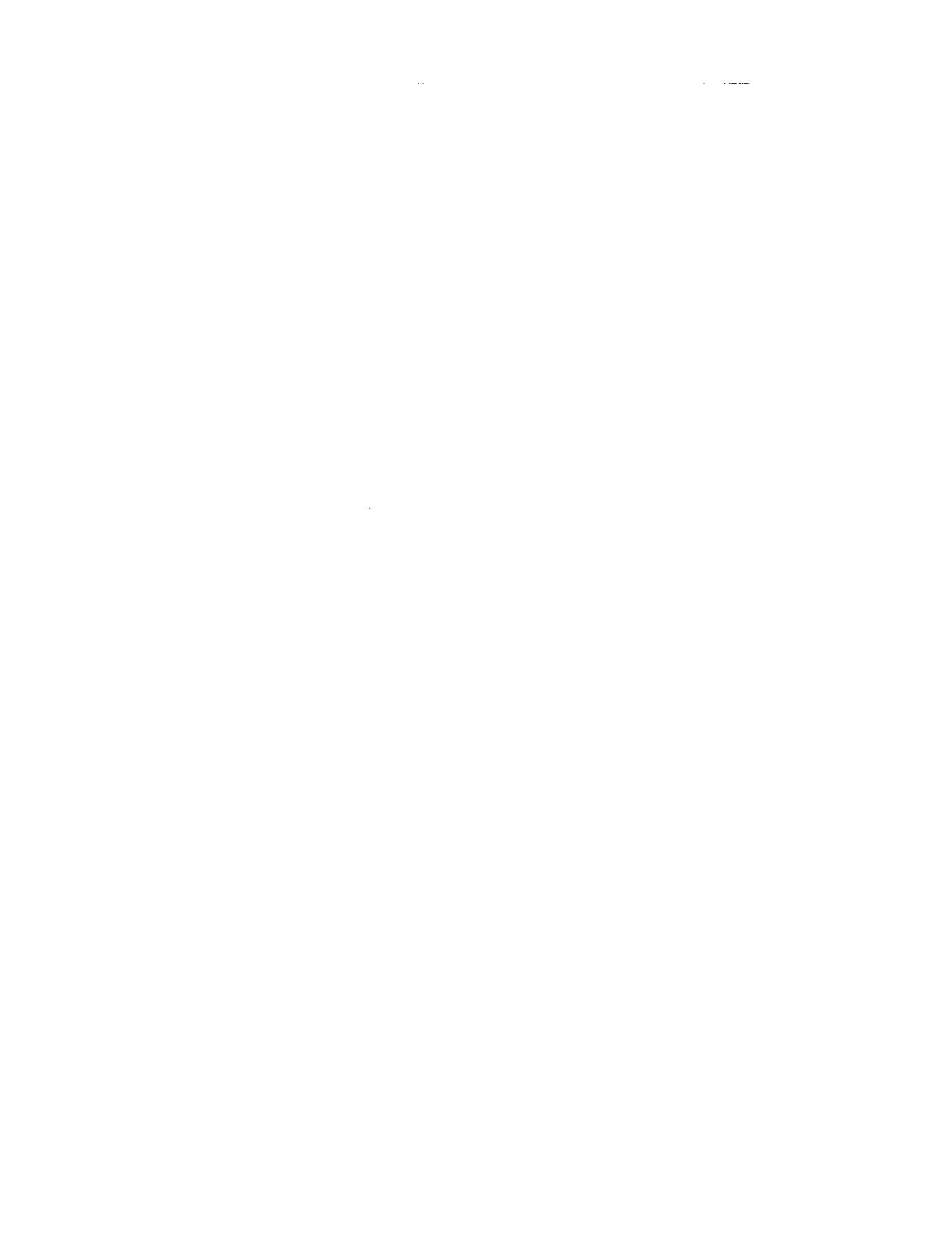
TABLE A-2. MASSACHUSETTS OBSERVED AND PROJECTED
AVERAGE LENGTH OF HOSPITAL STAY, AS A
PERCENT OF PREDICTED LENGTH OF **STAY**,
MEDICAID AND **NON-MEDICAID** PATIENTS,
DECEMBER **1972-DECEMBER** 1976 66

FIGURES

FIGURE 1. HOSPITAL UTILIZATION DATA FOR **KANSAS/**
NEBRASKA AND COLORADO IN DAYS OF CARE
PER 1,000 **MEDICARE** ENROLLEES; 1969-1975 ... 62

FIGURE 2. MASSACHUSETTS OBSERVED AVERAGE LENGTH OF
HOSPITAL STAY AS A PERCENT OF PREDICTED
LENGTH OF STAY: ADJUSTED FOR AGE AND
DIAGNOSIS; DECEMBER 1972-DECEMBER 1976 67

FIGURE 3. MASSACHUSETTS OBSERVED AND PROJECTED AVERAGE
LENGTH OF HOSPITAL STAY AS A PERCENT OF
PREDICTED LENGTH OF **STAY**: ADJUSTED FOR
AGE AND **DIAGNOSIS**; DECEMBER 1972-
DECEMBER 1976. 67



SUMMARY

The Social Security Amendments of 1972 established the Professional Standards Review Organization (PSRO) program in order to "promote the effective, efficient, and economical delivery of health care services of proper quality for which payment **may** be made under the Act." The PSRO program attempts to meet this goal by means of a peer review system that is funded by the U.S. Department of Health, Education, and Welfare (HEW). While the goals of the program are broad enough to include both reduction of expenditures and assurance of quality, the **primary** emphasis of the program has been to reduce utilization **of--and** thereby expenditures **for--short-stay** hospital care by means of "concurrent **review.**" Typically, PSRO concurrent review consists of examining hospital admissions to certify that, from a medical standpoint, they are appropriate and reassessing each case periodically to determine whether continued inpatient care is warranted.

Review and reanalysis of the research on the effectiveness of PSROs indicate that concurrent review is reducing the number of days of hospital care of Medicare enrollees by about 2 percent. This estimate has to be viewed with caution, however. Most extant evaluation studies are too flawed to be reliable, and furthermore, they yield inconsistent evidence. Even the best research **available--a** generally sound study conducted by **HEW's** Health Care Financing Administration (**HCF**A), on which the 2 percent estimate is **based--also** suffers from some important weaknesses.

Because of the lack of relevant data, it cannot be assumed that PSROs are equally effective in reducing utilization by other federal **beneficiaries** (primarily Medicaid patients) whose care is subject to PSRO review. Similarly, it is not clear what effects PSRO review would have on other groups (for example, veterans and private patients) if the **program's** authority were extended to them.

Although PSROs seem to be effective in reducing Medicare utilization, it is doubtful that they produce a net savings.

The recent HCFA analysis concluded that the monetary benefits of the Medicare portion of the PSRO program have been about 10 percent greater than its costs. That analysis implies an extremely small net savings relative to expenditures for services that are currently being reviewed by PSROs (less than 0.1 percent of relevant Medicare **reimbursements**). A CBO **reanalysis** of the data revealed no net savings at all; CBO has concluded that the best estimate is that the savings generated by the program are about 30 percent less than program costs. Both the CBO and HCFA estimates, however, rest on controversial assumptions and are open to considerable error.

A number of factors, including budgetary constraints, current concern with the containment of health-care costs, and continuing changes in the PSRO program, suggest that further evaluation of the effectiveness and cost-effectiveness of PSROs is needed. Moreover, the inconclusiveness of much of the existing research on PSROs indicates the importance of improving the quality of evaluations of the program. To some degree, quality can be increased by improving the research methods employed. However, the reliability of even methodologically sound **evaluations--for** example, the recent HCFA evaluation, which is for the most part a careful and well-designed **study--have** been limited by the way the program itself has been implemented.

Unless changes are made soon in both implementation and evaluation, future evaluations of the program will continue to be **unreliable--often** to such a degree as to be useless in formulating policy. This problem extends both to new PSRO activities (for example, review of long-term care) and to refinements of existing activities (such as focusing review on certain diagnoses, providers, practitioners, or patient groups that offer the greatest potential for a PSRO effect).

The most important improvement in the evaluation of PSROs would be a more careful use of comparison groups. When the effects of a certain component of the PSRO program are to be evaluated, that component must be implemented only in some areas (the "treatment" group), while other selected areas (the "comparison" group) are left without it. If the treatment and comparison areas are initially similar in all other respects, comparing them after the program is underway reveals whether seeming "effects" of the program are actually caused by other factors. For example, recent years have shown a general trend

toward a shorter average length of stay for hospitalized patients; use of comparison groups would avoid mistaking this trend, which began before the existence of PSROs, for an "effect" of the **PSRO** program. On the other hand, comparisons between areas with and without PSROs can be seriously misleading if the treatment and comparison areas were not equivalent (or nearly so) before the program. For example, if the program were implemented in areas already experiencing a decline in average length of stay, and the comparison areas were those in which average length of stay was stable, the comparison would show a spurious "effect" of PSROs on length of stay.

The way in which the PSRO program has been implemented has hindered reliable evaluation by preventing the creation of an appropriate comparison group. Ideally, the treatment and comparison areas should be chosen randomly; as a second-best alternative, they could be selected to be alike in as many respects as possible. To date, however, the implementation of the PSRO program has relied on **"self-selection"**: that is, areas have chosen on their own initiative whether or not to **participate**. Those that chose to participate became the treatment group, while those that chose not to participate became the comparison group. Self-selection virtually guarantees that the treatment and comparison groups will be dissimilar in many **respects--often** in ways that will cloud evaluation of the program.

Depending on what specific component of the program is involved, changing the manner of implementation to permit the use of good comparison areas might require legislative as well as HEW initiative. For example, several PSROs are currently pilot testing a new method of concurrent review that makes use of information on severity of illness and intensity of medical services as well as broad diagnostic categories. In contrast, the more traditional form of concurrent review is built around regional, **diagnosis-specific** norms for length of stay. The new method has received considerable attention as potentially cheaper and more effective than the traditional method. To test the new method reliably, one would randomly assign some PSROs to use it, while other areas would be left to use the old methods. Since the current statute gives individual PSROs the authority to choose their own criteria for review, however, HCFA would be unable to assign PSROs to the new system without legislative initiative.

Other improvements in the evaluation of the program could be made entirely on agency initiative. Multi-site evaluations should be stressed, and less emphasis should be placed on evaluations of individual PSROs. The measures of utilization employed should be comprehensive and should relate clearly to health-care costs. When feasible, utilization of health-care resources should be measured repeatedly over a considerable time span before the program is implemented; this allows one to assess pre-existing trends and clarify initial differences between the treatment and comparison areas, in order to avoid mistaking irrelevant patterns for PSRO effects. A few of the best evaluations of PSROs have incorporated some of these improvements, but further improvement is still greatly **needed**.

Reliable assessments of the effects of a given PSRO program component are often feasible only at early stages of that component's **implementation**. As implementation continues and the number of areas with that component **increases**, it becomes increasingly **difficult--and** eventually **impossible--to** create a reasonable comparison group. For that reason, if current or pending changes in the PSRO program are to provide reliable evaluations that are useful in formulating future policy, improvements of the sort discussed here must be made in the near future.

CHAPTER I. **INTRODUCTION--PSROs** AND THE REGULATION
OF MEDICAL CARE UTILIZATION

Since the enactment of Medicaid and Medicare in the mid-1960s, federal expenditures for health care have grown very rapidly. Congressional concern over the level of expenditures has been expressed throughout the period, and recently it has increased substantially. 1/

In response to this concern, several programs have been instituted to control expenditures for medical care. One of these efforts is the Professional Standards Review Organization (PSRO) program, established by the Social Security Amendments of 1972. 2/ The PSRO program is a type of peer review intended to "promote the effective, efficient, and economical delivery of health care services of proper quality for which payment may be made under the [Social Security] Act." "Proper quality" services are defined as those that meet the following three criteria:

- o They conform to appropriate professional standards;
- o They are provided only when deemed medically necessary;
- o They are **provided** in the most economical but nonetheless appropriate **setting--for** example, on an ambulatory rather than an inpatient basis, if appropriate.

Although the PSRO program has a broad range of goals, in practice it has primarily emphasized curbing certain types of inappropriate use of health-care resources. Considerable doubt has arisen about whether the program is meeting this goal.

1/ See, for example, Medicare and Medicaid: Problems, Issues, and Alternatives, prepared by the staff of the Senate Committee on Finance, 91st Cong. 1 sess. (1969).

2/ Public Law 92-603.

At the same time, the costliness of the program has raised concern that whatever benefits it may yield might not be sufficient to justify the costs. Budgetary constraints are currently holding the costs of the PSRO program to roughly \$150 million a year; this is approximately half of the amount projected by the U.S. Department of Health, Education, and Welfare for full implementation of current PSRO activities. 3/ However, extending PSRO review to ambulatory care of Medicaid and Medicare beneficiaries might increase current expenditures two- or three-fold. 4/ Extending PSRO review to long-term care and other services covered by the law would also be expensive.

The PSRO program is at an important juncture now because heightened interest in containing health-care costs has focused attention on the potential of PSROs to curb utilization of health-care resources. Depending on how effective PSROs are, the Congress may decide to speed up implementation of the program, hold back further implementation, reduce current activities, or restructure the program.

PLAN OF THE PAPER

The remainder of this chapter is devoted to background information on PSROs. It outlines why regulating medical-care practice may be desirable, and it describes those regulatory policies that preceded PSROs and those that continue to the present. The chapter also sketches the organization of the PSRO program.

Chapter II contains an analysis of three aspects of **PSROs' effectiveness**. First, it reviews a number of evaluation studies to assess the **effectiveness** of PSROs in reducing the use of hospital services by Medicare and Medicaid beneficiaries. Second, it presents calculations of the net savings yielded by the **program--that** is, the savings the program yields are compared

3/ See Professional Standards Review **Organizations**, Hearings before the Subcommittee on Oversight of the House Committee on Ways and Means, 95:2 (1978), p. 65.

4/ IOM, Assessing Quality in Health Care: An Evaluation (Washington: National Academy of Sciences, November 1976).

to its costs. Third, the magnitude of the **program's** net savings is compared to the size of the problem that the program is intended to help **solve--namely**, the extent of federal expenditures for **inpatient** care in short-term hospitals. The chapter also explains why so little is known and discusses factors that might be limiting the **effectiveness** of PSROs.

It is important to note that PSROs may affect the quality of medical care as well as utilization and costs. Both are important. It might be decided, for example, that a PSRO-type program that was not very effective in curtailing costs might nonetheless be worthwhile if it improved the quality of health care. Because research on quality effects is lacking, however, and because the program currently emphasizes health-care utilization and costs, this analysis focuses on the latter issues.

Chapter III discusses options for the program and for its evaluation and describes alternative patterns of future program implementation that would yield more reliable assessments of the program's **effectiveness**.

Appendix A reviews case studies of two PSROs. Appendix B discusses a number of technical points raised in the review of the PSRO rate study recently released by the Health Care Financing Administration (HCFA). 5/

THE NEED FOR REGULATION

The U.S. medical-care system is currently subject to various types of regulation. These include controls on prices (prospective reimbursement of hospitals, fee schedules for reimbursement of physicians), constraints on the construction of new facilities and the introduction of new services (health planning activities), standards of competence for the practitioners and providers of health-care services (licensing, **accreditation**), and limitations on the ways medical care is given. The PSRO program is one system for providing the last of these types of

5/ HCFA, Professional Standards Review **Organizations:** 1978 Program Evaluation (January 1979).

regulation. The program is designed to regulate the provision of medical care to most **beneficiaries** of federal programs that finance health services.

The regulation of medical-care practice is designed to alter the array of medical services delivered to patients. Given a standard of desirable care, an existing practice may be deemed inappropriate for one or more of the following five reasons:

1. Additional services could **significantly** improve the patient's prognosis;
2. A different course of treatment could improve the **prognosis**;
3. Some **services** are deemed "unnecessary" because they offer little if any improvement in prognosis;
4. Some services actually risk harming the patient while offering little medical benefit; and
5. Services delivered in a lower-cost setting (such as in a nursing facility, or at home) can be as effective as in a hospital.

Regulation holds out the prospect of cost containment if conditions 3, 4, or 5 exist, and sometimes if conditions 1 or 2 exist. It holds out the prospect of raising quality if conditions 1, 2 or 4 exist.

Inappropriate medical care may exist in an unregulated system for a number of reasons. Because patients usually lack the expertise to discern whether care is unnecessary and/or of poor quality, they depend on physicians to act as advisors in the purchase of medical services. Furthermore, convention among physicians discourages doctors from assisting patients in judging other **doctors'** work. Thus, physicians are responsible for the appropriateness of their services. A number of factors, however, impede their carrying out this responsibility.

Medical information diffuses slowly and unevenly. As a result, some techniques are used too long and others are not used soon enough. Physicians may be too busy to keep up with new developments. Furthermore, much of the information that is most

readily available to them is oriented toward promoting certain types of new **techniques--for** example, use of new drugs.

Financial incentives encourage the delivery of unnecessary services. Under the **fee-for-service** mode of payment, the physician usually gains **financially** from providing more services. In addition, patients' health insurance lessens their reluctance to use more services **because** of considerations of cost, and similarly, it lessens **physicians'** incentives to choose the most economical **setting** for treatment.

