THE HEALTH INSURANCE STUDY—A SUMMARY

PREPARED UNDER A GRANT FROM THE OFFICE OF ECONOMIC OPPORTUNITY

JOSEPH P. NEWHOUSE

R-965-I-OEO
MARCH 1974

Rand
SANTA MONICA, CA 90406
The research reported herein was performed pursuant to a grant from the Office of Economic Opportunity, Washington, D. C. The opinions and conclusions expressed herein are solely those of the author, and should not be construed as representing the opinions or policy of any agency of the United States Government.
THE HEALTH INSURANCE STUDY—A SUMMARY

PREPARED UNDER A GRANT FROM THE OFFICE OF ECONOMIC OPPORTUNITY

JOSEPH P. NEWHOUSE

R-965-1-OEO
MARCH 1974

Rand
SANTA MONICA, CA. 90406
This report describes the design of the experimental portion of the Health Insurance Study (HIS) and updates a previous report issued in November 1972. The HIS attempts to determine, among other things, the effects of alternative insurance plans upon the demand for medical services and the health status of the population. Answers to these questions can be sought through gathering and analyzing nonexperimental data or through an experiment. This report discusses the latter option, and has been published in Inquiry, March 1974. Other work in the HIS has analyzed nonexperimental data. This work includes


The HIS began under the auspices of the Policy Research and Evaluation Division of the Office of Economic Opportunity and is now being carried out under a grant from that Division. In mid-1973 administrative responsibility for the project was transferred to the Office of the Assistant Secretary for Planning and Evaluation, Department of Health, Education and Welfare. The HIS is a joint endeavor of The Rand Corporation and Mathematica, Inc., of Princeton, New Jersey. Mathematica has contracted with Rand to perform field operations, process claims, and assist with design and analysis.
SUMMARY

This report is a summary of the design of the experimental portion of the Health Insurance Study. The study has several objectives, including:

1. Estimating the responsiveness of demand for health services to insurance provisions, especially deductibles and coinsurance rates. Health services are defined broadly to include dental, vision and hearing, mental health, and other services frequently not covered by existing insurance.

2. Determining how responsiveness to insurance provisions differs with such factors as family income and severity of illness.

3. Ascertaining how the ambulatory system accommodates to varying levels of demand. This objective includes determining the effect of high demand on delays in making appointments, waiting times in the office, types of cases seen by the physician, revisit rates, and referral patterns.

4. Assessing the effects of various insurance plans on health status.

5. Learning how the quality of medical care is affected, if at all, by various insurance plans. If Professional Standards Review Organizations become prominent during the period of the study, before-and-after data will make possible an assessment of their effect. It should also be possible to compare the quality of care given by graduates of foreign medical schools with that given by graduates of American medical schools.

6. Finding out how much additional insurance a family will purchase if a government plan leaves it liable for considerable out-of-pocket payments, and how the amount purchased is likely to be affected by the tax treatment of health insurance premiums.

7. Examining three innovative types of insurance, including:
   a. Plans in which outpatient care is free but inpatient care is subject to deductibles or coinsurance, providing a positive incentive to obtain preventive care.
b. Health Maintenance Organizations (HMOs).
c. The placing of limits on family out-of-pocket expenditures in all plans where family payments are required. These limits, or maximum liability, will be determined as fractions of family income.

8. Gaining familiarity with the difficulties of administering insurance plans with income-related maximum liabilities. For example, if the maximum liability is defined as a fraction of annual family income, how is it to be changed if, for example, the family unit changes by legal separation at some point during the year?

To accomplish these objectives, some 7500 individuals in 2000 families at four sites (probably three urban areas and one rural area) will be enrolled in varying insurance plans, to include:

1. A plan in which care is free to the consumer.

2. Plans that require coinsurance rates up to 100 percent. In these plans the maximum out-of-pocket payments will be limited to 5 or 15 percent of a family's income. Also, in some of these plans the coinsurance will apply to inpatient services only; outpatient services will be free.

3. A plan that requires a deductible of $150 per person.

4. An HMO plan. To preserve the continuity of care, it is likely that families enrolled in this plan will already be enrolled in an HMO. These families will be chosen so that their demographic characteristics match those of other families in the study.

In addition, around 5000 individuals will be enrolled in a control group. Those individuals will remain on their old insurance plans.

Slightly over 50 percent of the families will participate in the study for five years, and the remainder will participate for three years. During these periods they will be interviewed periodically. The interviews will primarily gather information on health status (as perceived by the family). A medical history will be obtained from all families at the beginning of the study, and a physical screening examination will be given to all families at the end of the study. In addition, about one-half of the families will be given the screening
examination at the beginning of the study. The purpose of the examination is to obtain objective measures of health status; it is limited to half the families at the beginning to measure any potential effects that the examination may have on utilization. Most utilization data will be collected through claims for services filed with the study.

Eligibility for participation is quite broad. Those persons 65 years of age and over are not eligible; in addition, families with incomes greater than $25,000, members of the military, and the institutionalized population are not eligible. Families with lower incomes have a higher likelihood of selection.

To assure that a participating family actually has the insurance plan randomly assigned to it for the study, families are asked to assign the benefits of their present insurance to the study. If participation under these conditions could ever make a family worse off, it will be paid an amount of money sufficient to assure that it will never lose financially from participating. Also, families are urged to keep their present insurance; their share of any premiums paid to do so is reimbursed. Thus families will not become uninsurable by participating in the study and can return to their existing insurance if they wish to withdraw from the study.

Sites will be chosen so that the stress on the ambulatory system varies. For example, some sites will have long delays for an appointment with a primary-care physician; others will have little delay. The first site is Dayton, Ohio. One of the remaining sites will be in the West, and one will be in the South.

A pilot sample of 50 families has been enrolled in Dayton. These families were enrolled in November and December 1973, and their coverage was effective January 1, 1974. The purposes of the pilot sample are primarily to test our ability to enroll families and to test our operating systems. Experience has been encouraging on both points. Fifty-five offers were made to obtain the fifty families, and after two months of operation no families have withdrawn. The concern with our operating systems centers around our ability to gather data through interviews and through the claims system. To date, adequate information is being obtained, although followup of claims is necessary. It is, however, premature to make a definitive evaluation.
Assuming that experience with the pilot sample continues to be favorable, enrollment of approximately 500 families in experimental plans and 300 or more families in a control group will take place at Dayton in September 1974. The second site is to be chosen in the spring of 1974, the third site three months later, and the fourth site by the end of 1974. Enrollment is planned at the second site in June 1975, at the third in August 1975, and at the fourth in November 1975. The experience in Dayton will be reviewed by the Department of Health, Education and Welfare before enrollment is permitted to take place in the subsequent sites.

The annual cost of the fully implemented project is estimated to be $5 to $6 million, and the total cost over the lifetime of the project is projected at $32 million (FY1973 dollars).

