# DISCUSSION

Robert E. Schlenker, InterStudy

The three excellent papers presented in this session provide an overview of the type of empirical health policy research now underway. My comments will concentrate not on technical statistical issues, but rather on the problems empirical researchers face in the health policy arena. My comments could therefore be entitled: "If this research is so great, why don't policymakers use it more?"

The presentations of Peter Shaughnessy and Carl Morris illustrate two major categories of empirical health policy research now underway: the first is the type of study which evaluates on-going or past programs; this might be called "autopsy" research, and the evaluation of the Indiana hospital incentive reimbursement program described by Dr. Shaughnessy is an excellent example. The second type of research involves evaluation of an experiment deliberately designed to test a possible new policy or program. The RAND health insurance study described by Dr. Morris exemplifies this type of study. Unfortunately studies of both kinds encounter obstacles from policymakers with which empirical researchers must deal if their results are to be taken seriously. I will give three examples:

First, a common complaint is that such studies --especially those of the experimental variety --take too long. This might be termed the phenomenon of "evaporating relevance." That is, policy decisions follow political timetables and are often made before research results are available; the relevance of the study therefore evaporates before the study is completed. The negative income tax experiments in the income maintenance area are a prime example of this phenomenon, and RAND's health insurance study may suffer a similar fate if national health insurance is enacted in the near future.

A second general problem facing empirical researchers is that policymakers are often uncomfortable with the use of specific measures to indicate the success or failure of a particular program. There is considerable comfort in such vague program objectives as "medical care cost containment" and "improved access to medical care." But difficulties arise when specific measures of such program objectives are discussed. For example, how should the success or failure of a program aimed at containing hospital costs be measured? In terms of costs per inpatient day, costs per specific service, cost per case, or something else? How then should the quality of care be measured, so that the evaluator can determine if the quality of hospital care has been reduced to achieve a given degree of cost containment?

A third problem that empirical researchers often encounter is the argument that key program or policy outcomes are really non-quantitative and cannot be adequately measured. This is, of course, related to the preceding problem that specific measures never completely capture broad program objectives. However, since this argument is often based on faith rather than reason, it is very difficult to counter.

The problems I have just discussed affect both major categories of empirical health policy research (autopsy and experimental). In addition, each category carries some unique vulnerabilities. I'd like to provide a few examples; I'll start with the autopsy approach.

A major problem of this type of research is to identify the specific program being evaluated. Most programs are constantly changed, and it is consequently very difficult to determine the precise nature of the program. This, of course, makes it extremely difficult to generalize from the study results. An example is the federal government's Economic Stabilization Program of 1971-74 as it applied to hospitals. Throughout the period, federal regulations were frequently changed. In fact, I have heard it said that each time the hospitals began to understand a set of regulations, those regulations were changed. The difficulties of identifying program specifics for the evaluation of such a program are legion, and, unfortunately such difficulties are common to most health programs.

The other major problem I'll mention about the autopsy approach is that of identifying the appropriate comparison group. If one seeks to employ a cross-section approach and identify a reasonably similar control group not subjected to the program being evaluated, the argument can usually be made that confounding effects are still present and that untangling the program's impact from the effect of other variables is extremely difficult. In the same vein, if a time series approach is used, then the problem is that other variables also change over time. The longer the program being evaluated has been in effect, the more serious the problem becomes of disentangling the effect of the program being evaluated from the effects of all other changes occurring during the same time period. There are, of course, ways to deal with these comparison problems; however, the solutions do increase the costs of such research and thereby tend to try the patience of those funding the research.

Now, I'd like to mention a few of the problems encountered in the policy arena by the experimental type of research.

First, a health policy experiment is never "pure". Random selection is never completely

achieved, and differences between the treatment group and the control group will always remain, and these differences may be correlated with the variable or variables measuring the policy or program "treatment". In addition, the number of different treatments usually exceeds the number of experimental sites. For example, there is no single health insurance plan or single hospital incentive reimbursement scheme now being tested, but several varities of each. Policymakers generally wish to test a variety of program options, but funds are not available to test each option separately. Therefore, each resulting experiment usually tests a number of program variables at each site, and it is consequently very difficult to determine which of those variables is responsible for any measured differences between treatment and control groups.

A second problem encountered by the experimental approach is that the subjects of the experiment may behave differently in the experiment than they would under a permanent program. I have in mind here more than just the well-known Hawthorne effect. For example, hospitals participating in an incentive reimbursement experiment may try harder to contain costs than they would under a permanent program, since their objective may be to use the experiment to gain the knowledge necessary to later "beat the system" under an expected permanent program. An experiment would then overstate the likely impact of a permanent program. A contrasting possibility is that during the experimental period the participating hospitals may not have sufficient time to understand the program well enough to make the behavioral responses they would under a permanent program. In this case, an experiment would understate the likely effects of a permanent program.

This last point is related to the third problem of experiments: the problem of the long run. It can always be argued that no experiment is ever long enough to measure all the important long run effects of any major new program.

The final problem I'll mention for experimental research in the health policy area results from its expense. Because experiments are expensive, they acquire greater visability and draw more vehement defenders and detractors. This, in turn, increases both the pressures on the researchers to produce premature results and the difficulties of keeping those results objective.

The above are some examples of the problems encountered by empirical researchers in the health policy field. Although the problems are serious, progress is being made toward their resolution. Empirical research in the health care area, both of the autopsy and experimental varieties, will continue to grow, since the alternative of continued ignorance is even less palatable to policymakers. In addition, such research has already had, I believe, two major positive impacts on the policymaking process. First, it has raised the analytical level of the decision process. Second, it has demonstrated the incredible complexity of the health care system, and has thereby slowed down the decision-making process a bit. The resulting pause may be not only refreshing but vital.