Unnecessary services may also be induced by physicians' fears of malpractice claims. With patients well insured and technically ignorant, physicians are free to practice "defensive medicine," which **involves--among** other **things--more** diagnostic testing than is called for by best medical judgment.

A common response to problems of inappropriate care is to review the course of treatment prescribed by physicians. This method of regulating medical practice is usually called "utilization review" because it monitors **patients'** use of medical care. Utilization review activities vary widely in terms of the following **characteristics:**

- o Who does the reviewing?
- o At what stage of treatment is the review conducted?
- o What decisions about health-care use does the review focus on?
- o What is the extent and direction of **"targeting"--that** is, to what degree is review focused on specific diagnoses, providers of care, or treatments?
- o If inappropriate care is found, what sanctions are applied?

The major choice involved in who does the reviewing is between peer review and review by a third-party payer (usually an insurance company). Under peer review, a group of local physicians is ultimately responsible for review decisions. When

review is conducted by a third-party payer, it is that organization, whether governmental or private, that makes the ultimate decisions. The decision of whether or not peer review is employed should not be confused with whether or not physicians actually perform the review. Most peer review organizations use nonphysicians for screening in the early stages of review, and third-party payers may employ physicians in the review process. The difference between peer and third-party review is who sets the policies and the objectives being pursued.

Review activities vary in terms of the stage of treatment at which the review is conducted. In the case of hospital use, the review can be conducted on a prospective basis (before the **patient's** admission) for nonemergency cases, on a concurrent basis (during the hospital stay), or retrospectively (after discharge).

Review can also focus on many different decisions. The general course of treatment may be **questioned--for** example, is surgery necessary? Alternatively, the course of treatment may not be reviewed but the appropriateness of the setting questioned. Should this patient be hospitalized or should he be treated as an outpatient? Is the length of an **inpatient's** stay in the hospital too long?

Another aspect in terms of which review systems vary is the extent to which review is targeted. Review could be focused on certain physicians or hospitals, or on certain **diagnoses--for** example, acute myocardial infarction (heart attack). Similarly, certain procedures, such as tonsilectomies and **hysterectomies**, could be examined. Cost **effectiveness** may be increased by focusing on a small number of utilization decisions, rather than by reviewing all of them.

The final dimension is the nature of sanctions. Denial of reimbursement to a physician or hospital is the most common sanction available. Some reviewers use sanctions only rarely, preferring to induce compliance through education.

The federal government has been involved in health-care utilization review for some time. Since the inception of the Medicare program in 1965, utilization review by hospitals has been a condition of participation. Participation in Medicaid was made contingent upon utilization review in 1967. Medicare and

Medicaid regulations permitted wide latitude in the manner of review, creating difficulties in specifying the nature and extent of review activity in the typical hospital. There is evidence, however, that some hospitals conducted review programs similar to PSRO review.

A newly emerging type of utilization review is second opinions for surgery. Unlike formal review, the test of the appropriateness of a **physician's** surgical recommendation is whether it agrees with the opinion of a second physician. When the second physician disagrees, the patient then has to make the decision as to whether to proceed with the surgery.

PROFESSIONAL STANDARDS REVIEW ORGANIZATIONS

As stated earlier, the PSRO program is intended to lower health-care costs and assure the quality of care for beneficiaries of health programs under the Social Security Act by means of utilization review. PSRO review is distinguished from other **utilization** review by the administrative structure in which it operates, by the sanctions it can (or will be able to) bring to bear, and in many cases, by the nature of the review process itself.

Ultimately, PSROs are intended to review the full range of health-care services delivered under the Social Security Act. To date, however, PSROs have been concerned primarily with assessing the appropriateness of admissions to and lengths of stay in short-term general hospitals. The extension of PSRO review to other aspects of health **care--specifically** ambulatory care, long-term care, and ancillary services (that is, laboratory tests, X rays, and so **forth)--has** been very limited and is at present progressing little, largely because of budgetary constraints.

PSROs are **local--or,** in some sparsely populated areas, **statewide--organizations,** but the PSRO system involves state and national entities as well. As required by the statute, the Secretary of HEW divided the nation into 203 "PSRO areas." In each area, physician organizations could apply to HEW for

designation as that area's PSRO. 6/ All physicians in the area are **free** to join the local PSRO **after** it has been selected, and the majority of physicians in areas with PSROs are members. After an initial planning period, the PSRO is responsible for reviewing the appropriateness of health care provided under the Social Security Act in its area; the PSRO may devise its own criteria to use in that review. 7/ PSROs are advised by State **Professional Standards Review Councils** (in states with three or more PSROs) and Advisory Groups composed of nonphysician health-care practitioners and representatives of health facilities. In addition, the Secretary of HEW is advised by a National Professional Standards Review Council (NPSRC) consisting of physicians of recognized standing in the appraisal of medical practice. The NPSRC also provides technical assistance and information to PSROs and develops regional standards to be used by the PSROs.

All PSRO activities are federally financed, but in large part they are locally planned and administered. Review activities are financed by the Social Security Trust Fund, while management and **nonreview** costs are financed by direct appropriation.

Within **guidelines** established by the law, PSROs have some flexibility in determining how to review short-term hospital inpatient services. All PSROs, however, have adopted a plan suggested by HEW. This plan calls for three principal types of review activity:

6/ Although nonphysician organizations may also apply for PSRO status, the law prohibits the Secretary of HEW from designating such a group as a PSRO unless no qualified physician organization in the area has applied. No nonphysician organization has ever applied.

7/ In practice, most PSRO standards are based not on purely local criteria but on the American Medical Association criteria set and the Professional Activity Study regional **length-of-stay** norms. See Health Services **Administration**, Office of Planning, Evaluation, and **Legislaton** (OPEL), PSRO; An Initial Evaluation of the Professional Standards Review Organization (February 1978), Vol. I, p. 4.

- o Concurrent **review,**
- o Medical-care evaluations, and
- o Profile analysis. 8/

These activities are described in the remaining portion of this chapter.

Concurrent Review

The activity that has been most fully implemented, and the one that is the primary focus of PSRO activities at present, is concurrent review. Concurrent review has two components: review at admission and periodic rereviews (continued-stay reviews). Admission review, which generally takes place within 24 hours of a **patient's** admission, entails certifying that the admission is justified and setting a target date for the first continued-stay review. 9/ **Continued-stay** reviews are to determine the necessity of continued inpatient care. At both stages, the major focus of concurrent review is on whether the hospital is the appropriate setting for care. Assurance of quality is not an explicit aim of concurrent review, but quality may be affected by changes in utilization recommended by the PSRO reviewers.

PSROs carry out concurrent review in a variety of ways. Generally, initial screening is conducted by nonphysician "review **coordinators.**" In many instances these are nurses, but they may also be social workers or other types of personnel. Since only physicians are empowered to reject an admission or a continuation of stay, questionable cases are referred to a physician advisor. **Denials--that** is, determinations that admission or continued stay is **inappropriate--are** communicated to patients and their attending physicians. Patients, providers, and practitioners have the right to appeal at the local, state, and **national** levels.

8/ OPEL, **PSRO**, Vol. I, p. 49ff.

9/ In a few exceptional cases, pre-admission review is substituted for the normal post-admission review.

The direct effect of a PSRO denial is that, after a short grace period passes, reimbursement by Medicaid or Medicare for continued hospital care is prohibited. 10/ Additional legal provisions that have yet to be implemented will give HEW two stronger sanctions against providers or practitioners found by a PSRO to order or furnish unnecessary or inappropriate care frequently. First, the practitioner can be excluded from participation in the Medicare and Medicaid programs. Second, HEW can fine an offender as much as \$5,000 to recoup reimbursement for inappropriate care.

The persons actually carrying out concurrent review may be either hospital employees or members of the **PSRO's** own staff. The law requires that a PSRO delegate responsibility for review to hospitals capable of performing it. In June 1977, 76 percent of all hospitals under review were performing review themselves under contract from local PSROs. 11/

Medical Care Evaluations

The second type of activity conducted by PSROs is medical-care evaluations, which are retrospective studies of medical-care practices in a particular area. They are designed to uncover poor quality and ineffective administration. Results of medical-care evaluation studies may be used to make administrative changes to correct **deficiencies**, set standards for concurrent review, and focus concurrent review activities (discussed in Chapter II).

10/ At present, the statute (P.L. 95-142) mandates a single **day's** grace for Medicare patients and gives the PSRO the option of allowing up to two additional days. Medicaid patients, on the other hand, are not allowed any grace days in some states.

11/ HCFA, PSRO; 1978 **Evaluation**, p. 4. The requirement of delegation has been criticized on the grounds that hospital employees do not have the independence necessary to review utilization thoroughly.

Profile Analysis

The least developed activity is profile analysis. In this activity, statistical analyses of large numbers of **PSRO-reviewed** episodes are used to discern patterns of care. The object is to identify areas of health care in which utilization practices may be inappropriate in order to focus concurrent review activities and to suggest topics for medical-care evaluation studies.

* * * * *

Education of physicians, voluntary compliance, and deterrence are expected to be the major sources of changes in medical practice brought about by PSROs. The educational function is promoted by the process of developing and communicating local standards for efficient delivery of health care. The process may also speed dissemination of current medical information. A formal oversight system could serve as a deterrent by making problem providers easier to detect and hence harder to ignore. Providers on the margin of being cited may initiate self-improvements to avoid both penalties and notoriety.

Although the PSRO mechanism is **clearcut** in theory, there are reasons for skepticism about how it works in practice. Given the lack of reliable data on denial rates and **HEW's** delay in issuing regulations permitting the imposition of other sanctions, the extent to which the threat of denials and sanctions acts as a deterrent is highly uncertain. In addition, many PSROs have adopted standards suggested by the National Professional Standards Review Council rather than develop local ones. This presumably reduced the extent of the educational process.

CHAPTER II. THE EFFECTIVENESS OF PSROs: FINDINGS AND
METHODOLOGICAL LIMITATIONS

Three general questions should be asked about the PSRO program:

- o How effective is the program in reducing hospital utilization?
- o Are the savings associated with the program large enough to justify the costs of the program itself?
- o Are the **program's** net savings large enough to warrant the expectation that PSROs will play a major role in containing health-care costs?

With regard to the first question, review of the best existing research leads to the conclusion that PSROs have brought about a modest decrease in the use of short-stay hospitals by Medicare **beneficiaries**. Nationwide, the decrease in days of care is roughly 2 percent. 1/ Caution should be exercised in interpreting this conclusion, however, for several reasons. First,

1/ Difficulties in obtaining baseline data for **Medicaid** recipients have prevented a rigorous assessment of PSRO effects on Medicaid utilization. However, it would not be surprising if PSROs were less effective with Medicaid than with Medicare utilization because of differences between the people served by the two programs. The Medicare population, of course, primarily comprises the elderly; the nonelderly Medicare population is composed of persons who have received Social Security Disability Insurance for at least two years. In contrast, the Medicaid population consists primarily of children and women of child-bearing age. These two populations have very different patterns of illness, and their use of medical services differs accordingly. Evidence indicates that hospital admissions of Medicaid patients (continued)

the results of various studies have not been consistent with each other. The conclusion stated here does not derive from a consensus of research results; rather, it reflects a judgment that certain results deserve much greater weight than others because of relatively better data and methodology. Second, the rapid implementation of the program has created major obstacles to reliable measurement of program impact. The best evaluations of the program had to compare areas that had active PSROs at the time of evaluation with those in which PSROs had not yet become active. However, areas were not chosen randomly by HEW to **have--or** not to **have--the** early PSROs. On the contrary, the speed with which PSRO activity began in each area was a function of the interest of existing organizations in that area to become PSROs and their capability to take on that role. In other words, areas selected themselves to have early or late starts of PSRO activity. This fact tends to undermine the reliability of any observed difference in health-care utilization in areas with and without active PSROs. Also, there is little information on pre-PSRO activities under the utilization review program, further clouding the accuracy of measured utilization changes.

The second **question--concerning cost-effectiveness--is** even more difficult to answer. The HCFA Medicare rate study used its finding of a 1.5 percent utilization reduction to estimate that utilization savings exceeded review costs by 10 percent. 2/ Re-analysis of these data by CBO suggests that a better estimate would be a utilization savings about 30 percent less than review costs. Both estimates, however, are subject to wide margins of error.

1/ (continued) are primarily for routine deliveries and childhood **infections--conditions** in which one would expect little overutilization and little disagreement about the appropriateness of admissions and lengths of stay. Consistent with this expectation is a recent study of utilization review in 44 Massachusetts hospitals, which found that in almost all instances in which continued-stay review resulted in a denial of permission for continuation, the patients were over 65; the mean age of all patients denied was 79. See Paul M. Gertman and Michael E. Egdahl, "The Dynamics of Utilization Review: A Case Study of 44 Massachusetts Hospitals," Annals of Surgery, vol. 188 (October 1978), pp. 544-551.

2/ HCFA, PSRO; 1978 Evaluation.

Even if one accepts **HCFA's** higher estimate of net savings, the answer to the third **question--whether the program's** net savings will affect health-care costs **significantly--is** pessimistic. Any net savings achieved by the program are extremely small relative to federal expenditures for acute **inpatient** care. The HCFA estimate of net savings amounts to less than 0.1 percent of relevant Medicare **reimbursements**. At the time of the most recent measurement of PSRO **effectiveness--1977--the** program could not be expected to play a major role in **containing** expenditures for hospital care.

These conclusions are based entirely on an analysis of utilization in short-stay general hospitals by Medicare patients. Because of a lack of adequate data and research, these conclusions cannot be generalized to other patient groups or to other types of utilization (for example, ambulatory care).

This chapter reviews the evidence pertaining to these three questions about the effects of PSROs on utilization and costs, and discusses factors that might underlie the research findings. **Primary** emphasis is given to the one study that is the most reliable and that provided the major basis for the conclusions stated **above--namely**, the 1978 HCFA Medicare rate study. Other, generally less reliable, studies are discussed briefly to put the best study into perspective and to illustrate the methodological problems that make it difficult to answer the three questions **with** confidence.

DETERMINING PSROs' EFFECTIVENESS

PSROs and previous utilization review systems have generated many evaluation studies of varying scope and quality. Most have been evaluations of single review **organizations**, but a few have been national in scope.

Evaluations of Single Review Organizations

The body of research evaluating individual PSROs, PSRO prototypes, and pre-PSRO review systems is extensive. 3/ These

3/ The distinctions between PSROs, PSRO prototypes, and pre-PSRO review systems are often vague. Accordingly, the three are not treated separately here.

studies range from simple case studies to elaborate research designs. A review of this research points to two conclusions. First, the great majority of these studies are methodologically too weak to provide any but the most tentative assessment of the programs. (Indeed, many are too badly flawed to provide even a tentative **evaluation.**) Second, even the best studies, as a group, provide no consistent evidence that the programs are **effective.**

The 1977 evaluation of the PSRO program by the Office of Planning, Evaluation, and Legislation (OPEL) of the Health Services Administration includes a volume summarizing a number of the better evaluations of pre-PSRO systems and PSRO prototypes. 4/ It concluded that:

Generally speaking, the evaluations performed on programs that review utilization for necessity and duration of care have been inconclusive. Only one program [CHAP, in Sacramento, California] has demonstrated a significant reduction in utilization. Evaluations of other programs either found no significant effects, were met with serious challenges to their evaluation methodology, or were of such limited scope as to cast doubt on their value. One program evaluation [HAPP, in New Mexico] indicated a significant increase in the use of services and expenditures per Medicaid eligible. 5/

OPEL also noted that, although some studies reported that review was **cost-effective**, such estimates were frequently erroneous. 6/

A second review of studies of individual review organizations, including a number of PSROs, was the report prepared in 1976 by the Institute of Medicine (IOM) on assessing quality in medical care. The programs reviewed in the IOM report were not selected to be representative but rather to serve as examples of the best programs then in operation. Accordingly, estimates of program effects derived from this study would be expected to overstate the benefits of the typical review program.

4/ OPEL, PSRO, Vol. II.

5/ Ibid. , pp. 130-31.

6/ Ibid. , p. 132.

This sampling bias **notwithstanding**, the IOM report reached pessimistic conclusions. With regard to hospital **inpatient** concurrent review, IOM concluded **that**:

Because documentation is frequently inadequate, a definitive assessment of effectiveness is not possible at this time. **Nevertheless**, the information is sufficient to permit a preliminary assessment of current **programs....In** general, available information does not demonstrate convincingly the cost effectiveness of the concurrent review programs visited. Related literature is similarly **peessimistic**. 7/

IOM's pessimism generally reflected the lack of convincing evidence of reduced utilization, as well as the lack of convincing calculations of costs and savings.

The methodological problems that both OPEL and IOM cited as weakening the conclusions of the studies they reviewed apply, in varying degrees, to other PSRO evaluations as well. Many

7/ IOM, Assessing Quality, pp. 57, 73. One exception to the general bleakness of the IOM's conclusions was in the area of ambulatory care review, which is to be incorporated into PSRO review when the latter is fully implemented. A few of the programs studied by IOM incorporated ambulatory care review. The evaluations of these programs were "surprisingly consistent" (p. 76) in showing ambulatory care review to be both effective and **cost-effective**. IOM noted, however, that most of the savings appeared to accrue from traditional claims review functions (that is, denial or adjustment of claims) rather than from peer review itself, and in many instances did not reflect actual declines in utilization. These findings are only tentative, in part because of the small number of studies involved. Nonetheless, they have two implications for ambulatory care review: first, for this type of care, it may not be most efficient to supplement traditional claims review with peer review, and second, that the programs may be "cost effective for the fiscal intermediary, but not necessarily for society" (Assessing Quality, pp. 76-77).

reviews of individual PSROs that have appeared since those reports have been so seriously flawed that no confidence can be placed in their results. At the other extreme, the national PSRO **evaluations**, while unquestionably the best studies available, suffer somewhat from a number of the same problems, with the result that some doubt about the accuracy of their findings remains.