This study discusses the design of the experimental portion of the Health Insurance Study (HIS). In the following section the objectives of the experiment are considered in detail and placed in the context of existing knowledge about health care financing. Alternative means of attaining those objectives are also considered. The next section summarizes our methods of procedure. We conclude with brief remarks about the significance of the work. Table 5 at the end of the study summarizes the dimensions of the experiment.
ACKNOWLEDGMENTS

The earlier version of this report contained acknowledgments to fifteen individuals. Since that time the list of people who have contributed ideas to the design of the project has grown to well over a hundred, and any attempt to acknowledge some would certainly do an injustice to others. One individual should nevertheless be singled out. The original impetus to conduct a health insurance experiment came from Larry Orr, now at DHEW, and his ideas have shaped the design of the project in many important ways. I am deeply indebted to him and to my other colleagues who have brought the experiment from the drawing boards into reality.
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>PREFACE</strong></td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td><strong>SUMMARY</strong></td>
<td>v</td>
</tr>
<tr>
<td></td>
<td><strong>ACKNOWLEDGMENTS</strong></td>
<td>ix</td>
</tr>
<tr>
<td>I.</td>
<td><strong>OBJECTIVES</strong></td>
<td>1</td>
</tr>
<tr>
<td>I.</td>
<td>Effect of Price on Demand for Medical Services</td>
<td>1</td>
</tr>
<tr>
<td>I.</td>
<td>Effect on Demand of Changes in Coverage</td>
<td>3</td>
</tr>
<tr>
<td>I.</td>
<td>Effects of Insurance on Health Status</td>
<td>4</td>
</tr>
<tr>
<td>I.</td>
<td>Administrative Aspects of Health Insurance</td>
<td>4</td>
</tr>
<tr>
<td>I.</td>
<td>HMOs and Quality of Care</td>
<td>5</td>
</tr>
<tr>
<td>I.</td>
<td>Effects in Different Communities of Varying Physician Workloads</td>
<td>6</td>
</tr>
<tr>
<td>II.</td>
<td><strong>RESEARCH STRATEGY</strong></td>
<td>8</td>
</tr>
<tr>
<td>II.</td>
<td>Advantages and Disadvantages of Non-Experimental Data</td>
<td>8</td>
</tr>
<tr>
<td>III.</td>
<td><strong>METHODS OF PROCEDURE</strong></td>
<td>11</td>
</tr>
<tr>
<td>III.</td>
<td>Plan of Analysis of Demand for Services</td>
<td>11</td>
</tr>
<tr>
<td>III.</td>
<td>Dimensions of Experimentation</td>
<td>16</td>
</tr>
<tr>
<td>III.</td>
<td>Income Limits, Deductibles, and Coinurance</td>
<td>17</td>
</tr>
<tr>
<td>III.</td>
<td>Role of Control Group</td>
<td>19</td>
</tr>
<tr>
<td>III.</td>
<td>Period of Participation</td>
<td>22</td>
</tr>
<tr>
<td>III.</td>
<td>Statistical Experimental Design</td>
<td>24</td>
</tr>
<tr>
<td>III.</td>
<td>Breadth of Benefits</td>
<td>28</td>
</tr>
<tr>
<td>III.</td>
<td>Selection of Sites</td>
<td>30</td>
</tr>
<tr>
<td>III.</td>
<td>Rules of Operation</td>
<td>31</td>
</tr>
<tr>
<td>III.</td>
<td>Schedule and Costs</td>
<td>36</td>
</tr>
<tr>
<td>IV.</td>
<td><strong>THE SIGNIFICANCE OF THE RESULTS</strong></td>
<td>39</td>
</tr>
<tr>
<td></td>
<td><strong>REFERENCES</strong></td>
<td>41</td>
</tr>
</tbody>
</table>
For which of you, intending to build a tower, sitteth not down first and counteth the cost, whether he have sufficient to finish it? Lest haply, after he hath laid the foundation, and is not able to finish it, all that behold it begin to mock him, saying, this man began to build, and was not able to finish.—Luke 14:28, 29
1. OBJECTIVES

EFFECT OF PRICE ON DEMAND FOR MEDICAL SERVICES

A primary objective of the experiment is to measure how the demand for medical services varies with the price the consumer pays for medical services. We focus on price as a key variable for three reasons:

1. Since insurance operates to change the price, and since public policy can legislate the terms of insurance, public policy can exert considerable influence over price. Indeed, the ability of government to influence the terms of insurance is probably its most powerful available instrument for affecting the demand for medical services.

2. There is a well-established economic theory on the effect of price. This theory gives us some confidence in our ability to understand the phenomena we are measuring.

3. Price may be a common denominator for measuring the effects of many variations in the details of insurance policies. For example, deductibles may be per person or per family, per quarter or per year. To attempt to ascertain the effect of variation in detailed clauses of this kind—by observing actual variation with all else held constant—would require vast resources. One may be able, however, to approximate a particular variation as a variation in price; knowing something about the effect of a variation in price, one may then be able to estimate the effect of the particular provision.

We therefore assign participants randomly to insurance plans that systematically vary the price they face.

Measurement of the effect of price on demand can be refined in several ways. For example, we may wish to know whether the effect of price varies with income and total expenditure. Do poor families respond more readily to price than affluent ones? If so, they will

---

1 See Grossman (1972); Phelps and Newhouse (1972).

2 Another example of a policy detail is limits on specific services; some hospital policies may reimburse for a semi-private room; other may have a specific dollar maximum. Some may have a dollar limit for ancillary services; others may not have such a limit; still other policies may not cover such services at all.
benefit differentially from a generous health insurance plan. Do families with major illnesses respond differently from those with minor illnesses? If so, estimates of the effect of catastrophic health insurance must be based on data from only those with such illnesses.

The effect of price cannot be measured straightforwardly if the price changes with total expenditure. This would be the case if, as has been proposed, the federal government were to provide a health insurance plan that paid the entire bill after the family had spent a certain amount in a given year. The problem this poses for analysis may be grasped intuitively by noting that if a person expected to spend less than the amount at which the plan takes over, he would act differently than if he expected to spend more than that amount. The "price" he faces is therefore not obvious. This problem is discussed in greater detail below.

The effect of the price of particular medical services on the demand for other medical services should also be examined. For example, it is often alleged that the present relatively generous coverage of inpatient services shifts utilization from the office to the hospital. The HIS therefore covered all services uniformly; by comparison with a control group (many of whom will have the present kind of coverage) it may be possible to determine the effect of the scanty coverage of ambulatory services on the use of hospital services.

Another consequence of the scanty coverage of ambulatory services, it is often alleged, is a neglect of preventive medicine, leading to increased utilization later. To provide maximum incentive for the use of preventive services, a set of plans is included in which all outpatient care is free but inpatient care is subject to coinsurance. This will permit inferences about the effect of financial incentives in inducing the consumption of preventive services.²

²The limited duration of the experiment (five years) creates obvious difficulties for estimating the magnitude of the effect of any preventive services on the later consumption of medical services, but any immediate effect should be noticeable.
The HIS will also seek to measure the effect of price on demand for services other than typical hospital and physician services. It is generally agreed that hospital and most physician services should be included as part of a health insurance plan, but there is less general agreement on the inclusion of prescription drugs and vision and hearing, chiropractic, dental, and psychiatric and psychological services. Because little is known at present regarding the effect on demand of covering such services, most of them have been included in the scope of coverage.

EFFECT ON DEMAND OF CHANGES IN COVERAGE

While a major objective of the experiment is to improve estimates of the responsiveness of demand to price, we are ultimately interested in the demand for medical services under any particular insurance plan. This will be determined not only by the responsiveness of demand but also by the change in the amount of coverage that any particular plan will bring about. To compute this change, one must know the current coverage of the population; unfortunately, detailed information on this subject is rare. National Health Survey estimates can be used to determine the percentage of the population with no coverage, but they are not helpful if one wants to determine the improvement that would be caused by any particular plan—say, a full-coverage plan—among persons who now have partial coverage (as most do). The experiment will augment and update the existing scarce data on this subject (primarily the 1970 survey of the Center for Health Administration Studies).