The Lack of Adequate Comparison Groups. The weakest of the studies of individual PSROs simply compared utilization in one area before and after the implementation of utilization review. This is generally the least reliable method of assessing PSROs, because it confuses the effects of the program with other trends. For example, some of the studies of pre-PSRO review systems analyzed **in** the OPEL report compared average length of hospital stay in one area before and after the start of the program. OPEL noted that, since there was a nationwide trend toward decreasing length of stay during the period under consideration, any seeming "effects" of the program could simply be a reflection of the national trend. 8/ The

8/ OPEL, **PSRO**, Vol. II, p. 132. The national trend of decreasing length of hospital stay is just one of an almost unlimited number of trends that could be mistaken for program effects in this type of study. A short-term trend that can be problematic is the seasonal variation in the use of medical care. Hospital use is heavier during certain times of the year (especially during the winter). Comparison of utilization data from different seasons carries the risk of mistaking this seasonal pattern for a PSRO effect. This problem is easily avoided by comparing full years of data or by comparing data from the same season in different years. Other problems are posed by **people's** mobility, and by the fact that varying numbers of people in different years, while not moving, buy services in a PSRO area other than their own. (These problems were handled well in the national PSRO **evaluations**.) Changes in federal program regulations, economic conditions, and availability of doctors and hospital facilities can also markedly affect the use of medical care. Many other examples of misleading patterns could be cited.

same fundamental problem can be found in many of the more recent studies as well. 9/

In order to separate the effects of the PSRO program from the effects of unrelated changes, the better evaluations of the program have used areas that have no PSROs as a comparison. Detecting the effects of the PSRO program is then the same as any investigation that compares an experimental (or "treatment") group with a comparison (or "untreated") group. In the case of the PSRO program, the two groups are made up of PSRO areas, and the "treatment" is the implementation of PSRO review. If, before the program is implemented, utilization in the comparison group is the same as utilization in the treatment group, differences that appear after the program is implemented can be attributed to the effects of the program.

The key to the adequacy of this approach is the initial equivalence of the comparison and treatment groups. The groups must be equivalent not only with respect to utilization (or other outcomes of interest), but also in terms of other characteristics that influence utilization. If they are not initially equivalent, what appears to be an effect of the program could be an artifact of pre-existing differences.

The only fully adequate way to guarantee this equivalence is to assign "individuals" (people, patient groups, hospitals, PSROs, or whatever) randomly to each group. 10/ In the absence of random assignment, a variety of methods can be used to assign individuals to treatment and comparison groups. One of the least adequate methods is simple self-selection, such as was described earlier in this chapter; this approach is undesirable because areas that choose to have PSROs are likely to differ in important ways from those that do not. A somewhat better approach is to match the individuals in the groups being compared

9/ For example, some of the case studies cited in OPEL, PSRO, Vol. VI, and in PSROs, Hearings.

10/ In this case, the comparison group is called a control group.

on important **characteristics**. 11/ **Alternatively**, a variety of statistical methods can be used to remove some of the resulting distortion of the analysis. Repeated baseline (that is, **pre-implementation**) measures of utilization can also be useful, because they can reveal differences between the treatment and comparison groups in terms of pre-existing trends. None of these procedures, however, singly or together, is fully adequate to compensate for nonrandom assignment. Therefore, the further an evaluation deviates from random assignment, the less confidence can be placed in any seeming PSRO effect (or lack of effect).

Some PSRO evaluations have drawn their treatment and comparison groups from the same area. A group of patients whose cases are subject to utilization review are used as the treatment group, while other patients who are not subject to review serve as the comparison group. For example, a recent evaluation of the Massachusetts Commonwealth Health Agencies Monitoring Program (CHAMP) compared the utilization by **Medicaid** patients in certain hospitals with that of other patients in the same hospitals. 12/ It is virtually impossible, however, to obtain comparable treatment and comparison groups using this strategy. Most patients under review are federal **beneficiaries** (in most instances, Medicaid or Medicare patients) who are in a federal program for the very reason that they differ from the general population in certain **ways--for** example, in age, income, or health status. Therefore, similar individuals who would constitute an appropriate comparison group would generally be eligible themselves for the same federal program. Furthermore, changes in the federal programs **involved--for** example, **changes** in state Medicaid **regulations--during** the course of the study could change utilization **in** ways that might be misinterpreted as **effects** of utilization review.

11/ Matching before random assignment is not subject to the criticism discussed here. On the contrary, it is often superior to simple random assignment.

12/ Anita Fulchiero and others, "Can the PSROs Be Cost Effective? A Study of the Effect of the Commonwealth Health Agencies Monitoring Program on the Length of Stay of Medicaid Patients in **Massachusetts**," New England Journal of **Medicine**, Vol. 299 (September 14, 1978), pp. 574-80.

Many PSRO studies, including the national evaluations, have used an alternative method of constituting a comparison group. Utilization by one type of patient (for example, aged Medicare enrollees) is compared in areas with and without PSRO review. However, the Social Security Amendments of 1972, which established the PSRO program, called for a program implementation process that virtually insured that treatment and comparison groups constructed in this way would not be equivalent. The law mandated that the entire nation be broken into PSRO areas and that conditional PSROs be established "at the earliest practicable date after designation of an area." This mandatory speed and all-inclusiveness precluded assigning areas to serve in treatment or comparison groups on the basis of either randomization or careful matching. The areas first subject to PSRO review were those with physician organizations eager and qualified to perform peer review. These became the treatment group for PSRO evaluations. In those areas without such **organizations**, PSRO activity was delayed, and hence they served as the **comparison group**. Thus, active PSRO **areas--the treatment group--were** self-selected.

With such a selection mechanism, areas with active PSROs could be expected to differ from those without them. Statistical analysis confirmed that while the areas were similar in many respects, they had substantial pre-existing differences. In April 1977, when roughly half of all PSRO **areas** had an active PSRO, the active areas were drawn disproportionately from the Northeast and the West. The active areas also had a greater population density, a greater number of physicians per 1,000 individuals, a higher 1974 (baseline) hospital admissions rate, and a higher percentage of hospital stays of 22 days or longer. 13/

There is also evidence that the two sets of areas differed in potentially important but intangible aspects such as **physicians'** attitudes toward peer review. For example, the comparison group (as of April 1977) included one state in which no **physicians'** organization had ever indicated a willingness to serve as a PSRO. There may also have been differences in **pre-existing** trends in utilization, but this cannot be determined since national data exist only for a single pre-PSRO **year--1974**. Thus, a troubling aspect of the nonrandom assignment of PSROs is that

13/ HCFA, PSRO; 1978 Evaluation, p. 68.

the extent of pre-existing differences between the treatment and comparison areas can never be fully determined.

The absence of adequate comparison groups resulting from the way the program has been implemented limits the confidence one can place in evaluations of PSRO program **effectiveness**. While the severity of the problem varies with many aspects of evaluation design, generally one can place less confidence in **comparisons** between a single treatment area and a single comparison area than in evaluations that use information from many areas. The latter type of evaluation does not solve the problems of inadequate comparison groups, but the large numbers tend to reduce the magnitude of potential bias. This report therefore considers the evaluations conducted by HEW, in which a national sample of active PSROs was compared with a national sample of comparison **areas**, to be the best available test of the effectiveness of PSROs. 14/

Differences in Pre-Program Patterns of Utilization Change.
An additional, often very serious, problem can arise even when the treatment and comparison areas seem to be comparable just before the program is implemented. The potential problem is that different patterns of change may have been underway in the two areas that would not be apparent from a single, pre-program measurement. For example, two areas could show identical average lengths of hospital stay the month before PSRO review was started in the treatment area, even if average length of stay had been rapidly declining (from a higher initial level) in one area and increasing (from a lower level) in the other. In such a case, average length of stay in the two areas would be expected to become less and less similar as time went on, merely because of **pre-existing** trends. This increasing dissimilarity could easily be mistaken for an effect of the program.

Problems of this sort are particularly likely to occur in studies that involve only a few areas. They can, however, occur in large studies as well, if assignment to treatment and comparison groups is nonrandom. Even the national PSRO evaluations, which involved nearly 200 areas, may have been affected by distortion caused by pre-existing patterns. The only way to rule out these problems is to measure utilization (or other variables of interest) repeatedly for a considerable time before the start of the program.

14/ OPEL, PSRO, Vol. I & III; HCFA, PSRO; 1978 Evaluation.

The severity of this **problem** is illustrated by two evaluations of individual PSROs which did collect repeated, pre-program utilization measures that allow one to check for differences in **pre-program** trends. The first is **OPEL's** study of the Colorado Admissions Program (CAP). 15/ The CAP study is methodologically superior to most individual studies and contains extensive baseline data. The second is a study of the Massachusetts organization mentioned earlier, CHAMP. 16/

Colorado. CAP began before implementation of the PSRO program and became one of the first active PSROs as the program evolved. The evaluation of the CAP program used Kansas and Nebraska, which had no PSRO or PSRO prototypes, as a comparison group.

The most important analysis in the Colorado study looked at utilization by Medicare patients both before and after the implementation of CAP in 1973. Utilization rates (days of care per 1,000 Medicare enrollees) for the four years from 1969 through 1972 were averaged to establish a baseline, which was then compared with two post-implementation years, 1974 and 1975 (1973 was disregarded as a transitional year). The pattern of change in the Kansas/Nebraska comparison area was used to adjust the Colorado data for contemporaneous changes that had nothing to do with CAP. This analysis suggests declines associated with review of 7.6 percent from baseline to 1974 and 3.1 percent **from** baseline to 1975.

Examining year-by-year changes during the baseline period, however, rather than treating them as a single average, reveals a very different picture. The major divergence between the Colorado and Kansas/Nebraska utilization rates occurred in **1972--the year before implementation--and** in 1973, the transitional year. The first two post-implementation years show a leveling off and then a lessening of this difference. As a result, the Colorado-Kansas/Nebraska difference was virtually the same size two years after the program started (1975) as it

15/ OPEL, PSRO, Vol. V: A Comprehensive Case Study: The Colorado Experience.

16/ Fulchiero and others, "Can the PSROs Be Cost Effective?"

was both one and four years before the program (1972 and 1969). This pattern casts doubt on the conclusion that the reduction found by the first analysis was truly an effect of CAP. A point made in the abstract earlier is thus illustrated concretely: multiple baseline **measures**, extending for a considerable time before the implementation of the program, can be invaluable in sorting out the effects of the program from other misleading trends. They are particularly important when the treatment and comparison groups have not been randomly assigned and are not equivalent. The lack of multiple baselines accordingly is one of the major weaknesses of the national PSRO evaluations discussed below.

Massachusetts. CHAMP was a PSRO prototype, encompassing the entire state. The CHAMP study compared trends in utilization among Medicaid patients (subject to CHAMP review) with trends among non-Medicaid patients (not subject to review). A greater decline in utilization among Medicaid patients than among non-Medicaid patients (5.3 percent) was found and attributed to the CHAMP program.

Like the CAP study, the CHAMP study included more than one baseline measure (**unfortunately**, however, only two), and again the apparent effect of the program is called into question when one looks at the year-by-year pattern of change. Utilization by Medicaid patients was declining even before the **program's** inception, while that of non-Medicaid patients was increasing slightly. Medicaid utilization after the program began was at a rate that would have been anticipated from an extrapolation of the **pre-program** trend, whereas **non-Medicaid** utilization was at a lower rate than would have been expected. 17/

Problems of Inadequate Data. The preceding discussion has dealt with problems of research design, focusing mainly on comparison groups. Regardless of the quality of design, however, the results of a study are only as reliable as its data are good. **Unfortunately**, many evaluations of PSROs have suffered from severe data problems.

17/ For detailed analyses of the CAP and CHAMP studies, including graphic presentation, see Appendix A.

A major problem of data quality arises in choosing a measure of program impact. 18/ Accurate evaluation of reduced hospital utilization requires examination of its three **components--admissions, length of stay, and intensity of service** (that is, volume of medical service per patient day). In order to avoid mistaking a change in the size of the population under study for a **PSRO effect**, utilization should be expressed as a rate. That is, measures of utilization should take into account the size of the consumer **population--for** example, days of care per 1,000 Medicare enrollees. Inability to measure all of the components of utilization or to express utilization as a rate can detract from the reliability of measurements of PSRO effects.

In practice, however, no evaluations of PSROs have met all of the criteria outlined above. Comprehensive data on service intensity are lacking in all available studies. Furthermore, the use of rate data unfortunately has been quite rare. This is partly due to **the** difficulty of obtaining the necessary information about the number of people in relevant categories. For example, rate data on utilization by Medicaid eligibles have often been precluded by a lack of adequate information on the total pool of eligibles.

Other problems frequently encountered in PSRO evaluations are that data from either the pre-program or the post-implementation period are incomplete, or that the data from the two periods are in some way not comparable. For example, Medicare data for years up to 1974 are not comparable to data for later years because of major changes in the Social Security Administration's data collecting system. Lengthy claims-processing

18/ Ideally, a comprehensive evaluation of the effectiveness of PSROs would include measures of PSRO effects on both utilization and quality of care. Measurement of the quality of care is incomparably more problematic. The major reason is the absence of consensus on the measurement of quality, and the particularly intractable problem of aggregating better outcomes from a wide variety of medical procedures. This has not been a severe problem in practice, however, because PSRO evaluations generally have paralleled the program in placing primary emphasis on utilization rather than quality.

procedures prevent use of the most recent data. 19/ As a result, the most recent HEW evaluations were **restricted** to comparing 1974 with 1977.

Aside from these sometimes intractable data problems, inaccuracies have often occurred because sensible collection practices are not followed. For example, one evaluation used Medicare data that included hospital types not covered by PSROs and hospitals outside the PSRO geographical area. Two other evaluations compared total days of care during the pre-program period with just those days of care certified by the PSRO during the post-program period. 20/

Giving responsibility to individual PSROs to assess their own impact risks perpetuating errors such as those mentioned above, since PSROs tend not to have evaluation specialists on their staffs. Strong technical assistance from a central evaluation team is likely to be necessary to avoid such problems.

National Evaluations of the PSRO Program

The 1977 OPEL report and the 1978 HCFA report included a variety of studies, of which three are national evaluations of the **program's** effects on utilization.

The first of these studies was an evaluation of hospital discharge abstract data. 21/ (Abstract data are summaries of the medical records of a **patient's** stay.) Despite the ambitious scope of the study (over one million Medicare and Medicaid

19/ These lengthy claims-processing procedures can lead to serious distortions if data are used prematurely. For example, the General Accounting Office (GAO) examined the claimed cost savings of six PSROs and estimated the savings to be overstated by 672 percent. A major factor in the overstatement was the use of the latest year's data before late billings could be tabulated. See PSRO, **Hearings**, pp. 2-8.

20/ PSROs, **Hearings**, pp. 2-8.

21/ OPEL, **PSRO**, vol. 4: Acute Care Utilization Impact: Inferences from a Sample of Hospital Abstract Data.

patient discharge abstracts from PSRO and non-PSRO hospitals were included), it does not provide a reliable assessment of the **program's** impact. The study did not use rate data (that is, days of care per capita), but instead relied on **average-length-of-stay** data. In addition, there were serious sampling problems (including nonrandom assignment), and many of the PSROs involved had just begun operating at the time of assessment and may not have had time to become effective. Furthermore, the results of the study were inconsistent: PSROs were associated with a relative decrease in average length of stay of Medicare patients, but a relative increase in average length of stay of Medicaid patients; the combined result (with Medicare and Medicaid) was therefore very small and not statistically **significant**. No convincing explanation of the discrepancy between the two results was offered. Given all the problems in the study, neither the apparent positive PSRO effect on Medicare patients, nor the apparent negative effect on Medicaid patients, can be considered reliable.

The other **two** national **evaluations--one** in the 1977 OPEL report and one in the 1978 HCFA **report--were** methodologically much stronger. Both looked at Medicare rate data (specifically, total days of care per 1,000 Medicare enrollees) and are therefore referred to here as the "Medicare rate studies." Both used a nationwide sample of active and still inactive PSROs, the active ones being the treatment group and the inactive ones serving as the comparison group, and employed sophisticated techniques of data analysis.

The full report of the OPEL 1977 rate study has yet to be completed, and the only published information available on the analysis is a section of the Executive Summary. 22/ The information available, however, seems sufficient to support **OPEL's** conclusion that the analysis yields no reliable evidence of PSRO impact, either positive or negative.

The 1977 report, however, has become far less important in the light of the 1978 rate study, which did find a small but statistically significant reduction in utilization associated with PSRO review. The 1978 study can be seen as a refinement

22/ The Executive Summary is OPEL, PSRO, Vol. I; the full report will be Vol. III.

of the 1977 study; the sample of active PSROs was much larger, and the analytic design was improved in a number of respects. Although some problems **remain--largely** because of limitations of **data--the** HCFA 1978 rate study is unquestionably the best single evaluation of the **effectiveness** of PSRO concurrent review of acute-care hospital use. Furthermore, a better evaluation is unlikely to be forthcoming, because the continued expansion of PSRO review of this type will soon make devising any sort of reasonable comparison group impossible.

The 1978 analysis was based on all PSRO areas in the nation. The 96 areas that had begun concurrent review by April 1977 were classified as active PSROs; the remaining 93 areas were classified as inactive and served as a comparison group. 23/ As noted earlier, the active and inactive samples were not entirely **equivalent--as** one might expect, given the degree of self-selection. The active ones were drawn disproportionately from the Northeast and West, while the inactive ones were concentrated in the North Central and Southern regions. They also differed **significantly** as outlined below:

Characteristics in Baseline Year (1974)	PSRO Areas	
	Active	Inactive
Physicians per 1,000 Residents	1.7	1.2
Population per Square Mile	2,700	515
Hospital Admission Rate per 1,000 Population	312	328
Hospital stays of 22 Days or Longer (percent of total admissions)	13.1	11.6

SOURCE: HCFA, PSRO; 1978 Evaluation, p. 68.

23/ For technical reasons, a few PSROs were dropped from the sample, and a few others were consolidated. Therefore, the total sample (189) was slightly smaller than the national total of 203 PSRO areas.

Furthermore, **physicians'** attitudes in the two areas may differ as well; the inactive sample includes areas in which physicians have persistently (and, in some cases, successfully) resisted peer review. On the other hand, the actives and **inactives** did not differ substantially on a number of other **variables--for example, nursing home beds and short-stay beds per 1,000 population--which** might affect **utilization** or PSRO **effectiveness**.

In order to compensate for the differences between the two groups, HCFA used the statistical technique of multiple regression to analyze the **effectiveness** of the PSRO program. This method contrasted 1977 utilization rates in the active and inactive PSRO areas, after adjusting for **differences** in 1974 baseline (that is, pre-PSRO) utilization and a number of other **pre-existing** differences. 24/ This method is an appropriate choice, given the nature of the data. It should be emphasized, however, that no statistical adjustment procedures can be expected to adjust completely for pre-existing differences. Two of the remaining potential sources of distortion are discussed later in this section.

The most important analysis of these data **examined** the effects of PSROs on total days of care **per 1,000 enrollees**. The PSRO variable was expressed as months under review, rather than as a simple active-inactive dichotomy. The multiple regression analysis indicated that utilization was roughly 2 percent lower in the active (treatment) areas than in the inactive (comparison) areas. 25/ This difference was statistically significant. The

24/ More **specifically**, HCFA used forced-order regression. In this method, baseline utilization and other variables on which there are pre-existing differences are entered into the model in a first stage, with the PSRO variable entering in a second stage.

25/ This 2 percent estimated reduction in utilization is based on a reanalysis of the data by the authors of this report and is greater than the estimate (1.5 percent) given in the HCFA report. The difference stems primarily from two factors. First, the HCFA analysis did not adjust for regional differences. The **authors'** analysis incorporated a four-way regional contrast into the model. This increased the size of the PSRO effect. Second, a (continued)

effectiveness of the program, however, varied according to season. Almost all the reduction in use occurred during the first half of the year, while the program had very little effect during the second half of the year.

These findings that PSROs are associated with a decline in hospital use by Medicare recipients is subject to two qualifications. First, statistical techniques of this sort cannot be expected to adjust completely for differences that existed at the time of the single baseline measure. In particular, the **active** and inactive areas may have differed in ways not measured by the variables in the **model--for** example, there may have been differences in the rate of change of cancer incidence, or in **physicians'** attitudes. The model does account for most of the variation in utilization (93.5 to 95 percent), which would lessen the importance of this problem. However, given the extremely small percentage of the variation attributable to PSRO review (in statistical terms, the 2 percent decrease in days of care corresponds to about 0.2 percent of the variance), the addition of a single variable that is now omitted could be sufficient to remove the PSRO effect.

Second, this model looks only at static pre-existing differences at the time of the single baseline measure. It does not adjust for pre-existing differences in the pattern of change in utilization before PSRO review began. For example, if hospital discharges per 1,000 people were increasing at a faster rate in the inactive areas than in the active ones before 1974, and they continued to do so until 1977, the difference in rate would not be compensated for by the HCFA model and would produce an apparent PSRO effect without any real effect of PSRO review. The CAP and CHAMP evaluations described above are good examples of this sort of bias. Although the probability of such a bias is considerably less in the national evaluations because of the large number of PSROs involved, it remains a real possibility. The lack of additional baseline measures rules out refinements to adjust for this possibility.

25/ (continued) different method was used to estimate the percent reduction in utilization from the regression equation. These differences in technique are described in some detail in Appendix B.

In sum, while the **qualifications** described leave room for some doubt, the 1978 HCFA Medicare rate study provides reasonably firm evidence that PSRO review decreases utilization among Medicare enrollees. The data do not indicate what effect PSROs would have on hospital use by Medicaid eligibles, but other evidence suggests that the effect might be far smaller. Similarly, the data reviewed here provide no basis for estimating what effect PSRO review would have on hospital use by other patient groups, if the PSRO mandate were expanded to include them.

Are Some PSROs More Effective Than Others? Another issue of importance is that of "differential **effectiveness**"--that is, are some PSROs more effective than others? Information on the extent and causes of such differences could be invaluable in program management. It could also be helpful for evaluation efforts; case studies of particularly effective or ineffective PSROs could suggest hypotheses about the most effective **way** to run a PSRO, and those hypotheses could then be tested by more rigorous evaluation studies.

Questions of differential **effectiveness** are of two general types, seemingly similar but analytically very different. One type involves comparisons between individual **PSROs--for** example, a comparison between any two PSROs, or a ranking of all of them. The second type entails comparisons between types of **PSROs--for** example, between rural and urban PSROs, or between those that delegate most review to hospitals and those that do **not**.

Differences in effectiveness between individuals are no doubt substantial. The 1978 Medicare rate study attempted to rank all active PSROs in terms of the reduction in hospital use apparently produced by each, and the report notes the usefulness of information of that sort for program management. 26/ Unfortunately, the rate study in fact reveals very little about differences of this sort. While the analysis produced a reasonably reliable estimate of the **effectiveness** of all PSROs collectively, it yields only highly unreliable estimates of the **effectiveness** of individual PSROs. Estimates of the differences between individual PSROs are even more unreliable; indeed, they are of no

26/ See the Foreword to HCFA, PSRO; 1978 Evaluation, p. i.

real use for evaluation or management. (A conservative numerical estimate of the degree of uncertainty and an explanation of the statistical basis of the problem can be found in Appendix B.)

The rate study can, on the other hand, reliably answer the second type of question about differential **effectiveness**-- that is, comparisons between types of PSROs. 27/ One important question of this latter type is whether "mature" PSROs are more effective than new PSROs. The HCFA report states that the mature PSROs are indeed more effective. However, there are serious problems in the analysis leading to that conclusion (see Appendix B), and after those problems are corrected, there is no remaining evidence that PSROs grow more effective with time (within the range of zero to three years of experience). A recent descriptive analysis of Medicare utilization rates in PSRO areas reached a similar conclusion. 28/

Another question of this second type is whether PSROs are more effective in some areas than in others. The Medicare rate data do reveal striking regional differences in the effectiveness of PSROs. Specifically, PSROs were most effective in the Northeast and North Central regions. 29/ This difference must be interpreted with caution, however. An assumption underlying the CBO analysis of the data is that geographic region is not important in itself, but that it is important as a proxy for other variables that have not been included in the model. That is, if **PSROs** are more effective in one area than in another, it is presumably because the regions themselves differ in many

27/ Technically, with the exception of the question of PSRO "maturity," these are questions of the interactions between PSRO review and other variables. The results described here differ from those in the HCFA report for two reasons: differences between the models used, and a technical error in **HCFA's** procedure (as described in Appendix B).

28/ Paul M. **Gertman** and David L. **Rothberg**, The PSRO Program from 1975-1977: Exploratory Analyses of Activities Based Upon Secondary HEW Data Bases, prepared for the HEW, Health Care Financing Administration (Boston University, Health Care Research Section, 1978).

29/ This conclusion differs from that of the HCFA study; the basis for the differences is explained in Appendix B.

characteristics (for example, in demographic composition, health status, economic status, ethnic mix, **physicians' attitudes**, and so on), and it is one or more of these variables that influence PSRO effectiveness. If PSROs work better in one area than in another, the relevant task is to determine which of the underlying regional differences are responsible.

Several other questions of differential **effectiveness** were also examined in the HCFA study. Taken as a set, these tests for **differential effectiveness** were not statistically significant. 30/ Therefore, the results of these tests can be seen only as suggestive and must be interpreted with caution. They hint, however, that PSROs may vary in **effectiveness** as a function of characteristics of the communities in which they operate, but that those characteristics (for example, the proportion of the population that is 65 or over) are not of the sort that are amenable to policy intervention.

ARE THE SAVINGS THE PROGRAM YIELDS GREATER THAN ITS COSTS?

In order to relate the benefits of the PSRO program to its costs, reductions in utilization must be translated into cost savings. This is difficult to do accurately. One problem is that the cost implications of reduced admission rates are different from the implications of reduced length of stay, because the early days of a hospital stay tend to be more expensive than the later days. Regardless of this difference, the value of a one-day reduction in utilization is less than the average daily cost, because an important portion of hospital costs are fixed. Furthermore, a reduction in hospital use can lead to an increase in use of other types of services (for example, ambulatory care), and such compensatory increases are generally outside the scope of PSRO evaluations. Cost savings also depend on the types of patients involved; for example, if the least ill are denied admission, savings per day will be less than the average daily cost of hospital care. Factors such as these leave room for considerable ambiguity and controversy in calculating the monetary value of the PSRO **program's** benefits.

30/ Furthermore, none of the individual tests reached a level of significance sufficient to be considered reliable in the light of the large number of such tests.

The two Medicare rate studies included savings-to-cost analyses to assess whether utilization reductions caused by PSROs generated enough program savings to justify the costs of PSRO review. Since the 1977 OPEL rate study found no utilization reductions, its savings-to-cost analysis could only calculate how large a utilization reduction would be necessary for benefits to equal savings. The savings-to-cost analysis in the 1978 rate study is of greater importance since there was a utilization reduction on which to base such an assessment. It is the focus of this section.

In the 1978 HCFA rate study, overall savings from Medicare utilization reductions were estimated to be \$50.5 million. This is 10 percent greater than the estimate of \$45.9 million for Medicare review costs, yielding a savings-to-cost ratio of **1.1-to-1**. These estimates required a number of assumptions, such as the proportion of hospital costs that are fixed and the increase in outpatient and long-term care costs associated with reductions in inpatient use. The estimates are highly sensitive to the assumptions made, and many of them lacked adequate factual underpinning. Consequently, the savings-to-cost ratio is subject to a high degree of error.

For the most part, the HCFA evaluators were careful and fair in their choice of assumptions. Although better assumptions could have been used, some would have increased the ratio while others would have decreased it. On balance, however, the ratio of **1.1-to-1** appears somewhat too high. As explained below, a ratio of **0.7-to-1** is more realistic, although the errors in this **reestimation** are also likely to be sizable.

Methodology

In the HCFA study, separate estimates of PSRO costs and savings were made. The cost estimates are more **straightforward** and are discussed first.

The objective of the cost calculation was to isolate the costs of concurrent review from other PSRO costs. The starting point was reports to HCFA by active PSROs. Management functions were allocated to concurrent review according to the proportion of direct costs involved in this activity. Since these costs were for fiscal year 1977, a series of adjustments were made to

make them comparable to the benefits, which were from calendar year 1977. No attempt was made to subtract costs that would have been incurred for pre-PSRO utilization review activities. 31/ For the 96 PSROs identified as active in the study, costs of concurrent review were estimated to be \$45.9 million.

Calculation of the savings was somewhat more complex. The first task was to compute the number of inpatient days saved in each PSRO area. This figure was the difference between the number of Medicare days of care expected in the absence of PSROs and the number expected with PSROs. 32/ On the basis of the Medicare formula for hospital reimbursement and Medicare cost reports, a reimbursement saving was calculated. Underlying this calculation were the following five assumptions:

- o Each hospital in a PSRO area had the same proportional reduction in patient days.
- o All hospital costs are fixed (so that what appears as savings from reduced Medicare utilization are really just costs transferred to other patients and third-party payers, usually insurance companies).
- o The days of care eliminated had the same volume of ancillary charges as the average of all Medicare days.
- o For every 100 days of reduced hospital utilization, nursing home utilization increased by 15 days.
- o For every \$100 reduction of inpatient ancillary reimbursements, ambulatory ancillary **reimbursements** increased by \$50.

31/ Pre-PSRO expenses for utilization review were reimbursed on the basis of cost along with all other hospital expenses, and hospitals had little incentive to separate review costs from other costs. There are no reliable estimates of those costs available.

32/ Both of these expected values were calculated from the regression model.

Savings from reduced use of attending physicians were not estimated. The result of the calculation was an estimated savings of \$50.5 million.

Analysis

A number of the assumptions listed above deserve careful scrutiny. Some appear to make the savings-to-cost ratio too low, others make it too high, and still others could bias the estimate substantially in either direction. Discussion of the likely distortions follows.

Assumptions Causing the Savings-to-Cost Ratio to be Understated. The assumption that all hospital costs are fixed is too conservative. Studies of hospital costs suggest that about 60 percent are fixed. ^{33/} HCFA based its contrasting assumption on the fact that the average PSRO hospital experienced only a 0.5 percent reduction in days of care (Medicare accounts for roughly one-third of all short-term hospital days) and that such a small change is unlikely to affect hospital staffing and purchasing decisions. However, the studies of hospital costs mentioned above, which uniformly show fixed costs to be less than 100 percent of total costs, reflect the experiences of many hospitals with small changes in utilization over time. Assuming instead that only 60 percent, rather than all, of costs are fixed increases the estimates of savings.

PSRO impact may spill over to affect hospital use by non-Medicaid and non-Medicare patients, but no allowance is made for this. PSROs are alleged to work by educating physicians as well as by deterrence. Any changes in medical practice with respect to Medicare patients may carry through to treatment of private patients, but it is difficult to quantify this effect without a study. (The effect, however, could have been measured with a

^{33/} See Joseph Lipscomb, Ira E. Raskin, and Joseph Eichenholz, "The Use of Marginal Cost Estimates in Hospital Cost-Containment Policy" in Michael Zubkoff, Ira E. Raskin, and Ruth S. Hanft, editors, Hospital Cost Containment; Selected Notes for Future Policy (New York: Prodist, 1978), pp. 514-37 for a review of this literature.

small extension of the Medicare rate **study**.) Any such **effect** would increase estimated benefits, although it is impossible to say how much.

Costs may have been overstated because some may be attributable to the **start-up** of PSRO activities rather than to continuing concurrent review activities. Also, all concurrent review costs were included, rather than only those that were incremental to pre-PSRO utilization review activities.

Reduced hospital utilization may curtail attending physician services even more than it increases outpatient use. If so, savings would be increased.

Assumptions Causing the Savings-to-Cost Ratio to Be Overstated. The HCFA evaluation defines a savings as a reduction in Medicare **reimbursements**, rather than as a reduction in hospital costs. A mere shifting of costs from Medicare to Blue Cross or other third-party payers is recorded as a PSRO savings, although such a shift accomplishes nothing toward the ultimate goal of containing health-care costs. It seems clear that the federal government should not be investing substantial resources merely to shift costs onto the private sector. A reduction in resources used in hospitals is a more appropriate measure of savings to be compared with PSRO review costs. 34/ Using this more appropriate definition of a savings reduces estimated savings substantially.

No allowance was made for induced increase in **utilization** because of the so-called "**Roemer effect**"--that is, the tendency of empty hospital beds to generate demand for their use. 35/

34/ The **inappropriateness** of HCFA's definition of savings is seen in conjunction with their assumption of all hospital costs being fixed. In this case, it would be impossible for PSROs to reduce hospital expenditures. The savings-to-cost ratio would be zero. Nevertheless, a substantial dollar savings was erroneously calculated.

35/ The effect is named for Milton Roemer, an authority on the organization of medical care, who first suggested the existence of this phenomenon. See Milton Roemer, "Bed Supply and Hospital Utilization: A National Experiment," Hospitals, vol. 35 (November 1961), pp. 36-42.

Although hospitals cannot fill all of their emptied beds by inducing utilization increases, the research literature suggests that 40 percent will be filled. While the **Roemer** effect on Medicare **patients** is already taken into account in the rate study, the induced increase in non-Medicare utilization could offset 31 percent of the Medicare savings. This would reduce the estimated savings substantially.

In projecting the savings achieved for the entire PSRO system (assuming the inactive ones will become active), it is inappropriate to apply the savings-to-cost ratio calculated for PSROs that are now active to those that are currently inactive. As an example, hospital costs in areas where PSROs were inactive in 1977 are lower than in areas where they were active. Consequently, if the utilization reduction were the same, the savings would be smaller in the inactive areas, causing the savings-to-cost ratio to be lower.