The change in coverage any legislation causes will also be partly determined by the degree to which consumers will purchase additional insurance to supplement a government policy, if the government plan stops short of full coverage (as, for example, it did with Medicare). Supplementation is an especially important issue in plans with large deductibles, which could leave the consumer bearing considerable risk. The HIS will offer the participant the chance to supplement his experimental insurance in the final year of the experiment, and so will be able to generate evidence on the degree to which consumers would purchase additional insurance if a national health insurance plan were passed that did not provide full coverage.
How much additional insurance the consumer decides to purchase, if any, may well be a function of the tax treatment of insurance. Currently, the tax system provides a considerable subsidy to the purchase of insurance, since employer-paid premiums are not taxable income, and individually paid premiums are 50 percent deductible (up to $150). This subsidy exists to stimulate the purchase of private insurance, but if a national plan were available the rationale for the subsidy could disappear. As a result, some have proposed changing the tax treatment of insurance.\(^1\) By altering the terms at which supplementary insurance is offered, the HIS can determine the effect of various tax treatments of insurance premiums.

**EFFECTS OF INSURANCE ON HEALTH STATUS**

Estimating the response of demand to insurance is necessary (though not necessarily sufficient) for estimating the costs of alternative financing legislation, or the resources that the nation would devote to health care. To determine the benefits of alternative legislation, one would like to know something about the effect of a change in utilization on the health status of individuals. This issue is often described as "necessary" rather than "unnecessary" care, and the problem is to define how much additional health care of each kind is induced by more generous insurance. Improving the state of knowledge regarding this issue is a principal objective of the experiment, but it will not be discussed further since it has been considered elsewhere.\(^2\)

**ADMINISTRATIVE ASPECTS OF HEALTH INSURANCE**

Another objective of the experiment is to contribute information on the administrative feasibility of various aspects of national health insurance. Obviously, the scale of the experiment will not permit replication of how a large organization, private or public, might administer a national plan, but it will permit some testing of various

---

\(^1\) Feldstein and Allison (1972); Mitchell and Vogel (1973).

\(^2\) Kisch and Torrens (1974).
technical details. For example, the accounting period and the accounting unit must be determined for those plans that use deductibles or coinsurance that is eliminated after a certain amount of expenditure. Although the HIS will use a uniform set of definitions, we hope that the data we collect can be used to simulate the consequences of alternative definitions. We also expect to contribute information on details of claims processing. If, for example, a large deductible applies to the family, what claims volume might be expected? What information can be obtained from claims forms that will permit ongoing evaluation of the insurance plan? Although there is now some information of this kind among fiscal intermediaries, it is generally relevant to other types of plans than those we are considering in the experiment.

HMOs AND QUALITY OF CARE

In recent years HMOs have been advocated as an important reform for the delivery system. Although they encompass a wide variety of arrangements, a distinguishing feature is the prepayment of the physicians for services. Thus, the method of reimbursement of the physician differs from the fee-for-service system. There is a growing volume of evidence that this change in the method of paying the physician considerably reduces the amount of utilization. However, there is very little evidence on the effect of such a reduction on health status. We therefore plan to assign certain of our enrollees to such organizations to measure the effects of HMOs on both utilization and health status.

Interest in the quality of care is related to the debate over HMOs. It is feared that if physicians are given a fixed amount of money to care for their patients, they will be motivated to stint on services, thereby jeopardizing the quality of care. Considerable emphasis is

1The accounting period defines whether the deductible is annual, quarterly, or tied to some other period, whether the period is a moving average, and so forth. The accounting unit defines whether the deductible applies to the family or to the individual, and, if it applies to the family, how the family unit is defined.

2Donabedian (1969).
therefore being placed on the development of measures of quality. Since one of the objectives of the experiment is to assess the effect of pre-payment, it will be necessary to address the issue of quality-of-care measurement within the experiment. It is hoped that this effort will contribute to the growing amount of methodological work on the issue.

The principal instrument currently envisioned for improving the quality of care is some kind of peer review. There are those who doubt that peer review can be effective. By choosing sites where the extent of peer review varies, we hope to generate some evidence on relative effects.

EFFECTS IN DIFFERENT COMMUNITIES OF VARYING PHYSICIAN WORKLOADS

A final objective relates to measuring the effects of the varying workloads of physicians in different communities. The types of services delivered by physicians when they are busy may differ from the types delivered when they are not. For example, waiting time to an appointment, waiting time at the office, time spent per patient, revisit rate, and profile of medical problems seen may all change as the physician becomes more or less busy.\(^1\) This in turn may affect who gets what kind of medical services for what kinds of medical problems. A number of financing proposals would increase the demand for ambulatory physician services significantly, so that understanding the effect of a physician's being more or less busy is an important issue in the debate over financing; it is also important in formulating health manpower policy.

The experiment can contribute to understanding these issues, since physicians are considerably busier in some parts of the United States than in others. For example, the best endowed of the four Census regions has around half again as many physicians per person as the worst endowed. Yet the best endowed region has only 10 to 20 percent more physician visits.\(^2\) This suggests that where there are few physicians per person those physicians are likely to be 30 percent or more busier than

\(^1\) Enterline et al. (1973).

\(^2\) National Center for Health Statistics (1968); American Medical Association (1966).
their colleagues in well-endowed areas. By selecting sites with varying degrees of physician workload, we should be able to measure the effects of physician workload.
II. RESEARCH STRATEGY

ADVANTAGES AND DISADVANTAGES OF NON-EXPERIMENTAL DATA

The use of existing non-experimental data has certain advantages over conducting an experiment. In this section we consider whether an experiment is a sensible research strategy.

The principal advantages of existing non-experimental data are that the analysis is faster and that it is cheaper than with experimental data. Both are significant. If it could promise the same kind of results, the analysis of existing data would be preferred. In addition, both existing and prospective non-experimental data avoid two difficulties found with experimental data, the Hawthorne Effect and the transitory problem. According to the Hawthorne Effect, individuals behave differently when they are being observed (in an experiment) than when they are not being observed. The transitory problem is analogous to the Hawthorne Effect; individual behavior in an experiment of limited duration differs from behavior in a national program that the participant expects to last indefinitely. However, as discussed in the next section, these two problems have solutions; so the important issues are speed and cost relative to the quality of results.

To achieve the objectives discussed above, both experimental and non-experimental research should be pursued. But existing non-experimental data—from surveys and claim files—simply cannot provide answers to many of the questions posed above. Where the data can provide some answers, the precision is likely to be low, because non-experimental data have a variety of weaknesses:

1. They generally provide little or no information on the health status of individuals. As a result, it becomes impossible to measure the consequences of more or less utilization. It is reasonable to believe that more rapid progress will be possible on this issue with a project of the scope described here.

2. Survey data rely on the ability of individuals to recall details of their utilization from memory. There are obvious possibilities for error here, the more so if utilization farther in the past
must be recalled.\(^1\) (The National Health Survey estimates physician visit rates using data from the past two weeks because of the magnitude of recall error.

Administrators of survey data either rely upon the respondent for details of his insurance policy or attempt to obtain the policy from an independent source. Such data are essential if the effect of changing the terms of the policy is to be measured. Yet so few respondents know the details of their policy that the National Health Survey has stopped asking such questions. I know of only two national probability sample surveys that have tried to obtain the policy from an independent source: the 1963 and 1970 surveys of the Center for Health Administration Studies. Two surveys represent only a small amount of data.