Assumptions **Likely** to Affect the Savings-to-Cost Ratio in an Important but Undetermined Manner. Those patients accounting for the PSRO-induced reduction in utilization are likely to have different use of hospital **services--in** particular, ancillary **services--than** the average patient. These patients probably use fewer services than average, given the judgment that they do not need **hospitalization**. Some patients not needing hospitalization, however, make heavy use of ancillary services as outpatients. Overall, it is difficult to estimate the net tendency.

The assumption of a 50 percent substitution of outpatient for **inpatient** ancillary services has little empirical basis. **Unfortunately**, estimated savings are very sensitive to this assumption, but there is no basis for any alternative assumption.

Recalculation with More Appropriate Assumptions

The effect of changing the assumptions that underestimate or overestimate the savings-to-cost ratio can be roughly calculated. On the savings side, replacing the HCFA estimate of a **1.5** percent decline in utilization caused by PSROs with the **reestimate** of 2.0 percent increases savings by 33 percent. The HCFA savings estimates are based, however, on an inappropriate method of calculating individual PSRO effects (see **Appendix B**). Correcting this error reduces the estimated savings by 19 percent. Changing the fixed-cost assumption together with counting hospital cost

reductions rather than Medicare reimbursement reductions causes a net additional reduction of savings of 24 percent. The **Roemer** effect accounts for a 31 percent reduction. The fact that hospital costs are lower in the inactive PSRO areas than in the active areas reduces projected **systemwide** savings by 11 percent. Effects on private patient utilization through education are assumed to be negligible. Adding up all the adjustments in savings yields a total reduction of 50 percent.

More realistic assumptions reduce estimated program costs as well. Assuming that 10 percent of management costs are start-up costs results in a 3 percent reduction in total PSRO costs. Subtracting costs that would have been incurred under pre-PSRO utilization review is difficult but very important to do. Although in theory pre-PSRO review was similar to review by PSROs, it is **generally** acknowledged that this was not the case. Little information is available, however, on the relative costs of pre-PSRO and PSRO review. Assuming that pre-PSRO concurrent review costs were about 20 percent of PSRO **concurrent** review costs (including local management and support) reduces estimated program costs proportionately. The total cost reduction is then 22 percent. Combining the savings and cost reductions, the net effect is a reduction in the savings-to-cost ratio to **0.7-to-1**.

The savings-to-cost ratio calculated by HCFA is too high. Rather than producing savings slightly higher than their costs, PSROs yield savings that appear to fall far short of their costs. While both of the benefit-cost estimates are subject to substantial error, the range that has been established is not a very favorable one for the program.

THE MAGNITUDE OF NET SAVINGS RELATIVE TO THE SIZE OF THE PROBLEM

The reanalysis of the HCFA report described above concludes that the PSRO program probably yields a net loss. Because of uncertainties in estimating the savings-to-cost **ratio**, however, it is useful to compare the HCFA estimate of net savings with the magnitude of the problem the program is meant to **solve--that** is, the magnitude of Medicare reimbursements for short-term hospital care.

The HCFA report noted that PSRO concurrent review costs accounted for 0.75 percent of relevant Medicare reimbursements

(that is, reimbursements for inpatient care in short-term general hospitals) in the active areas. HCFA's savings-to-cost ratio of **1.1-to-1**, therefore, indicates that net savings are 10 percent of costs, or less than 0.08 percent of relevant **reimbursements**. Thus, even HCFA's more optimistic figures indicate that the PSRO program at its present level of **effectiveness** cannot be expected to cause a substantial reduction in Medicare expenditures.

WHY AREN'T PSROs MORE EFFECTIVE?

The PSRO system has been less successful to date in curbing utilization than many proponents had expected, and it has been even less effective in achieving a net reduction in expenditures. These failures cannot be attributed to a lack of inappropriate days of inpatient hospital care; on the contrary, various studies have estimated that 24 to 30 percent of all hospital stays involve inappropriate days of care and that about 11 percent of all days are inappropriate. 36/ Because PSROs and similar review systems still play a prominent role in many proposals concerning the health-care system, it is important to summarize some of the factors that may be limiting the **effectiveness** of PSROs. In some instances, clear evidence supports these **explanations**; in other cases, however, one can only speculate.

Characteristics of the Medical Care System

Some of the factors that may be limiting the **effectiveness** of PSROs are characteristics of the medical-care system itself.

36/ Such estimates vary substantially according to the criteria used and the population under study. These figures are therefore presented only as a rough index of the magnitude of the problem of excessive inpatient care. They are taken from two studies: Paul M. Gertman and Joseph D. Restuccia, Appropriateness Evaluation Protocol Development and Methodologic Testing of a New Technique for Studying Inappropriate Hospital Utilization, prepared for HEW, Health Care Financing Administration (1978); and Joseph D. Restuccia and Don C. Holloway, "Barriers to Appropriate Utilization of an Acute Facility," Medical Care, Vol. 14 (July 1976), pp. 559-73.

Some patients spend unnecessary days in the hospital not as a result of their own choice or the discretion of their physicians, but because of the way the medical-care system works.

A recent study by Restuccia and Holloway of a sample of patients in a single hospital clearly illustrates this point. 37/ They found that 201 (11 percent) of the 1,902 total **patient** days were inappropriate according to Medicare **levels-of-care** criteria; 53 (24 percent) of the 218 patients spent at least one inappropriate day as inpatients. Identifying and classifying the causes of inappropriate hospital use revealed that physicians were responsible for 42 percent of all inappropriate inpatient days and that problems in the health system caused another 42 percent. Most of the latter group involved unavailability of beds in skilled nursing facilities. Other examples included problems in hospital scheduling for tests and procedures and poor discharge planning.

Another characteristic of the medical-care system that might limit the **effectiveness** of PSROs is the Roemer effect discussed earlier in this chapter. Although PSROs might be able to reduce the size of the Roemer effect, no evidence yet confirms this possibility. Pre-PSRO data indicate that a 10 percent increase in empty beds induces, on average, a 4 percent increase in inpatient days. 38/ Thus, PSROs might have to produce a large gross reduction in days of hospital care to achieve a net reduction that would be substantially smaller. **Specifically**, the

37/ Restuccia and Holloway, "Barriers to Appropriate Utilization."

38/ For a discussion of the magnitude of the Roemer effect, see Martin S. Feldstein, "Hospital Cost Inflation: A Study of **Non-Profit Price Dynamics**," American Economic Review, vol. 61 (December 1971), pp. 853-73; J. Frederick McNeer and others, "Early Discharge After Myocardial Infarction: Reply," New England Journal of Medicine, vol. 298 (May 25, 1978), pp. 1199-1200; and J. P. Newhouse and C. E. Phelps, "New Estimates of Price and Income Elasticities of Medical Care Services," in Richard N. Rosett, **ed.**, The Role of Health Insurance in the Health Services Sector (New York: National Bureau of Economic Research, 1976).

2 percent net reduction estimated in the HCFA Medicare rate study would require a 3.3 percent gross reduction. 39/

Because of the **Roemer** effect, the net long-term reduction in utilization produced by PSROs depends in part on the activities of Health Systems Agencies (HSAs), the regional health planning agencies established by the National Health Planning and Resources Development Act (P.L. 93-641). There is little evidence, however, assessing the degree to which HSAs and PSROs have been coordinating their work so far. Because HSAs are responsible for planning and developing health facilities, they could be instrumental in eliminating beds made unnecessary by PSRO-induced reductions in utilization. The more stringent HSAs are in encouraging the closing of unused facilities and discouraging the construction of new facilities, the greater the resulting long-term cost reduction from PSRO activities. Conversely, if HSAs are not strict, the additional unneeded beds will generate additional inappropriate use, cutting into the effectiveness of PSRO review.

PSRO **effectiveness** is also related to the policies of agencies that set hospital rates. If hospital rates are set on a per diem or per admission basis, the hospitals have an incentive to increase utilization. In contrast, if total expenditures are the focus of the rate setters, the incentives will be in the opposite direction. PSROs could possibly be more successful with the latter type of rate setting. 40/

All these factors would contribute to limiting the effectiveness of PSROs in reducing utilization. In addition, some characteristics of the medical-care system may limit the cost savings yielded by whatever net reduction in utilization PSROs produce. The large proportion of hospital costs that are fixed is one example. Another is what can be termed the "spillover" effect: increases in other types of care (for example, ambulatory care) that result from, and partly compensate for, savings in hospital use.

39/ $3.3\% - (0.4 \times 3.3\%) = 2.0.$

40/ The HCFA Medicare rate study showed utilization to be higher in areas with mandatory hospital rate setting, but it did not find any relationship with PSRO effectiveness. See HCFA, PSRO! 1978 Evaluation, p. 85.

Characteristics of PSROs

The effectiveness of PSROs may also be limited by characteristics of the PSROs themselves. Some of these traits are inherent in peer review in general and **might** be intractable; others are specific to the PSRO system and could be changed more easily.

In two different senses, peer review as a method of controlling utilization relies on physicians and hospital staffs to act against their own **self-interests--in** the view of some observers, "placing the foxes in charge of the chicken **coop.**" First, it is clear that both physicians and hospital staffs working in a **fee-for-service** system often stand to lose financially as a result of decreases in utilization. Although individual physicians cannot review cases in which they are involved, physicians as a group would stand to lose if PSROs were too stringent. Second, medical care **use--especially** hospital **use--is** determined in large measure by **physicians'** standards. Peer review may alter utilization by patients of physicians whose standards are substantially different from the norm, but such review is unlikely to effect major changes in the standards of physicians as a group.

One characteristic of the PSRO system exacerbates this problem. The law mandates that PSROs "delegate" responsibility for review to hospitals that demonstrate the ability to perform review. Review was fully or partially delegated to 76 percent of all hospitals in which PSRO review had been implemented as of June 30, 1977. 41/ While delegated review is somewhat less expensive than review by PSRO personnel, 42/ it may also be less effective because hospital review staffs are likely to have stronger incentives to maintain high levels of utilization.

The results of the 1978 Medicare rate study are consistent with the speculation that physician and hospital control of PSROs may be limiting their **effectiveness**. As was noted earlier, the **effectiveness** of PSROs varied seasonally; they appeared to

41/ HCFA, PSRO; 1978 Evaluation, p. 4.

42/ As of fiscal year 1977, the medians were \$8.76 for delegated and \$10.94 for nondelegated concurrent review. HCFA, PSRO; 1978 Evaluation, p. 152.

have substantial effects in the first half of the year but very little in the second half. This pattern mirrors seasonal differences in utilization rates; in this sample, **utilization** was about 8 percent higher in the first half of the year than in the second. The link between seasonal fluctuations in utilization and PSRO **effectiveness** may derive from the behavior of hospital staffs and physicians. Some observers maintain that physicians are much more amenable to curtailing excess utilization during periods when hospitals are in heavy use; decreasing utilization at such times enlarges the supply of empty beds for their patients and makes scheduling procedures easier. Likewise, hospital administrators have stronger incentives to maintain high utilization during periods of low occupancy and they may even pressure physicians to that end.

Although **fee-for-service** practice in some instances gives providers and practitioners **incentives** to provide too much care, the PSRO law sets up sanctions that ideally should discourage such excesses. As described in Chapter I, HEW may, on the advice of a PSRO, exclude offending physicians and hospitals from the Medicare and Medicaid **programs**, and the department may exact a \$5,000 fine. It is possible that the failure to date of the department to implement these provisions of the law may have lessened the effectiveness of the PSRO system.

Regardless of the incentives built into peer review and the sanctions available, there are practical limits on the potential **effectiveness** of any utilization review system caused by the way review is carried out. Reviewing each case on every day of a **patient's** hospital stay would be prohibitive, so all utilization review systems examine each case only on a few **days**. Typically, a PSRO reviews a patient within a day of his admission and a date is then set for the first "extended-stay" or "continued-stay" review. The date is chosen to correspond to a given percentile point in the local distribution of lengths of stay **for** the diagnosis of the patient whose care is under review. The percentile points used vary from PSRO to PSRO. For example, the Southeastern Wisconsin Foundation for Medical Care sets its first extended-stay review at the median length of stay based on regional norms. 43/ Accordingly, any inappropriate hospital

43/ OPEL, PSRO, Vol. VI: Case Studies of PSROs: A Contextual Analysis, p. 1-16.

days occurring before the median day will go unnoticed by the review process. This problem can be minimized by shortening the time between rereviews of each case, but that would increase the cost and therefore possibly undermine the **cost-effectiveness** of review.

Another factor that may limit the **effectiveness** of PSROs in reducing use and costs is that, until recently, **PSRO** utilization review has been largely unfocused. That is, PSROs were instructed to try to review all admissions and extended stays. In many instances, this resulted in the review of cases in which there was no room for a PSRO effect. This could be due to a variety of factors, such as the routine nature of a procedure, the particular diagnosis involved, or the policies and customary practices of a given physician or institution. Such useless reviewing presumably increases the costs of the PSRO program without improving its **effectiveness**. Current policy, shaped in part by budgetary constraints, is to increase the degree to which review is focused on those cases in which review is likely to be most effective. If PSROs are successful in reducing unnecessary review without simultaneously eliminating some of the needed and effective review activities, the cost-effectiveness of the program will be increased. Chapter III, however, explains why the current method of changing to focused review virtually guarantees that it will be impossible to evaluate the impact of this change on the effectiveness of the program.

Those activities of PSROs that are primarily oriented toward quality **assurance--in** particular, Medical Care Evaluation **Studies--may** also lessen the **effectiveness** of the program in reducing costs. These activities may increase utilization in some instances, as physicians attempt to meet newly established criteria of quality. Concurrent review, however, which has been the primary emphasis of the PSRO program to date, is unlikely to have a major effect of this sort. Whether Medical Care Evaluations will exert appreciable upward pressure on utilization as they become more common in the future is beyond **the** scope of this report.

The two principal conclusions of Chapter **II--the** small size of the apparent effects of PSRO concurrent review and the uncertainties in the evaluation of those **effects--point** to two questions about the future of the program:

- o What are the options for the program? Should it be continued as is, modified substantially, or terminated altogether?
- o What are the options for future evaluations of the program?

These two questions are closely interrelated. Reliable assessments of any changes in the program would be invaluable as a guide to the future design of this and similar programs. Furthermore, how program changes are **implemented--regardless** of what changes are **made--determines** whether conducting reliable and useful evaluations of the effects will be possible.

PROGRAM OPTIONS

PSRO concurrent **review--and** utilization review in **general--** is just one of many ways to try to curb health-care use and expenditures. For example, efforts to control expenditures have included the following four approaches:

- o Supporting the growth of Health Maintenance Organizations (HMOs) as an alternative to **fee-for-service** arrangements;
- o Establishing controls on the revenues of health-care **providers;**
- o Limiting the supply of providers and practitioners by means of manpower and health-facilities planning (for example, through Health Services Agencies); and

- Structuring health insurance in ways that would discourage excessive utilization. 1/

Although a discussion of these alternative approaches is beyond the scope of this paper, it is important to note that the importance and effectiveness of PSRO concurrent review may depend as much on other cost reduction efforts as on changes made in the PSRO program itself. For example, Chapter II suggested that the long-term **effectiveness** of PSROs might be enhanced if Health Services Agencies coordinated their work with that of local PSROs and were stringent in closing down any health-care facilities made superfluous by PSRO-induced reductions in utilization. Conversely, a major increase in the number of HMOs could make PSRO concurrent review superfluous, since HMOs have strong incentives of their own to limit utilization and costs.

A number of options for changing the PSRO program are apparent. One is to end PSRO concurrent review altogether. The analysis in Chapter II suggests that this might save a small amount of money in the short run. It would also, however, eliminate the possibility of future savings resulting from the development of more **cost-effective** methods of review.

If PSRO concurrent review is continued, the issue is whether the program can be changed to make it more **cost-effective**. While the research now available offers little guidance for choosing the most **cost-effective** new directions for the program, several options are clear.

1/ Ways of restructuring health insurance to discourage excessive health-care use involve more extensive use of cost sharing by the patient. Cost-sharing can be obtained through deductibles, coinsurance, or indemnity **provisions**. Deductible provisions require the patient to pay the first so many dollars of covered medical expenses per year. Coinsurance requires that the patient pay a percentage of the expenses (for example, 20 percent). Indemnity provisions require the patient to pay the difference between the price and an allowed amount for each unit of service.

Focusing Concurrent Review

Rather than reviewing all patients in a given region, PSROs could focus concurrent review on those areas (diagnoses, physicians, **institutions**, or patient groups) in which utilization seems most excessive or in which PSRO review has the greatest effect. Budgetary constraints have contributed to the rapid shift toward focused review now under way. Development of the **PSROs'** profile analysis should further facilitate this trend.

Focusing clearly holds a promise of increasing cost-effectiveness. There are many ways to focus review, however, and which way is best is not clear. For example, certain methods of focusing might inadvertently reduce some necessary and effective review in the course of eliminating superfluous review. Likewise, it is not known how narrow the focus can be before concurrent review loses its presumed deterrent effect. It is also unclear how much focusing would increase **cost-effectiveness** even if it were carried out in an optimal way. Unfortunately, the way focused review is being **implemented--with** individual PSROs free to choose not only how to focus, but also how much to **focus--**virtually guarantees that these questions will not be answered reliably. As a result, designing the most effective focusing system will be hindered.