Existing data are by definition limited to the range of insurance policies currently in effect. Such policies seldom have deductibles greater than $100 per person and coinsurance rates greater than 25 percent. The effect on demand of a plan containing a large income-related deductible or coinsurance rate is therefore very hard to assess. Also, as pointed out above, existing policies may vary along numerous dimensions. One policy may pay up to $60 per day for a hospital room, another may pay the full cost of a semi-private room, and still another may pay 80 percent of whatever the room charge is. Attempting to measure the effects of such policies is difficult. Finally, existing policies typically do not cover some medical services, such as psychiatric services; it is therefore difficult to learn much about the consequences of such coverage.

3. Data from the claim files of existing insurance plans are, of course, restricted to existing kinds of plans and therefore have the weaknesses just described. In addition, claim files usually contain little detailed information on the patient beyond age and sex. Without data on income, such data cannot be used to answer questions relating to the effect of insurance on different economic groups. In addition, if there is a deductible or upper limit in the policy, there will be no information on services consumed below the deductible or

\(^{1}\)Balamuth (1965).
above the limit. Thus, total demand may not be well measured. Finally, the necessary data in many claim files are not in machine-readable form. To reduce the data to this form is an expensive and time-consuming operation, tending to defeat the advantage of non-experimental data.

4. Consumers have by and large self-selected their insurance. Since there is evidence that those in poorer health have better insurance, 1 use of existing data may overstate the responsiveness of demand, if insurance were extended to the entire population. While the overstatement can be corrected with enough data and appropriate estimation techniques, the cost of obtaining accurate estimates from non-experimental data is raised.

Another alternative to an experimental strategy would be to gather new, richer data from either insurance claims or survey data. To adopt such a strategy, however, is to give up most of the speed and cost advantages of non-experimental data. In addition, certain inherent weaknesses of existing data, as noted, keep the value of this strategy quite low. The weaknesses were apparent to the Social Security Administration when it responded to a 1971 Congressional request to analyze the cost of various national health insurance plans. In providing estimates, it commented: "The [percentage increases in demand] were chosen after a review of past experience, but no claim is made that any of [them] are based on solid empirical foundations." 2

1Phelps (1973).
2Social Security Administration (1971).
III. METHODS OF PROCEDURE

PLAN OF ANALYSIS OF DEMAND FOR SERVICES

We plan to estimate three different models of demand for medical services. The simplest, the easiest to estimate, and the least informative is an analysis of a covariance model, which estimates the mean annual cost of each plan, adjusting for measurable differences in demographic characteristics among individuals assigned to different plans. The model to be estimated is

\[ Y_i = \sum_k \delta_k D_{ik} + \sum_j \beta_j Z_{ij} + \epsilon_i \]  (1)

where \( Y_i \) is a general response variable (such as annual expenditure on medical services of the \( i \)th person); \( D_{ik} \) is a set of dummy variables, which take the value unity if the \( i \)th person is enrolled in the \( k \)th plan and the value zero otherwise; and \( Z_{ij} \) is a vector of demographic characteristics for the \( i \)th person (such as age or income). If no \( Z_{ij} \) are included, the estimated \( \delta_k \) are simply the mean level of expenditures in each plan.

Demographic variables such as age and income are probably best entered in dummy-variable interval form to allow for non-linearities. To predict the cost of any particular plan when such variables are included, one must obtain the predicted cost for each demographic group and then weight such costs by the proportion of the total population that the demographic group represents.

The model specified in Eq. (1) does not allow for interactions between the demographic variables and the insurance plans. Tests will be performed for the existence of interactions of interest (such as whether the effect of a particular plan differs by income group). If such interactions prove to be significant, they will be included.

There is no doubt that such a model can be estimated and that the results will be of interest. The results of estimating Eq. (1), however, are not as useful as the results of estimating two other kinds of models, for two reasons. First, no theory underlies the model specified in
Eq. (1), which would afford an insight into the effect of health insurance on demand. And, second, the model specified in Eq. (1) estimates discrete points on a response surface. It provides no basis for interpolating between points and cannot readily be used to predict what would happen if the underlying plans were changed.

The second model is a traditional model of demand based on economic theory. An example might be of the form:

$$Y_i = \alpha + \sum_j \beta_j Z_{ij} + \delta_1 P_{iI} + \delta_2 P_{i0} + \epsilon_i$$

(2)

where variables are defined as before except that $Y_i$ is expenditure for a specific service rather than total expenditure, $P_{iI}$ is the price facing the ith individual for inpatient services, and $P_{i0}$ is the price for outpatient services. $P_{iI}$ and $P_{i0}$ are functions of the insurance plan. The estimated values of $\delta_1$ and $\delta_2$ can then be used to derive own-price and cross-price elasticities of demand for that service. Again, no interactions have been specified in Eq. (2), but we will test for their presence. Specification of a linear form is not meant to be restrictive; logarithmic, polynomial, and other non-linear functions of $P$ can readily be estimated.

$Y_i$ must be disaggregated by service to prevent bias in $\delta$. Further disaggregation of expenditure into variation in quantity (physical units) and price (or style) may also be informative.

As with the first model, there is little doubt that this model can be estimated, since a similar one has been estimated using existing survey data. The model specified in Eq. (2) is not much more difficult to estimate than that in Eq. (1), and meets the two objections to that model. Since there is a well-articulated economic theory of the effect of price on demand, some explanation of the underlying phenomena is possible. Moreover, the model is embedded in a functional form, so that interpolation to other prices is possible.

---

1 Newhouse and Phelps (forthcoming a).
2 Newhouse and Phelps (forthcoming b).
3 Ibid.
The estimated coefficients of the Z vector (β) in Eq. (2) will also be of interest. The Z vector will include a dummy variable that takes the value unity for those assigned to a prepaid (or HMO) plan, and the P variable for such individuals takes the value zero. The estimated coefficient of the dummy variable will then show the effects of prepayment, conditional upon zero price. We do not expect to assign individuals to prepaid plans that contain fees; if there is any interaction between price and prepayment, our design will not measure it.

The Z vector will also include estimates of community demand and supply and their interaction with price. It is hoped that such estimates will show how services are rationed as community demand increases relative to supply and therefore will indicate what might happen if national legislation were to change either element. Obviously, our estimates will be reliable only over the range of variation in community demand and supply observed in the sample. By deliberately selecting sites with extreme values and by taking account of seasonal changes in demand, we hope to observe considerable variation; but national legislation representing particularly marked change from the status quo may well raise community demand beyond any level we will be able to observe. In this case, prediction on the basis of our data will be hazardous.

Variables measuring the individual's price of time and its effect on demand will also be included in the Z vector. National legislation that reduces the money price and thereby increases demand may well increase queueing and thereby the amount of time required for medical services; this in turn would decrease the demand for services. Demand could be expected to decrease most among persons who set high prices on their time. The sensitivity of particular groups to changes in the time price could have a considerable effect on how a given supply of services is distributed in the population.  

1Although the experiment can measure the responsiveness of demand to changes in the time price, the amount by which any plan will affect the time price must also be estimated if the information is to be useful. This amount can be assessed to a limited degree in the study by observing the time price across communities with varying amounts of insurance coverage and varying supply. The experiment is not well suited, however, to measuring what would happen to time price if large
Measures of health status will also be included as part of the Z vector. Such measures have been found to be among the most important in explaining variation in demand in cross-sectional studies. For the purpose of estimating Eq. (2), a general measure of health stock seems appropriate. This could be subjective ("Would you rate your health as excellent, good, fair, or poor?") as well as objective (derived from physical examination of the individual). Inclusion of such variables will make the measurement of price effects more precise.