Ending or Curtailing Delegated Review

That delegated review is, on average, cheaper than non-delegated review is clear. On the other hand, delegated review presumably involves more potential for conflict of interest since it is conducted by hospital personnel, and it might therefore be less effective. At present, however, there is no conclusive evidence on the relative **effectiveness** of delegated and **nondelegated** review. ^{2/} Accordingly, what the most cost-effective mix of delegated and nondelegated review would be is unclear.

^{2/} The HCFA Medicare rate study reported no difference in **effectiveness** between delegated and nondelegated review (PSRO; 1978 Evaluation, p. 79). That analysis, however, is inconclusive for several technical reasons.

Shifting Responsibility for Utilization Review

Since insurers presumably have stronger incentives to curb utilization and costs, utilization review could be conducted by insurers rather than physicians. In the case of, Medicare and Medicaid patients, the insurer is the government. The impact of such a change on the **effectiveness** of review is difficult to predict, however. Furthermore, many observers would argue that decisions about standards of the appropriateness of care are more properly made by organizations of physicians.

OPTIONS FOR EVALUATION

Any changes made in the PSRO program can be linked to improvements in the evaluation of the program. Improvements in evaluation would be important because the shortcomings in the existing **research--and the** caveats necessarily attached to even the best existing **evaluations--seriously** limit the usefulness of the research in designing a more **cost-effective** program. Moreover, if changes in evaluation strategies are not made soon, future evaluations of the program will probably be no **more--and** sometimes considerably **less--reliable** than the existing studies.

The 1978 HCFA Medicare rate study, while the best study of the effectiveness of PSROs now available, has a number of major shortcomings that suggest ways future research could be improved. First, the reliability of its conclusions is limited for two reasons: the areas with active PSROs (the treatment group discussed in Chapter II) were self-selected, and there are no data on patterns of change in hospital utilization before the PSRO program. Second, the **study's** scope was in some respects circumscribed: the analysis was restricted to concurrent review of inpatient care of Medicare **beneficiaries.** **HCFA's** conclusions cannot be generalized to other PSRO activities or to other populations. Third, the study was limited by the absence of important background material, such as information on the relationship between utilization and costs.

Outlines of several possible improvements in each area follow. Some of these could be made on agency initiative alone, while **others--those** involving changes in **implementation--might** require legislative action as well.

Improvements in Program Implementation and Evaluation Design

Accurate and reliable program evaluations require certain strategies of program implementation, and usually it is not possible after the fact to compensate fully for implementation strategies that are undesirable from the point of view of evaluation. This can be problematic in that the methods of implementation that are best for purposes of evaluation are sometimes poor with respect to political or administrative **considerations**. Hence, reliability of evaluation often has to be weighed against efficiency in administration. Factors going into such decisions might include the urgency **of** the program's implementation, prior assumptions about the **program's** value, and the loss of accuracy from limitations in evaluation design.

The manner in which the PSRO program has been implemented so far (as noted in Chapter II) has made reliable evaluation difficult to achieve. Partly because of the method of implementation, many evaluations are simply too unreliable to be of much use in deciding the program's future course. Even the 1978 HCFA national Medicare rate study suffers from limitations caused by the program's implementation; these problems place in some doubt the **study's** conclusion that the program is effective in curtailing utilization. Consideration should be given to implementing future PSRO activities in such a way as to permit more reliable evaluation.

Tailoring implementation plans to suit the needs of evaluation would involve switching to a course of planned partial implementation, with some areas (the treatment group) deliberately chosen to have new aspects of the program and others (the comparison group) chosen not to have them. In all cases, the optimal strategy for purposes of evaluation is random assignment (of hospitals or PSROs, in this case) to treatment and control groups. Other, less restrictive strategies exist, but the reliability of evaluations falls off very rapidly as studies depart from random assignment. 3/ The greater the concern about

3/ See, for example, D. T. Campbell and R. F. Boruch, "Making the Case for Randomized Assignment to Treatments by Considering the **Alternatives**: Six Ways in Which Quasi-Experimental Evaluations in Compensatory Education (continued)

the accuracy and reliability of **evaluations--as** distinct from other factors of **importance--the** more strictly the implementation should adhere to random assignment.

Depending on the evaluation design employed, it can be important to measure utilization (or any other outcome measure) repeatedly over a considerable period before implementing new program components. (In **the** case of utilization, which shows a marked variation over time, it is best to have such repeated measures extending back several years **before** program implementation.) 4/ If assignment is truly random and the sample used is large, such repeated pre-program measures are not essential, though they can nevertheless be helpful. The further the research design is from random assignment, however, the greater the need for repeated pre-program measures. This is strikingly illustrated by the CAP and CHAMP **PSRO-prototype** evaluations discussed in the previous chapter; one can reach entirely different conclusions, depending on whether or not one looks at patterns of change before implementation of the program.

These **implementation** and evaluation strategies can no longer be applied to PSRO review (as it is now conducted) of admissions to short stay hospitals and continued stays, since this activity has already been implemented in most of the nation. Accordingly, the 1978 HCFA rate study is likely to remain the most reliable evaluation ever produced of this type of PSRO review. But as PSRO review is extended to the **full** range of services covered in the **law--that** is, ambulatory care, hospital ancillary services, and long-term **care--HEW** will have the opportunity to tailor the phasing in of each new aspect of review to fit the needs of evaluation. In each case, the opportunity will be short-lived, because as implementation **proceeds**, constructing an appropriate comparison group will become increasingly **difficult--** and eventually impossible.

3/ (continued) Tend to Underestimate Effects," in C. A. Bennett and A. A. **Lumsdaine, eds., Evaluation and Experiment** (New York: Academic Press, 1975), pp. 195-285.

4/ This need not always involve waiting several years before implementation, for in some cases the relevant data may already exist.

Several extensions and alterations of the program that are currently under way are good candidates for a strategy of planned partial implementation. Review of ambulatory care has been undertaken in a few demonstration projects and in a few other PSROs (roughly 10 in all), but it has been stalled by lack of funds. Review of ancillary services has also been undertaken by only a very few PSROs. Review of long-term care has been undertaken by a few more PSROs (roughly 30 in all), but so far enough PSROs have not been involved to make it practical to arrange large treatment and control groups. Focused review, in which PSROs review only a fraction of all cases, is somewhat further along. HCFA is urging PSROs to focus review, and budgetary constraints will force some degree of focusing. How to focus, however, as well as to what extent, is for the most part left to individual PSROs. For example, PSROs with high per-review costs would be expected to focus more in order to save money; PSROs in low-utilization areas can afford to focus more because the need for review is presumably less severe. This is, of course, an extreme case of **self-selection**, and no reliable comparison of areas with and without focusing will be possible if the present course continues. Some manner of planned partial implementation, however, may still be feasible. For example, different PSROs could be randomly assigned to undertake different degrees or types of focusing.

Recent innovations in the criteria used in concurrent review could also be subjected to vigorous evaluation. Concurrent review is usually based in large part on **diagnosis-specific** norms for length of stay. Several PSROs, however, are pilot testing an alternative **system--devised** by a private consulting **firm--that** uses severity of illness, intensity of service, and discharge screening criteria as well as information on diagnosis. The new system has received considerable attention as potentially more effective and less expensive than traditional review. 5/ To date, however, there has been no rigorous evaluation of the system. To institute such an evaluation might require legislative action, since current law gives PSROs considerable autonomy in choosing their own criteria for review.

5/ See, for example, Suzanne LaViolette, "New Criteria Cut PSRO Cost and Length of Stay," Modern Healthcare, May 1979, p. 42.

Planned partial implementation with random assignment has an additional advantage that is particularly important in times of limited resources. In short, partial implementation can be inexpensive because the program can initially be restricted to a small number of representative sites. Nationwide implementation can await the results of the evaluations made possible by partial implementation and can be restricted to those program components that have already been demonstrated to be effective.

Broadening the Range of Research Questions

Overall evaluation of **the** PSRO program would also be greatly strengthened by broadening the range of questions addressed. **Specifically**, six additional areas could be investigated:

- o Background research,
- o The PSRO review process,
- o Additional types of PSRO review,
- o Use of additional outcome measures,
- o Spillover effects, and
- o The relationships between utilization of health-care resources and costs.

In some instances, research in the areas mentioned above has been or is now being conducted, and the task is to **integrate--and** perhaps to **supplement--existing** research rather than to initiate wholly new efforts.

Background Research. As indicated in Chapter I, the potential **effectiveness** of PSROs as a mechanism for controlling health-care expenditures depends on the nature, scope, and causes of inappropriate utilization. The term "background research" is used to encompass investigation of the inappropriate utilization that PSROs are meant to affect. Within this broad category, two types of questions need to be addressed: the extent and the causes of inappropriate utilization.

In order to gauge the extent of inappropriate utilization, it is first necessary to assess whether a reasonable consensus exists in the medical community that would permit the establishment of firm, explicit standards of appropriate utilization. Assuming there is not, how important are those areas in which a consensus is lacking?

If explicit standards could be set, one would have to ask: What is the scope of overutilization? How many hospital inpatient days or visits to physicians are involved? What are its most important forms (for example, overuse of ancillary services, or overly long length of **stay**)? What are the correlates of overutilization: Does it vary with the type of hospital? With **patients'** characteristics? With physicians' characteristics? With diagnoses? Does overutilization in the case of Medicare and Medicaid patients differ qualitatively or **quantitatively** from **overutilization** in other segments of the patient population? While the cost-containment focus of the PSRO legislation would no doubt be reflected in a preponderance of research on overutilization, poor quality of care associated with underutilization should be investigated as well.

Once inappropriate utilization has been identified, it is crucial to explore its causes. To what degree is it attributable to matters of **physicians'** discretion? To hospital policy? To families' or patients' preferences? To what extent is it not a matter of **anyone's** discretion, but rather a function of lack of access to beds in skilled nursing or intermediate care **facilities** for federal program **beneficiaries**?

Research on such questions need not be prohibitively expensive, since a relatively small sample of hospitals, carefully selected to represent hospitals nationwide, would be adequate for most research of this type. 6/

6/ An example of productive background research is Restuccia and Holloway, "Barriers to Appropriate **Utilization**," described in Chapter I. This study provided valuable information on the extent of inappropriate utilization, its relationship to length of stay, **and--perhaps** most importantly--its causes.

Research on the PSRO Review Process. The OPEL evaluation was designed to measure the incremental effect of PSRO review over and above the effects of pre-existing utilization review. Neither PSRO nor non-PSRO review, however, is a homogeneous category. For example, the formal review process clearly differs from one PSRO to another. ^{7/} Perhaps more important, PSRO review is likely to differ, even within a given PSRO, from case to case. **Nevertheless,** even **OPEL's** case studies do not **provide** a means of gauging the degree of case-to-case variation. Comparison (non-PSRO) areas also differ in terms of the nature and extent of review activities. Accordingly, pinpointing precisely what is involved in a contrast between PSRO and non-PSRO review is impossible.

This gap in understanding of the contrasts involved is crucially important for two reasons. First, it would tend to obscure true PSRO incremental **effects--though** to an unknown degree. Second, PSRO review comprises myriad diverse activities that presumably differ in their effectiveness as well as their costs. For example, review of admissions might have more effect on total days of care than does review of continued hospital stay. From a **policymaker's** perspective, the relative cost-effectiveness of various PSRO components is of great importance.

Accordingly, it would be a substantial improvement if future PSRO evaluations were to examine not only the outcomes of PSRO review (in the aggregate), but also the specific processes that are part of review. The range of processes that could be assessed profitably is large; it would include, for example, the timing of review, the focusing of review on diagnosis and other dimensions, the nature of the feedback process, the types of care given to replace inappropriate care that has been prevented, and so forth. **As** in the case of background research, research on specific processes should be tailored to address questions posed by current or planned changes in the PSRO program.

Assessment of Additional Types of PSRO Review. In the future, as PSRO review is extended to care in settings other than short-term hospitals, opportunities will arise to evaluate new review activities effectively.

^{7/} See OPEL, PSRO, Vol. VI.

HEW has already initiated evaluations of some new review activities. For example, the agency has undertaken demonstration projects in long-term and ambulatory care. The **issue** of new areas of evaluation is therefore raised in this study for only two reasons. The first is the concern stated earlier that if these new evaluations are not linked to appropriate strategies of planned partial **implementation--at** the time when new activities are first being **instituted--the** results are likely to be unreliable. Second, evaluations of new areas of review present two important **possibilities**: they can contribute to an assessment of "spillover" effects (discussed below) and, if properly integrated, they can permit measurement of intensity of services per day of care, an important aspect of overall utilization.

Use of Additional Outcome Measures. So far, PSRO evaluations have focused primarily on hospital admissions, length of stay, and days of care. A complete evaluation of **PSROs'** impact requires the use of another variable as well: intensity of services. The rapid growth in the use of diagnostic tests and X-rays in recent years suggests a potential for significant cost savings by peer review.

Some writers have argued that PSRO activities could not only reduce utilization by identifying unnecessary services but also increase utilization by drawing attention to underutilization or by raising standards of care. 8/ This compensatory increase seems unlikely to be an appreciable factor so long as PSRO activities are restricted to the sort of concurrent review activities now being conducted. It would seem to be a real possibility, however, in the case of quality-oriented medical-care evaluations. Accordingly, it would be good in future evaluations to include measures of the variability of utilization, rather than to assess only differences in mean levels of utilization. In this way, researchers might be able to infer PSRO impact from an observed reduction in variability even if the mean utilization remained unchanged.

8/ For example, C. C. Havighurst and J. F. Blumstein, "Coping with Quality/Care Cost Trade-offs in Medical Care: The Role of PSROs," Northwestern University Law Review, Vol. 70, No. 1 (1975), pp. 6-68.

Assessment of Spillover Effects. In order to gauge the total effectiveness (and cost-effectiveness) of PSROs, so-called "spillover," or secondary, effects must be assessed. An example of a spillover would be an increase in ambulatory care that might accompany a decrease in the use of short-term inpatient care. The former increase would have to be subtracted from the latter decrease to obtain a measure of total effectiveness.

The magnitude of spillover effects cannot yet be estimated precisely, but preliminary estimates suggest that the effects may be large. The Institute of Medicine study of quality assessment cited two studies of prepaid health plans which indicated that savings effected in inpatient care were entirely offset by increases in primary and other ambulatory care. 9/ The benefit-cost analysis was shown to be highly sensitive to assumptions concerning the magnitude of spillover.

Assessment of spillover effects is dismissed by some as prohibitively expensive because of the complexity and volume of data required. A properly constructed representative sample, however, could be an adequate substitute for a comprehensive nationwide assessment, reducing this problem substantially. A more troubling problem is that the opportunity for sound evaluation of many of the most important spillover effects has already passed. For example, since concurrent review of short-term hospital care is now underway in the great majority of PSRO areas, it is no longer feasible to assess the degree to which the decreases in hospital days of care induced by that activity were accompanied by increases in days of care in nursing homes. **Nevertheless,** important possibilities for assessing spillover effects remain. For example, if PSRO review is extended to long-term care, it would be **feasible--and important--to** assess concomitant changes in ambulatory care.

Research on the Relationship Between Utilization and Cost. Chapter II outlines a series of reasons why cost reductions were not likely to be proportional to utilization reductions. There are a number of research projects that could be performed independently of individual PSRO evaluations that might improve the reliability of cost-reduction estimates.

9/ IOM, Assessing Quality, pp. 60-61.

First, knowledge of the relationship between fixed and variable costs in hospital care is limited and often misinterpreted. Most studies are conducted with annual or semi-annual time series data. 10/ Their estimates of the fixed-cost proportion are relevant for fluctuations in occupancy rates. They are not particularly germane to the gradual but permanent declines in use that may be induced by PSROs. In such cases, the proportion of fixed costs is likely to be lower. Studies more appropriate to the longer time frame associated with PSRO activities would be useful.

A research topic that should get higher priority is the relationship between different types of utilization reductions and cost reduction. Such analysis would compare the savings from admission-rate reduction and **length-of-stay** reduction, for example. In addition, reduction in intensity of service should be examined. Paralleling all this, relationships between utilization reduction and cost reduction should be estimated separately for various types of patients.

The most practical method of investigating these relationships might be to combine hospital discharge abstract data (already in use in PSRO evaluations) with hospital charge data. The latter should be adjusted for variation in the ratio of charges to costs on a **department-by-department** basis. Such a data base would allow researchers to contrast various types of utilization reductions by tracing through reductions in services charged for and translating them into cost reductions. Needless to say, such information would be of use to PSRO managers in allocating review resources as well as to evaluators.

10/ For a review of this literature, see "Use of Marginal Cost Estimates," Lipscomb, Raskin, and **Eichenholz**.

APPENDIX A. ANALYSIS OF THE COLORADO ADMISSIONS **PROGRAM** AND
COMMONWEALTH HEALTH AGENCIES MONITORING PROGRAM

The Colorado Admissions Program (CAP) and the Commonwealth Health Agencies **Monitoring** Program (CHAMP) were cited in Chapter II as particularly important evaluations of individual PSRO prototypes because of their prominence and because of the methodological concerns they illustrate. In particular, they exemplify the importance of repeated baseline measurements when comparing changes in **non-equivalent** treatment and comparison groups. When the groups are not equivalent, they may show different patterns of change before the start of the program that may not be apparent from a single baseline measure. This was the case with CAP and CHAMP, and the conclusions one reaches about their effectiveness depend entirely on whether one considers **pre-existing** patterns of change or merely **pre-existing** static differences.

CAP

The CAP program was begun by the Colorado Foundation for Medical Care (CFMC) before the establishment of the PSRO program, but its activities were essentially the same as PSRO concurrent review, and it became one of the first official PSROs once the program was underway. An extensive analysis of CAP was included in the 1977 OPEL **report. 1/** The analysis examined rate data for both Medicare and Medicaid populations. Kansas and Nebraska together, where no PSROs or PSRO-type organizations were active, served as a comparison area.

Two methods were used in analyzing the Colorado Medicare data. The first was a simple pre/post design with no control, in which utilization rates before and after the implementation of CAP were compared. The second method used **Kansas/Nebraska** as a comparison area, since changes in utilization in those states could not have been caused by CAP.

1/ OPEL, **PSRO**, Vol. V: A Comprehensive Case Study: The Colorado Experience.

The pre/post, no-control analysis warrants little attention. The general weakness of such an approach is its tendency to confuse historical trends with program effects. Furthermore, in this specific instance, a pre/post comparison would be of interest only if it showed that the implementation of CAP was accompanied by an acceleration of the historical trend toward lower utilization rates. The report showed just the opposite, however: a tapering off of this trend, with essentially no change in utilization from the introduction of CAP in 1973 to 1975.

Because of severe **limitations** in the availability of **Kansas/Nebraska** Medicaid data, OPEL found it impossible to perform a rigorous comparison of Medicaid utilization in Colorado and Kansas/Nebraska. Hence, the Medicaid analysis is also a pre/post, no-control design and is accordingly not considered here.

The key section of the OPEL report is that in which changes in Colorado Medicare days-of-care rates were adjusted for simultaneous changes in **Kansas/Nebraska**. That analysis was appropriately accompanied with the caveat that "the results of this analysis are a **direct** function of the choice of a normalization [adjustment] model and the selection of the base **period.**" 2/ In other words, the adjusted effects obtained in this type of analysis depend on which areas one chooses as a comparison (see Chapter II) and the period of time selected as the baseline for the analysis. In order to illustrate the importance of these choices for the interpretation of the results, a key finding from the Colorado report is analyzed here in some detail.

The analysis in the Colorado report that appears to be most important is one in which days of care per 1,000 Medicare **enrollees** in 1974 and 1975 (after the establishment of CAP) were compared to the rate during a four-year (1969-1972) baseline, adjusting for comparable trends in Kansas/Nebraska. The method of adjustment was to calculate the percent change in days of care in Colorado for a given period and to subtract from that figure the comparable percent change in **Kansas/Nebraska**. In the OPEL report, the 1974 and 1975 rates are compared to a single average rate for the four-year baseline. Tabulated in this fashion,

2/ Ibid. , p. 64.

utilization in Colorado (before adjustment) showed a 12.0 and 10.5 percent reduction from baseline in 1974 and 1975, respectively. After adjustment for trends in Kansas/Nebraska, the figures shrank to 7.6 and 3.1 percent, respectively.

The data can present a substantially **different** picture, however, if they are examined on a year-by-year basis rather than **with** the average of 1969 through 1972 as a single baseline. Figure 1 shows the Colorado and Kansas/Nebraska utilization rates, along with the difference between the two. Table **A-1** presents the corresponding figures, along with the percent changes in each from year to year. It is clear that the divergence of the Colorado and Kansas/Nebraska rates occurred during 1972 and 1973. The difference merely returned to its 1969 level in 1972, and 1973 saw a substantial further widening of the gap. In 1974, the difference essentially remained at the 1973 level, while in 1975 it returned to close to the 1972 and 1969 levels. Thus if one compares the adjusted utilization for the most recent year (1975) to the last baseline year (1972) rather than to the four-year baseline average, one finds very little **effect--just** 0.8 percent. If one compares 1975 to the "transitional" year of 1973, one actually finds an increase in utilization of 3.5 percent, which is somewhat larger than the 3.1 percent decrease found when 1975 was compared to the 1969-through-1972 average.

These figures illustrate the importance of the second of **OPEL's** caveats noted above: the results of this type of analysis depend on the selection of a base period. Depending on that choice, one can find a negative program effect, a positive effect, or no effect at all. Which of these conclusions can be accepted as the most reasonable?

To resolve this question, it is helpful to view Table A-1 and Figure 1 in the context of the CAP **program's** history. The CFMC was established in mid-1970 as a statewide peer review organization, with control and reduction of utilization (initially, among Medicaid patients) as one of its charter functions. CAP was set up by CFMC as a mechanism for reviewing inpatient stays of both Medicare and Medicaid patients. In mid-1973, CAP was awarded an Experimental Medical Care Review Organization (EMCRO) grant; the phasing in of CAP review began in mid-1973 and was completed in January 1974. In July 1974, CAP was designated a conditional PSRO. OPEL reported that this designation did not result in immediate changes in CAP operation.

TABLE A-1. COLORADO UTILIZATION DATA COMPARED WITH KANSAS/NEBRASKA DATA:
DAYS OF CARE PER 1,000 MEDICARE ENROLLEES

	1969	1970	1971	1972	1973	1974	1975
Colorado	4,394	4,171	3,963	3,780	3,612	3,587	3,647
Percent change from previous year	N/A <u>a/</u>	-5.1	-5.0	-4.6	-4.4	-0.7	+1.7
Kansas/Nebraska	4,868	4,454	4,270	4,241	4,233	4,261	4,126
Percent change from previous year		-8.5	-4.1	-0.7	-0.2	+0.7	-3.2

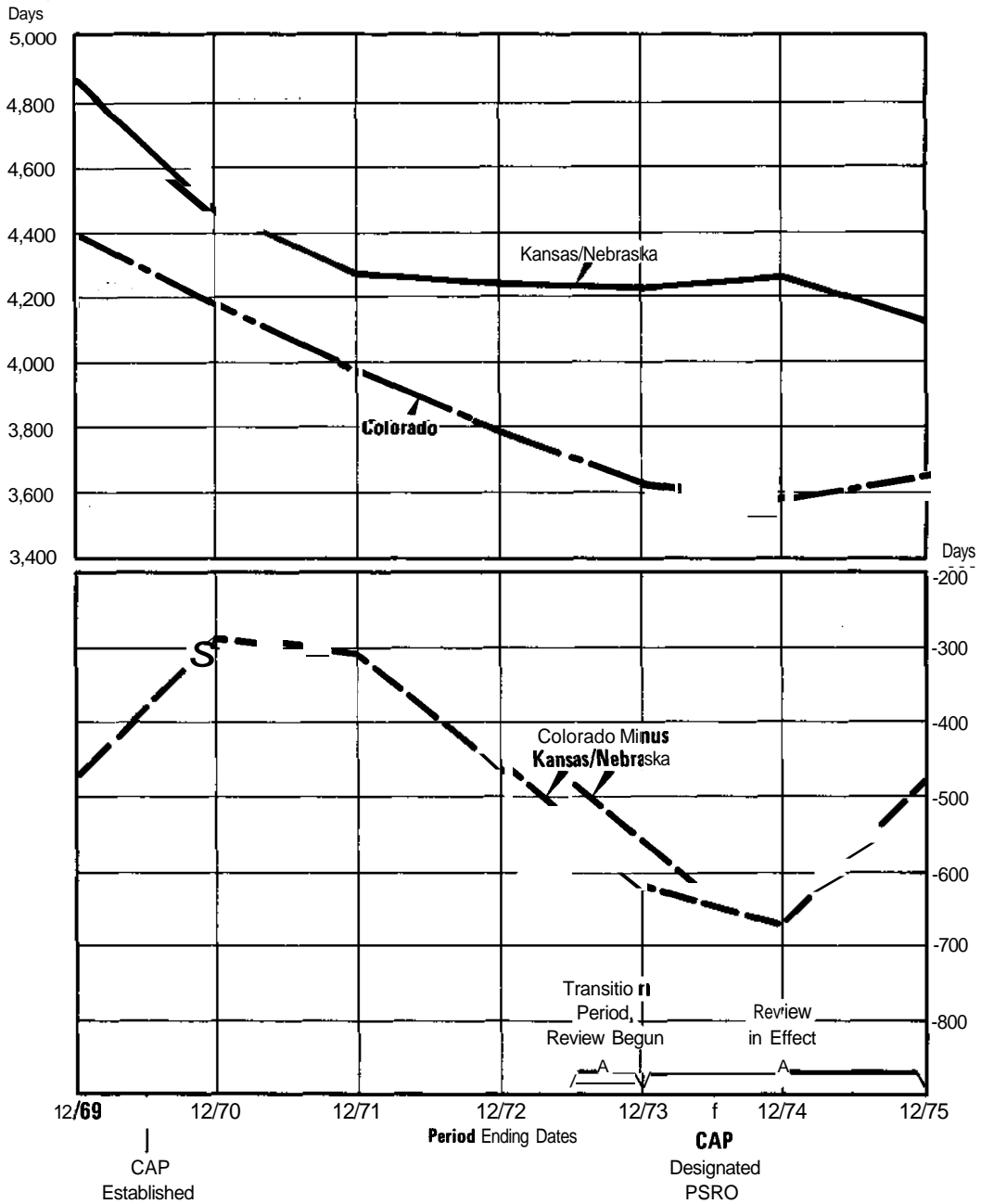
Colorado Minus Kansas/Nebraska	-474	-283	-306	-461	-621	-674	-479
Percent change from previous year	N/A	-40.3 <u>b/</u>	+8.5	+50.2	+34.7	+8.5	-28.9

SOURCE: Adapted from OPEL, PSRO, Vol. V, Table 5-E, p. 69.

a/ Not applicable.

b/ A negative sign in this row indicates a negative program effect; that is, that the Colorado utilization rate rose relative to the Kansas/Nebraska comparison rate during the year in question.

Figure 1.
 Hospital Utilization Data for Kansas/Nebraska and Colorado in
 Days of Care per 1,000 Medicare Enrollees; 1969-1975



The **classification** of 1973 as a transitional year--**neither** pre- nor **post-implementation--seems** appropriate. Similarly, 1972 could be called the last pure baseline year. Thus, the year-by-year analysis of changes in the Colorado-Kansas/Nebraska difference shows that the biggest decline in Colorado utilization rates during this span occurred the last baseline year, and the second largest decline occurred during the transitional year. The first post-implementation year showed a further small decline, followed by a larger increase in the second year.

The OPEL analysis of the data has the effect of masking these year-by-year patterns. The 1972 decline in utilization becomes an indistinguishable part of the single 1969-1972 baseline measure. Since 1973 was treated as a transitional year and therefore neither pre- nor **post-**, it did not appear on the OPEL table; 3/ the effect is to make the 1973 decline in utilization appear as a component of the 1974 figures. This approach is not justified in the light of the year-by-year trends described **here.** 4/

These data are open to a number of alternative explanations that explicitly do consider the year-by-year variation. One interpretation would be that the 1972 and 1973 changes reflect disparities between the experimental group (Colorado) and the comparison group (Kansas and Nebraska). That is, the areas may have differed in ways that had nothing to do with the PSRO program but that caused different patterns of changes in their utilization rates. A second interpretation would attribute the change to non-CAP activities of **CFMC--for** example, educational efforts. A third view explains the 1973 changes as a true effect of CAP; that is, it argues that CAP had real effects on utilization but that those effects were primarily manifested during the transition year rather than in the first full years of program operation. This last interpretation, however, must be qualified by noting that the transition year gains were not **augmented--** indeed, they were eroded **somewhat--during** the first two post-

3/ OPEL, Vol. V, Table 5-J, p. 77.

4/ In principle, if not in practice, OPEL agreed with this criticism. See, for example, Vol. I, p. 106, and Vol. V, p. 15.

transition years. Furthermore, it does not explain the 1972 (pre-transition) drop in utilization, which was larger than the 1973 change (see the bottom row of Table A-1).

The Colorado report does not provide a firm basis for choosing between these alternative interpretations. Data comparing Colorado with Kansas/Nebraska on a few relevant points were presented, however, and those few points indicate that in some **significant** respects, Colorado differed from Kansas/Nebraska not only in baseline **characteristics**, but also in the processes of change that occurred over the period of the study. For example, during the years 1972 to 1975, Medicare enrollment as a percent of total population was 47 to 50 percent higher in Kansas/Nebraska than it was in Colorado. A potentially significant difference was that, although the number of Medicare-certified short-term beds per 1,000 enrollees declined in all three states, the decline was steepest in Colorado (approximately 10.14 percent in contrast with 7.85 percent). This could clearly affect utilization, by means of the **Roemer** effect (defined in Chapter II). On the other hand, Colorado had consistently more **Medicare-certified** short-term beds per 1,000 enrollees (but fewer per 1,000 general population). Yet another potentially significant factor was the opening of a major Health Maintenance Organization (Kaiser) in the Denver area in the early 1970s. These differences raise the clear possibility that part or all of the observed differences between the Colorado and **Kansas/Nebraska** trends might be an artifact of other differences between the states that have nothing to do with CAP.

In sum, the evidence on Medicare utilization in the Colorado report is inconclusive. The year-by-year pattern of change casts doubt on the assertion that the relative decline in utilization in Colorado was truly an effect of CAP, and there is a clear possibility that the observed patterns were at least partly caused by other differences between the experimental (Colorado) and comparison (**Kansas/Nebraska**) areas.

CHAMP

The CHAMP program was a PSRO prototype covering the entire Commonwealth of Massachusetts. A recently published program evaluation attributed a substantial reduction in **utilization--**

5.3 percent--to the program. 5/ The CHAMP study is like the CAP study, however, in that a careful look at trends over time throws the reported finding into doubt.

The CHAMP study compared trends in utilization among Medicaid patients (subject to CHAMP review) with trends among non-Medicaid patients (not subject to review). The measure of utilization **was** average length of hospital stay, standardized and expressed as a percent of the length of stay predicted on the basis of all observations over all time periods. The authors found **that:**

The average length of stay of Medicaid patients decreased by 11.9 percent relative to the norm, whereas the non-Medicaid length of stay decreased by only 6.6 percent. We infer that the Program may be credited with the 5.3 percent differential decrease. 6/

The data (which fortunately include two baseline measures) show a disturbing trend, however: even before the startup of CHAMP, the length of stay of Medicaid patients (the treatment group) was declining, while that of non-Medicaid patients (the comparison group) was, if anything, slightly on the rise. The data are presented in Table A-2 and Figure 2. The two measures between July 1972 and June 1973 predated the program. The following year (for which no data are presented) was the transitional year. **Pre-implementation** measures began with July-December 1974. The **post-CHAMP** decline in Medicaid utilization is apparent. The data also show that the major change in Medicaid **utilization--relative to non-Medicaid--occurred** during the pre-CHAMP period and during the transitional period of implementation. During the post-implementation period, the Medicaid and non-Medicaid lines were nearly parallel.

In order to assess the possible effects of the pre-CHAMP trends, Medicaid and non-Medicare length of stay have been projected on the basis of two pre-CHAMP observation periods (see Table A-2 and Figure 3). These projections can be interpreted

5/ Fulchiero and others, "Can the PSROs Be Cost Effective?" pp. 574-80

6/ Ibid.

TABLE A-2. MASSACHUSETTS OBSERVED AND PROJECTED AVERAGE LENGTH OF HOSPITAL STAY,
AS A PERCENT OF PREDICTED LENGTH OF STAY, MEDICAID AND NON-MEDICAID
PATIENTS, DECEMBER 1972-DECEMBER 1976

Date (Endpoint of Six-Month Interval)	Medicaid			Non-Medicaid		
	Observed	Projected	Percent Difference	Observed	Projected	Percent Difference
Pre-CHAMP						
12/72	117.0	--		103.4	--	
6/73	114.9	--	=	103.6	--	--
Implementation Period						
12/73	--	112.8	--	--	103.8	--
6/74	--	110.7	--	--	104.0	--
Post-Implementation						
12/74	106.5	108.6	-1.93	98.3	104.2	-5.66
6/75	106.0	106.5	-0.47	98.2	104.4	-5.94
12/75	103.7	104.4	-0.67	96.5	104.6	-7.74
6/76	100.9	102.3	-1.37	95.3	104.8	-9.06
12/76	101.1	100.2	+0.90	93.9	105.0	-10.57

SOURCE: Adapted from Fulchiero and others, "Can the PSROs Be Cost Effective?"

-- denotes missing data.

Figure 2.