While the model specified in Eq. (2) represents an improvement over that in Eq. (1), two broad difficulties remain concerning the definition of price. The first arises most prominently with hospital services, where the relevant price variable for using the service at all may differ from the relevant price variable for decisions on how much of the service to use. The decision on whether to use hospital services at all has an all-or-nothing character about it, and it should therefore be estimated with a dichotomous dependent variable (with an appropriate estimation technique such as logit or probit) and a price variable that measures the price facing the individual for his total stay in the hospital. Decisions on how much of the service to use (once having decided to use it) may be more appropriately taken to be a function of the cost of remaining an extra day in the hospital.

These two prices are not necessarily proportional, since out-of-pocket costs become zero after a certain level of expenditure. As a result, hospital expenditure variation should be disaggregated into variation along several dimensions, each of which should be analyzed separately: (1) admissions, with total price as an explanatory variable; (2) length of stay, conditional on admission, with the cost of an additional day (marginal price) as an explanatory variable; (3) room and board price per day, with the coinsurance rate for an additional day as an explanatory variable; and (4) ancillary services expenditure per

---

1Anderson and Lee (1970); Grossman (1972); Newhouse and Phelps (forthcoming b).
day, again using the coinsurance rate as an explanatory variable.\(^1\) A similar disaggregation into quantity and price can be carried out for many different services (for example, physician visits, drugs). However, the disaggregation of quantity into separate decisions on use or no use, and then on quantity conditional on use, is probably not necessary for services other than hospital services, since the total price and the marginal price are probably approximately proportional. In sum, this difficulty can largely be handled within the context of Eq. (2).

The second difficulty is more fundamental. The theory underlying the model given in Eq. (2) is an application of standard economic theory.\(^2\) Unfortunately, standard theory assumes price per unit to be a constant; but that is known not to hold in the HIS. Since plans for the HIS limit out-of-pocket expenditures to a fraction of income (because of policy interest in this type of plan), the price falls to zero after a certain amount of expenditure, and the assumption of constancy does not hold. There is little or no prior work in modeling demand in this type of situation. Heuristically, the problem occurs because the consumer may take account of the expenditure limit in deciding how many services to consume, and therefore he may act as if he faces a price other than the one he pays for any particular service.

Because of this problem, we are working on the development of a third model that differs from the first two by analyzing expenditure by episode. It views the consumer as incurring an illness episode, then making a decision on whether to seek care. (Chronic illness is treated as an episode that continues throughout the accounting period.) If the consumer does seek care, he or his physician then makes a decision on how much care by taking into account the probability that the consumer will exceed the expenditure limit at some time during the accounting period—and the value of the services consumed if he does so. The probability that the consumer will exceed the expenditure limit will depend on how far he is from the limit, the length of time

\(^{1}\) Newhouse and Phelps (forthcoming b). The product of the first two dimensions represents the rate of patient days in the population; the sum of the third and fourth dimensions represents price per day. The product of the patient-day rate and price per day represents expenditures.

\(^{2}\) Ibid., Appendix A.
remaining in the accounting period, his underlying health status, the severity of his current problem, and the probability distribution of expenditures for others in his family unit. The operational implication of this model is that expenditures must be classified by episode, and the data collection system is set up to do so. This model is described elsewhere in more detail.1

DIMENSIONS OF EXPERIMENTATION

Given that we wish to estimate the above three models, we have structured the sixteen insurance plans shown in Table 1. These plans are of three basic types. The first, represented by Plans 1 through 8, consists of plans in which the fraction of the bill the consumer must pay (or coinsurance) varies between zero in Plan 1 and 100 percent in Plans 6, 7, and 8. In Plans 2 through 7 the maximum amount that the family can spend on medical care in a year is limited to 5 or 15 percent of its income; in Plan 8, the limit is $150 per person or $450 per family. The variation in utilization and health status observed across these plans will be the primary means of inferring the general effect of price.

The second type of plan (Plans 9 through 15) differs from the first in that all outpatient care is free, but inpatient care is subject to coinsurance or deductibles. Comparison of individuals enrolled in these plans with those enrolled in Plans 2 through 8 will permit statements about the possibilities for and consequences of shifting the place of care from an inpatient to an outpatient basis by using insurance. In addition, comparison of individuals in Plans 9 through 15 and Plan 1 with individuals in plans with positive coinsurance rates for all services (especially Plans 6, 7, and 8) should provide evidence on whether insurance affects demand for preventive services and, if it does, whether preventive services affect health status.

The third type of plan is Plan 16, the HMO plan. Ideally, a variety of HMOs would be included in the study, ranging from well-established prepaid group practices, such as Kaiser-Permanente, to medical foundations and new HMOs. Moreover, difficulties in finding sites with HMOs

1Newhouse (forthcoming).
Table 1
COINSURANCE AND MAXIMUM DOLLAR EXPENDITURE VARIATION IN EXPERIMENTAL INSURANCE PLANS

<table>
<thead>
<tr>
<th>Insurance Plan</th>
<th>Inpatient Coinsurance Percentage</th>
<th>Outpatient Coinsurance Percentage</th>
<th>Maximum Dollar Expenditure (percentage of income)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>25</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>50</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>50</td>
<td>15</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>100</td>
<td>15</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>100</td>
<td>D</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>11</td>
<td>50</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>12</td>
<td>50</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>13</td>
<td>100</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>100</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>100</td>
<td>0</td>
<td>D</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>HMO</td>
</tr>
</tbody>
</table>

NA = Not applicable.
D = $150 per person deductible, $450 per family maximum.
HMO = Health Maintenance Organization.

and establishing data collection systems within HMOs to provide data comparable to those of the fee-for-service system will probably force limitation to one or two HMOs. Because no site for the experiment that contains an HMO has yet been selected, it is impossible to be more specific about procedures in this area at present.

In the remainder of this section I discuss, first, our choice of particular fraction-of-income limits, deductibles, and coinsurance rates, then the role of the control group, the period of participation, and the question of supplementary insurance.

INCOME LIMITS, DEDUCTIBLES, AND COINSURANCE

Speaking formally, our choice of fraction-of-income limits of 5 and 15 percent and coinsurance rates of 0, 25, 50, and 100 percent means
that we have constrained the design so that no families can be allocated to any other design point. Some such constraint is necessary for computational reasons. Our first task, therefore, was to decide how many design points we wished to have—that is, how many points we wished not to constrain to be zero. The answer to this question depends upon the functional relationship between the experimental "treatment" (in this case, coinsurance rates and expenditure limits) and the response (in this case, demand or health status). If, for example, the relationship were linear, maximum efficiency would be achieved by allocating all of the sample to the two extremes of the range of interest. If some of the sample were to be allocated to intermediate points, efficiency would be lost. If, by contrast, the relationship were non-linear, an optimal design would include some intermediate points, the exact number depending upon the non-linearity. If the functional relationship was in fact non-linear but the sample was allocated wholly to extreme points, the non-linearity could not be measured. In either event, however, extremes of the range one is interested in would be included. Thus, plans calling for zero coinsurance and for 100 percent coinsurance were included. There is no reliable evidence as to what the functional form actually is. Because it was felt that the zero to 25 percent range might be of greater interest than higher ranges of coinsurance, a plan with 25 percent coinsurance was also included. Finally, a plan with 50 percent coinsurance was included to give some notion about possible non-linearities in the zero to 100 percent range.