Massachusetts Observed Average Length of Hospital Stay as a Percent of Predicted Length of Stay: Adjusted for Age and Diagnosis; December 1972-December 1976

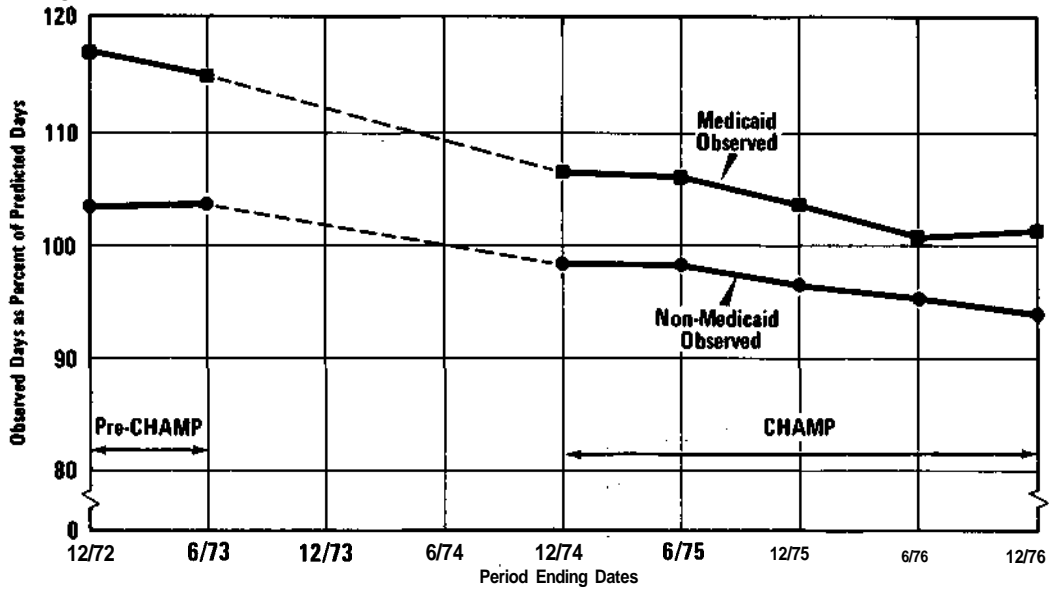
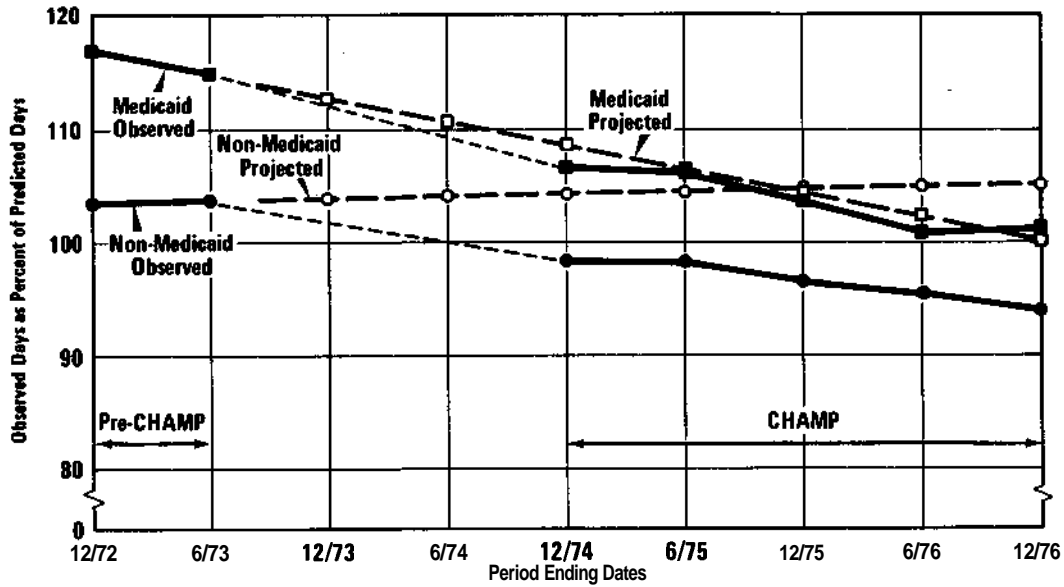


Figure 3.

Massachusetts Observed and Projected Average Length of Hospital Stay as a Percent of Predicted Length of Stay: Adjusted for Age and Diagnosis; December 1972-December 1976



as rough estimates of what utilization might have been if there had been no CHAMP program. One can see that the observed utilization by **Medicaid** patients (the treatment group) is almost exactly what was projected, while observed **non-Medicaid** (comparison group) utilization was substantially lower than projected.

Presented in this fashion, the data suggest a negative effect of CHAMP. As Table A-2 shows, in the last observation period the Medicaid rate was 0.9 percent higher than projected, while the non-Medicaid rate was 10.6 percent lower. This suggests a differential of +11.5 **percent--that is, a relative increase of 11.5 percent** attributable to CHAMP.

The alternative **analysis** presented here is also not without its **weaknesses--particularly** projecting from two data points, which is a risky method. It is at least as risky, however, to ignore the apparent trend suggested by those points. The most prudent conclusion would be that no effect, positive or negative, has been reliably demonstrated. 7/

All in all, the usefulness of the CHAMP study is primarily to reemphasize a point made about the CAP report: when treatment and comparison groups are not essentially equivalent, examining differences in patterns of change in both groups before program implementation is critical. Without such examination, conclusions can be entirely misleading.

7/ This study also suffers from the obvious **noncomparability** of its treatment and comparison groups. Any change in Medicaid **regulations--for** example, a change in eligibility or reimbursement **policies--could** produce changes like those apparent in Figure 2.

This appendix presents the details of the HCFA Medicare rate analysis, some problems with its methods, and the results of a **reanalysis** by CBO of the HCFA data.

THE REGRESSION MODEL

The HCFA analysis of medicare rate data described in Chapter II used forced order regression. The unit of observation was the PSRO area. The dependent variable was 1977 Medicare days of care per 1,000 enrollees. Independent variables were entered in the following stages:

Stage 1: Base utilization rate (1974 Medicare-paid days of care per 1,000 Medicare **enrollees**);

Stage 2: Demographic (control) variables:

- o Proportion of total population age 65 or over (1974 to 1976 change),
- o Short-stay hospital beds per 1,000 population (1974 to 1976 **change**),
- o Nursing home beds per 1,000 population (1973 to 1976 **change**),
- o Population per square mile,
- o Proportion of total hospital days accounted for by Medicare enrollees,
- o Physicians per 1,000 population (1974 to 1976 change),
- o Hospital occupancy rate,
- o Proportion of families with incomes under \$5,000;

Stage 3: "Longevity" (months of PSRO review; zero in in-actives);

Stage 4: Base utilization by longevity interaction;

Stage 5: Longevity by demographic interactions.

This model is problematic in one **important** respect: it offers no direct control of regional differences between the active and inactive areas, even though calculations based on HCFA data showed the regional differences to be highly significant by a chi-square test. Accordingly, the data were reanalyzed with regional dummy variables inserted in a separate stage just before Stage 3. The result was an increase in both the magnitude and the statistical significance of the PSRO effect. 1/

This seemingly paradoxical result, in which "controlling for" a confounding variable increases the estimated effect, is well known in the literature as "traditional" or "classical" suppression. The best estimate of the "pure" effect of PSROs in this case is the higher estimate resulting from the CBO version of the model.

A second reanalysis involved the base measure. The HCFA model that used 1977 total days of care per 1,000 Medicare enrollees (TDOC) as a dependent variable also used TDOC (1974) as a baseline measure. The baseline data revealed an interesting pattern, however: the active and inactive PSRO areas were quite similar in terms of TDOC, but they differed on the components of TDOC: Medicare average length of stay (ALOS) and Medicare discharges per 1,000 enrollees (DISC). The actives were significantly lower in DISC but **higher--although not significantly so--in** ALOS; these differences tended to wash out in TDOC, which is their product. It is therefore possible that using only TDOC as a baseline would be an inadequate control for pre-existing

1/ Thanks are due to HCFA and its consultant, Mitchell Dayton, for computing the additional regressions. Needless to say, HCFA was pleased with the results and included them in its report. See HCFA, PSRO: 1978 Evaluation, pp. 81-84.

utilization differences. ^{2/} Accordingly, the model was re-estimated with **ALOS** and **DISC** added in Stage 1. Regional dummies were included, as in the previous analysis. The addition of **ALOS** and **DISC** made no appreciable difference.

ESTIMATING PERCENT REDUCTION IN UTILIZATION

The estimate of greatest relevance to policy is the reduction of utilization that can be expected nationwide as a result of full implementation of PSRO review. Taking a different approach, HCFA estimated instead the reduction attributable to PSRO review in the areas where the program had been implemented as of April 1977. Although the two estimates can be based on the same regression equation, they require very different computational procedures.

In order to estimate the nationwide reduction in utilization caused by PSRO review, two hypothetical PSROs are created: a typical active and a typical inactive. Both are given the overall (active plus inactive pooled) mean values on all variables (baseline, demographic, and regional) that precede PSRO longevity in the model, thus controlling for the effects of those variables. The typical inactive is assigned a value of zero on longevity, and thereby values of zero on all of the subsequent variables (which are product terms involving longevity). In contrast, the typical active PSRO is assigned the mean longevity of all active PSROs on longevity and the mean of all PSROs on other variables. These two hypothetical PSROs are then plugged into the regression equation to get predicted utilization rates, and the difference between the two estimated values is expressed as a percent of the **inactives'** predicted value. The resulting percent change is the model's best estimate of a "pure" effect of PSROs, controlling for all of the stated pre-existing differences between the active and inactive areas.

^{2/} The TDOC variable, while a product of **ALOS** and **DISC**, is not the interaction between them. Rather, the interaction would be the variance predicted by the product after partialling out the effects of **ALOS** and **DISC**.

The HCFA approach, which they called a "disaggregated" method, was markedly different. For each of the active PSROs only, two values were calculated: a predicted active and a predicted inactive. The active value was predicted by plugging into the equation each **PSRO's** actual values on all variables, including longevity (as opposed to the overall and **within-group** means used in the above method). The inactive value was calculated by using each PSRO's actual values on all variables up to but not including longevity, and assigning zeros thereafter. These two predicted values were used to calculate an estimated percent reduction in each active PSRO; these reductions were then weighted by the size of the enrolled population in each PSRO and pooled across all active PSROs to get a mean percent reduction.

Under ideal **circumstances**, the HCFA approach would have two advantages. First, it would allow estimation of real benefits (that is, the benefits from PSRO review where already implemented) rather than potential benefits (that is, likely benefits when PSRO review is implemented either nationwide or in a random sample of areas). Second, it would allow one to take into account differences between PSROs in hospital costs and in total population when calculating savings-to-cost ratios. If, for example, utilization reductions were greater in areas where daily costs were higher and the population larger, the national estimate of savings calculated by this method would ideally reflect this fact. In practice, however, neither of these potential advantages are realized.

First, so-called "real" savings are not the estimate of principal policy interest. Rather, it is most important to estimate the overall effects of national PSRO **implementation**, as mandated by the statute. Whether estimated savings in those areas that happened to be first in setting up PSRO review is different from the national estimate is a question of far less importance.

Second, the increased accuracy in savings-to-cost calculations that the HCFA method would seem to **offer** is only present if the estimates of the **effectiveness** of individual PSROs are themselves reliable and accurate. **Unfortunately**, they are not. The margin of error in estimating the **effectiveness** of an individual PSRO is certainly many times the size of the margin of error

in estimating the national **effectiveness** of the program. 3/ That is, one can have far more confidence in an estimate that all PSROs have an average effect of decreasing days of care by 2 percent than an estimate that PSRO A saves 2 percent.

The problem of error in the HCFA method, however, is even more severe than these figures would estimate. This is because their method depends on the accuracy, not only of estimates of individual **PSRO's effectiveness**, but also of the estimated differences in **effectiveness** between two PSROs. That is, for their method of estimating to be superior to the "national" method, one would have to be able to place confidence in their calculations showing that one PSRO (perhaps one with high daily costs) is more effective by a given amount than another PSRO (perhaps one with lower daily costs). The formula for the standard error of a difference shows that the margin of error in such a comparison is necessarily even greater than the margin of error in each of the individual PSRO estimates.

Finally, the HCFA method of estimation involves a misinterpretation of their regression model, as a result of which spurious differences between PSROs will be estimated. PSRO review was expressed as "months under review" in the principal model, with inactives having a value of zero. In a supplementary analysis, HCFA divided this longevity variable into two parts: a **dichotomous** active-versus-inactive variable, and a second variable that expressed months under review for the actives. The analysis showed that the dichotomous variable was significant, while the second variable had virtually no additional effect. In other words, the fact of having review did indeed lower utilization, but longer review did not predict greater utilization reduction. **HCFA's** disaggregated analysis, however,

3/ A precise estimate of the margin of error in assessing the **effectiveness** of an individual PSRO is complex to calculate and would vary from one PSRO to another. It would depend on the distance of the **PSRO's** score on each independent variable from the mean of that variable, the multiple correlations between each of the independent variables and all of the others, and the standard partial regression coefficients of each independent variable in each of those multiple correlations.

failed to take this important finding into account. Since that analysis estimated each PSRO's **effectiveness** on the basis of its actual longevity, older PSROs were falsely credited with **larger** utilization reductions. In other words, **HCFA's** disaggregated method forces an apparent finding that longevity increases effectiveness, even though the supplementary analysis described above revealed no such relationship.

In contrast, the method used by CBO uses only two values of longevity; all inactives are assigned zero, and all actives are assigned the mean longevity of the actives. Longevity is therefore treated as the dichotomous variable that the HCFA supplementary analysis showed it to be.

This error in **HCFA's** method has **implications**, not only for comparisons between PSROs, but also for their aggregate estimates of the **program's effectiveness** and savings-to-cost ratio. This is because their aggregate estimates were obtained by adding up all of the disaggregated estimates for all active PSROs. If longevity is correlated with the size of the PSRO's Medicare population, then the aggregate estimate of PSRO-induced reductions in utilization will be biased. Similarly, a correlation between longevity and hospital costs will bias the aggregate estimate of the program's savings.

RANKING PSROs IN EFFECTIVENESS

To rank PSROs in terms of their **effectiveness** would require some sort of disaggregated analysis. The section above discusses two reasons why **HCFA's** disaggregated analysis cannot provide a reliable ranking of this sort: The wide margin of error in the estimates of the **effectiveness** of individual PSROs, and the systematic bias introduced by the handling of the longevity variable. The problem of providing a meaningful and useful ranking, however, goes beyond these two problems.

Assuming that the data allow one to obtain individual PSRO estimates with a reasonable margin of error, the differences in **effectiveness** between PSROs can be broken into several **components**:

1. Differences attributable to variations in program characteristics (organization, management, and so

forth). For example, PSROs which delegate review might be more or less effective than those which do **not**.

2. Differences attributable to characteristics of the **PSROs'** settings. For example, PSROs might be more effective in urban than in rural communities, or **vice-versa**.
3. Differences attributable to "error."

Ideally, components 1 and 2 should be tested by means of variables in the model. That is, one should measure the relevant characteristics of the PSROs and their settings and include those measures in the equation. The third, or "**error**," component includes both measurement error and the effects of variables which are not included in the analysis.

These three components can be combined in different ways, depending on the purpose to which the information is to be put. For example, if one wanted to pick out the PSROs with the most effective management strategies, one would want to isolate component 1. On the other hand, if one wanted to channel funding into those settings where PSROs work **best--independent** of program **management--one** would attempt to isolate component 2.

HCFA's disaggregated analysis ranks PSROs primarily in terms of the second **component--differences** attributable to the **PSROs'** settings. The model used includes only one characteristic of the PSROs themselves: their longevity. Accordingly, all other characteristics of the PSROs were in effect relegated to the error component. The error component was then excluded by the method of calculating each **PSRO's effectiveness**. By elimination, the differences in **effectiveness** estimated by the procedure are attributable to the characteristics of the **PSROs'** settings 4/ and PSRO longevity. Since the effect of longevity was estimated incorrectly by this procedure, the only component of the ranking which is not spurious is that attributable to the **PSROs'** settings.

4/ These enter into the estimate through the **longevity-by-covariate** interactions.

VARIATIONS BETWEEN TYPES OF PSROs IN EFFECTIVENESS

A related, but technically very different issue is whether it is feasible to estimate differences in **effectiveness** between types or groups of PSROs. Questions of this sort involve interactions between PSRO review and other variables. 5/ Such questions take the form: Do PSROs that differ on variable A (region, number of beds per 1,000, or whatever), as a group, differ in the **effectiveness** of review as well? If the interacting variables can be regulated or are relevant to policy for other reasons, such interactions can be of great importance.

The appropriate way to answer questions of this sort is to examine the interaction terms in the equation. A significant interaction would be interpreted as indicating that PSROs differing on the variable in question differed also in the effectiveness of review, after removing the effects of other variables in the equation on which the PSROs differ. That is, it is a "pure" measure of the influence of the interacting variable, holding constant all other variables in the equation.

In contrast, the HCFA report measured differences in PSRO **effectiveness** by a very different method: each individual PSRO was assigned an estimated effect by the disaggregated method described above, and then the average effect of PSROs high on a given variable was compared to those low on that variable. This was the method used to calculate regional differences in PSRO **effects**.

This approach produces misleading results because it leaves the interactive effects of the variable in question confounded with the other variables in the equation. 6/ For example, if **region** were associated with Variable X, then the regional differences calculated this way would reflect the influence of both

5/ The question of whether older PSROs are more effective than new PSROs, while conceptually similar, involves main effects rather than interactions. It is therefore discussed separately above.

6/ Technically, if Variable X is anything other than longevity, the confounding is between the two PSRO-by-covariate interactions and does not involve main effects. When (continued)

region and Variable X. By failing to produce the purest possible measure of the effects of the interacting variable, this disaggregated approach yields biased estimates of differences in the effectiveness of PSROs.

6/ (continued) Variable X is longevity, the confounding involves both the longevity main effect and all **longevity-by-covariate** interactions in the model.

○