Only two fraction-of-income limits were included, to permit relatively good estimates of the effect of any given plan. To have included a larger number of limits would have greatly increased the number of plans and would have had the practical effect of making the estimation of the model specified in Eq. (1) meaningless, since precision in any particular plan would become low. It was felt that the speed and ease of estimating Eq. (1), together with its heuristic value, meant that an effort should be made to keep the number of plans relatively small.¹

Five percent of income was chosen as the lower limit, because it was thought unlikely that a national plan would have a substantially lower

¹Ibid.
value, unless all care were free (Plan 1). Fifteen percent of income was chosen as the upper limit, because (1) there was some evidence that few families would reach the limit (especially in plans with 25 and 50 percent coinsurance), and so the variation in price would be preserved; (2) it seemed unlikely that a national plan would have a significantly higher expenditure limit; and (3) inclusion of a higher limit would be more expensive, because of the experimental reimbursement method described in a later section. Both limits have been modified to be the lesser of the appropriate percent of income and an absolute dollar amount. The effect of this "truncation" is to eliminate price variation in high expenditure ranges. There would be few observations in these ranges, and the cost of each observation would be large if the truncation were not included.¹

Plans with a $150-per-person deductible were included because it was thought that such a deductible could possibly be included in a national plan. The effect of expenditure limits related to income is to introduce a rather large family deductible, and it would be difficult to estimate the effects of a per-person deductible if it were not included explicitly. The deductible was set at $150 because it seemed unlikely that a national plan with deductible unrelated to income would have a significantly higher figure. A $450 out-of-pocket limit on family expenditures was included to save costs (given the nature of payments to families, as described below) and because such a limit could well be part of a national plan.

ROLE OF CONTROL GROUP

The experimental design includes a control group whose purposes are five-fold. The first is to provide a basis for measuring the Hawthorne Effect, the effect of observation of the families on their utilization. The method for doing this is to vary the frequency of interviewing in the control group. If greater involvement with the families has any effect on utilization, there will be differences in

¹Ibid. Initially, this "truncation factor" has been set at $1000. Work is now in progress to determine an optimal truncation factor by plan, being higher in the high-coinsurance plans.
behavior among families who are interviewed with varying frequency. If the null hypothesis of no difference cannot be rejected (and the power of the test is reasonable, as it should be), one can be reasonably sure that the Hawthorne Effect is not large.\footnote{If the null hypothesis is rejected, the next step is to determine whether there is any interaction between frequency of interviewing and generosity of insurance. To do this, the control group will be stratified in such a way as to guarantee variation of insurance coverage within classes of frequency of interviewing. If the null hypothesis is not rejected (that is, if there is no significant interaction), then the effect of interviewing would only be to change the intercept in our various models. If the null hypothesis is rejected, estimates will have to be made of the interaction (from control-group data) and appropriate adjustments of price elasticities made.}

The second purpose of the control group is to provide a basis for measuring transitory demand in the experimental group. Transitory demand occurs when consumers "take advantage of a sale"—that is, purchase at a different rate from their rate in the steady state. (In the high-deductible plans there can be negative transitory demand.) The problem occurs with services that produce a steady stream of benefits in the future and whose consumption is affected by the money price. Treatment of a self-limiting illness, for example, would not exhibit transitory demand, since the treatment does not produce benefits beyond the period of the experiment. Similarly, if the consumer would not purchase eyeglasses at any price, there is no transitory demand. But suppose that his purchase of eyeglasses is affected by their price, and that if they are not covered by insurance he would replace eyeglasses only infrequently. Then if he is assigned to a generous plan, his rate of purchase would be expected to increase during the period of the experiment relative to what it would otherwise have been. But because the eyeglasses will continue to produce benefits after the experiment ends, his rate of purchase may exceed what it would have been in the steady state with generous insurance. The problem of transitory demand seems important for benefits relating to vision and hearing, dentistry, psychiatry, and elective surgery.

One method for assessing the degree to which transitory demand leads to an overstatement of price elasticities is to include a control group
whose members continue coverage under their existing insurance.\textsuperscript{1} To the extent that there is variation within existing insurance for these kinds of services, one has some assistance in estimating the amount of overstatement of the elasticities. Such estimates will not be very precise in view of the difficulties described above of measuring the effects of existing insurance, but they should provide some information.

Third, the control group will help in measuring the effect of specific exclusions in existing policies, a type of price variation not included among the experimental "treatments" because of the need to minimize the number of treatments. The effects of two exclusions common to existing policies are of interest. Psychiatric services are frequently not covered. It is sometimes alleged that this exclusion has the effect of raising the demand for other medical services. The implication is that the additional cost from covering psychiatric services is smaller than it might seem, because fewer services would be demanded from other physicians if they were covered. This hypothesis can be tested at sites where psychiatric coverage is quite rare; then the use of medical services in the control group (who by assumption do not have psychiatric coverage) can be compared with the use in the experimental group. Rarity of psychiatric coverage at a site is necessary to guard against adverse selection in the group without psychiatric coverage. (That is, those with psychiatric problems may seek employment where there is coverage, leaving a non-representative group without coverage.)

Preventive services (physical examinations, inoculations, and so forth) are also frequently not covered. It is sometimes alleged that this exclusion prevents early discovery of pathology and leads ultimately to poorer health and higher costs. Since the experimental "treatments" cover all services of physicians, comparison with the control group can establish the effect of covering only curative services rather than both curative and preventive services.\textsuperscript{2}

\textsuperscript{1}Since the present coverage is not randomly distributed, it will be necessary to treat existing insurance as endogenous when analyzing data from the control group.

\textsuperscript{2}If almost all members of the control group do not have coverage for either psychiatric or preventive services, comparisons of total
Fourth, the control group will help measure the effect of covering inpatient services but not outpatient services. The experimental "treatments" include equal coverage of both services, and favorable treatment of outpatient services. We rely on the control group to help measure favorable treatment of inpatient services, since that kind of coverage is conventional in the population.

Fifth, unlike the administrators of a national plan, we will not have detailed fee profiles on all providers and therefore may be charged higher prices than would be charged under a national plan. Efforts will be made to eliminate this possibility, but it is unreasonable to expect that they can be entirely successful. Moreover, the incidence of abnormally high prices is likely to be greater in the more generous plans—implying that the estimated price elasticities are overstated. The control group can provide some evidence on the degree to which this has taken place. Since the claims filed by the control group are paid by existing intermediaries, comparisons of prices for specific procedures charged to the control group with those charged to the experimental group (controlling for generosity of insurance) will throw light on any differences in charges between our methods and those of existing intermediaries.

**PERIOD OF PARTICIPATION**

Those persons enrolled in the experimental group will participate for either three or five years. The purpose of varying the participation time is to acquire better evidence than the control group is expected to provide on the extent of transitory demand at the end of the experiment. By comparing the behavior in the third year of those enrolled for three years with that of those enrolled for five years, the degree to which individuals "crowded in" or postponed services in the terminal year should be apparent.

Utilization between the experimental and control groups will not, strictly speaking, help in identifying which exclusion is responsible for the results. In that case, comparisons must be made for those services thought likely to be affected by one exclusion or the other.
Three years is the least period of time that can provide what we regard as a reasonable amount of steady-state data. We expect transitory demand at both the beginning and the end of the experiment; based on available data, we estimate that this demand will occur primarily in the first and last six months. Therefore, we will have about two years of steady-state data from the three-year group. Five years was chosen as the period of participation for the other group for two reasons. On one hand, it was sufficiently long to provide a good baseline against which to measure transitory demand in the three-year group. On the other hand, to have chosen a longer period would have reduced the number of individuals participating (given a fixed budget), and therefore the amount of information at any point during the project.

We expect about 52 percent of the sample to be enrolled for five years and the remainder for three years. This division is statistically optimal if the first six months of data are not useful because of catch-up demand and if the last six months of data are not useful because of transitory demand.

In the final year of the experiment, those enrolled in plans with coinsurance and deductibles (that is, Plans 2 through 15) will be permitted to supplement their experimental insurance at rates that simulate varying tax treatment of insurance premiums. This will provide evidence on the amount of additional insurance families might purchase if a national plan requiring out-of-pocket payments were enacted. After two years we will have a reasonably good estimate of the actuarial value of the plans and so will know what premium to charge for supplemental insurance. Permitting supplementation will introduce adverse selection into the data; however, this can be controlled for by using simultaneous-equation estimators. In any event the data from the last year may not be very useful for estimation purposes because of the problem of transitory demand. The issue of measuring demand for supplementary insurance is discussed at greater length elsewhere.

---

1. Sparger and Anderson (1973); Strauss and Sparger (1971).
STATISTICAL EXPERIMENTAL DESIGN

In this section we discuss methods of selecting families and allocating them to insurance plans, problems of refusal and attrition, and the expected precision of our estimates at the end of the experiment. These issues are discussed in greater detail elsewhere.¹

Four sites will be chosen for the experiment. From each, a clustered random sample of roughly 6,000 families will be chosen. These families will be given a screening interview to define the family unit; determine eligibility; and gather information on income, education, age, sex, race, and self-perceived health status. Using this information, we will choose a subset of 2,000 families to receive a lengthy baseline interview. These families will be chosen in accordance with the Finite Selection Model developed by Morris.² Given a fixed budget and an equation or a set of equations to estimate, the model chooses families that yield the optimal variation in explanatory variables to estimate the equation(s). The baseline interview verifies the information from the screening interview (going into considerably more detail on such variables as income and self-perceived health status) and asks questions about prior utilization and insurance. These data are important for selecting families and will also be a source of non-experimental data for analysis (analogous to the 1963 and 1970 surveys of the Center for Health Administration). From the 2,000 families given baseline interviews in each site, approximately 500 will be chosen for enrollment in the experimental plans and 300 assigned to the control group. Thus, with four sites, there will be some 2,000 experimental families and 1,200 control-group families in all.

The number of families assigned to any one plan will be determined by the Conlisk-Watts model;³ the resulting allocation is optimal for estimating an equation or set of equations. It is similar to the Finite Selection Model except that variables related to particular insurance plans, rather than demographic characteristics of families, are used in the optimization process.

¹Newhouse (forthcoming); Morris (forthcoming a, b).
²Morris (forthcoming a).
The families selected for enrollment will be allocated "randomly" to plans. While the Finite Selection Model can, in principle, select optimal family-plan combinations (given a model to be estimated), random allocation offers some protection against latent variables. (That is, one can be reasonably sure that any such variable will be balanced among the treatment groups.) By allocating randomly, one pays a price in efficiency of estimation (if there are no latent variables). The price paid can be kept small if random allocation is made subject to a constraint of near balance among the treatment groups.

In enrolling families, we make them an offer that ensures that they will be better off financially under any circumstances. As a result, we expect any refusals to participate to be random. This may not be correct, however; in addition, refusals to participate in screening and baseline interviews may also not be random. To test this, a larger payment will be offered by way of persuasion to a sample of those who refuse (at each stage). Analysis of data from these individuals should show whether there was any significant selection bias and, if so, should provide corrections for it. Attrition differs from initial refusals, since we have some data on those who leave; hence, we do not plan to re-enroll anyone who drops out.

What precision might be expected from data on 2,000 experimental families, 52 percent of whom are enrolled for five years and the remainder for three years? Because of uncertainties in estimating precision, we have defined "pessimistic" and "more likely" expected standard errors. In explaining this method of estimating precision, I discuss our ability to predict total expenditures; and at the end of this section I give expected precisions for particular medical services.

Two major factors affect our definition of "pessimistic" and "more likely" estimates. The first is the coefficient of variation. In the 1963 Center for Health Administration Studies Survey, the coefficient of variation across 2376 families for total expenditure was 1.34. We have used this figure to obtain a "pessimistic" estimate. However, the

---

1 The idea contained in this paragraph was suggested by Bradley Efron.

2 Stated formally, the constraint is that there is near orthogonality between the demographic and plan variables.
experimental expenditure data may have less measurement error, the data will be site-specific, and some of the variation will be explained. All of these factors will tend to reduce the residual coefficient of variation, the appropriate figure. As a "more likely" estimate, we have chosen 1.00, recognizing that even this may be high. The second factor concerns the model we wish to estimate. Our "pessimistic" estimate uses the model described in Eq. (1). A "more likely" model would assume that price can be represented by four parameters (rather than the 16 in Eq. (1)), in which case standard errors would be halved. Using these two definitions, Table 2 shows the estimated precision at the end of the experiment. With the 0.022 figure, a 95 percent confidence interval around expenditure in an "average" plan would be less than ± 5 percent of mean expenditures in that plan.

Table 2
ESTIMATED COEFFICIENTS OF VARIATION FOR TOTAL HEALTH EXPENDITURE USING EXPERIMENTAL FAMILIES

<table>
<thead>
<tr>
<th></th>
<th>Coefficient of Variation of Predicted Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Pessimistic&quot; Estimate</td>
<td></td>
</tr>
<tr>
<td>Assumes model specified in Eq. (1) and population coefficient of variation equal to 1.34</td>
<td>0.060</td>
</tr>
<tr>
<td>&quot;More Likely&quot; Estimate</td>
<td></td>
</tr>
<tr>
<td>Assumes price represented by four parameters and population coefficient of variation equal to 1.00</td>
<td>0.022</td>
</tr>
</tbody>
</table>

We next attempt to give some idea of the expected effects of other factors that could affect precision. Although their net affect is unknown, it is reasonable to believe that on balance they leave the above estimates approximately unchanged.

There are four reasons why these standard errors might be too low. First, some attrition may occur, so that the final sample will be fewer than 2,000 families. If the attrition occurs early, additional families
can be enrolled; later in the experiment, control of attrition may be better and many of the data will already be included. If there is a 10 percent (net) loss of family years, standard errors would rise by 5 percent. Second, although the above calculations have assumed that family years are independent, there is likely to be considerable serial correlation in health expenditures for families. If the correlation from year to year is 0.3, as might be expected, standard errors would increase by 10 percent. Third, the data from the beginning of the experiment may be useless because the steady state is not achieved or because of the Hawthorne Effect. If the first year's data are useless, standard errors will rise by 15 percent. There will also be some problem at the end because of transitory effects and because families will be able to supplement. However, the three- and five-year variation should permit measurement of the transitory demand, and the use of simultaneous-equation methods when families have chosen their insurance should mean relatively little degradation. As a result, precision should fall by substantially less than it would if the data were useless (15 percent). Finally, a number of parameters other than price will be estimated, which will cause a loss of degrees of freedom; estimation of 400 nuisance parameters (which seems high) will cause standard errors to rise by 2 percent.

A number of factors also operate to reduce the standard errors below those reported. First, the standard errors as reported assume a strictly random choice of families and allocation to plans. Based on Monte Carlo results, use of the Finite Selection Model and the Conlisk-Watts Model should reduce standard errors by at least 25 percent. Second, we plan to use Empirical Bayes methods in the analysis, particularly for site-specific parameters. A conservative estimate is that such methods could reduce standard errors by 10 percent. Empirical Bayes methods may be particularly useful for coping with the problem of within-family correlation of utilization. Third, the control

---

\[1\text{In the Palo Alto data described in Phelps and Newhouse (1972), the correlation between physician expenditures in 1966 and 1968 was 0.15, and the correlation for the number of visits was 0.3.}\]

\[2\text{Efron and Morris (1972).}\]
group and the baseline interview will provide additional data, although how much is unknown. Finally, part of the general philosophy of the experimental design is sequential implementation. Although this can provide no improvement per se in the estimates given above, it forms the basis for altering the design if the assumptions on which it was based prove to be inaccurate. For example, if it should appear that standard errors had been underestimated based on data from the first site, eliminating half the plans in the last three sites would reduce standard errors in Eq. (1) by 24 percent.

As might be expected, the coefficients of variation for expenditure on specific services are higher than the coefficients for total expenditure. Table 3 shows the factors by which standard errors must be increased if one is interested in expenditure on a specific service. These factors are the ratio of the coefficient of variation for the specific service to the coefficient of variation for total expenditure (1.34) in the 1963 Center for Health Administration Studies Survey. Of course, since dollar expenditure on such services is smaller than total expenditure, in most cases the absolute errors (in dollars) will be smaller than those for total expenditure.

Table 3

<table>
<thead>
<tr>
<th>Service</th>
<th>Increase Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital nonobstetric days</td>
<td>2.5</td>
</tr>
<tr>
<td>Physician office visits</td>
<td>1.2</td>
</tr>
<tr>
<td>Surgeon in-hospital expenditures</td>
<td>3.2</td>
</tr>
<tr>
<td>Prescription drugs</td>
<td>1.3</td>
</tr>
</tbody>
</table>

**BREADTH OF BENEFITS**

The following criteria were used in making decisions about which medical services should be covered: (1) The more likely a service would be covered under a national plan, the more likely the experiment would cover it; (2) the less the transitory or "catch-up" demand for
a service, the more likely the experiment would cover it; and (3) the lower the cost of the service, the more likely the experiment would cover it.

The rationale for our decisions will be explained in detail elsewhere.¹ In general, most services are covered, since it was felt that those making decisions about coverage as part of a national plan would value information on the effect of insuring many kinds of services. The way to generate such information is to include them in the scope of coverage.

There were some exceptions, however. It was felt that drugs not prescribed by a physician would never be covered in a national plan, so such drugs are excluded. (Prescribed drugs that could be obtained without a prescription are tentatively included.) Although dental services for children are treated like any other service (except orthodontia, which is excluded on the basis of all three criteria), a compromise was reached on dental services for adults. Such services may not be included in a national plan, are subject to catch-up demand, and do significantly affect cost; however, it was felt that some evidence on the responsiveness to price of dental services for adults would nevertheless be useful. Therefore, some or all of the families in Plan 1 (the free-care plan) will receive dental coverage for adults (again excluding orthodontia); the exact number will be chosen later. In addition, some families in Plans 4 and 5, the 50 percent coinsurance plans, may receive dental coverage, although in this case their out-of-pocket expenditures will not count toward the Maximum Dollar Expenditure (to preserve comparability with the other plans). Since the remainder of the sample will typically not have dental coverage, comparison of the utilization of those with experimental coverage with the utilization of the remainder of the sample will provide evidence at a relatively low cost on the responsiveness of demand for dental services to insurance.²

¹Clasquin (forthcoming).
²The possibility of adverse selection in the choice of dental insurance (those who expect to use more services having better insurance) seems unlikely, but will be tested by using a simultaneous-estimator.
Intensive psychoanalysis was excluded by limiting the coverage of outpatient psychiatric services to 52 visits per year; this was done on the basis of all three criteria. Those who use psychiatric services intensively (even for fewer than 52 visits) can be expected to satisfy their expenditure limit, which would mean that they face zero price for all other services. As a result, it was decided not to count expenditures on psychiatric services toward the expenditure limit, although they are covered at the coinsurance rate specified in the policy.

**SELECTION OF SITES**

There are two problems in site selection: determining the optimal number of sites, and then choosing the actual sites. The optimal number of sites can be determined from an extension of the Conlisk-Watts model.\(^1\) The number of sites is chosen that minimizes the sum of between-site and within-site variances, subject to a budget constraint and a cost function of the form

\[
C_s = C_f + \sum_{i=1}^{n} c_i
\]

where \(C_s\) is total cost per site, \(C_f\) is the fixed cost per site, and \(c_i\) is the expected cost of an additional observation in any site. For estimated values of between- and within-site variances in the experiment and estimated costs, the optimal number of sites appears to be about four.

Given that we wish four sites, we elected not to choose them randomly.\(^2\) With only four sites to be chosen, a strictly random selection could result in quite bad estimates of national parameters. In addition, purely random selection is impractical, because the cooperation of state and local officials and local providers is essential if the experiment is to succeed. Sites are therefore chosen purposively. The desirability of potential sites will be assessed by using the

\(^1\) Newhouse (forthcoming); Morris (forthcoming c).

\(^2\) Morris (forthcoming b).
Finite Selection Model. Site-dependent variables will be entered, and costs per site will be made proportional to a medical-care price index for the site. Two variables are of interest. One is the capacity utilization of the local ambulatory care system. This is being measured by a telephone survey of physicians asking about work loads, scheduling of patients, and waiting times for appointments. The other is the presence of an HMO, Experimental Medical Care Review Organization, and medical foundation. The inclusion of the latter two institutions is to test for the possible effects of peer review upon utilization outcomes.

In addition, there are certain constraints on the choice of sites. Since there is evidence of regional disparities in the utilization of health services as well as in capacity utilization, regional diversity will be sought in the sites selected. As noted, the cooperation of state and local officials (including the Governor, the Insurance Commissioner, the Mayor, and the local Community Action Program) and local providers is a precondition in the choice of any site.

The first site chosen for the experiment is Dayton, Ohio. The remaining sites have not yet been selected, but one is to be in the West and one in the South.

RULES OF OPERATION

One of the major advantages of conducting an experiment over analyzing non-experimental data is that one is forced to consider a great many detailed rules of operation that are generally not considered in the non-experimental analysis but may in fact be significant determinants of demand. Although there are too many such issues for full discussion here, a few are of sufficient importance to warrant brief mention. (A fuller discussion of the reasoning behind the actual rules of operation will be contained in a report now in preparation. ¹

In any plan containing income-related clauses, the family unit whose income is pooled, and the income itself, must be defined. Our definition of the family unit is based on the notion of an economic

¹Clasquin (forthcoming).
unit; persons who share income and expenditures are treated as one unit since this was felt to be the relevant decisionmaking unit for medical expenditures. Thus, we consider a self-defined head and his or her spouse to be the nucleus of a unit. In general, persons dependent upon the head or spouse and residing with them are treated as part of the same unit, their income and expenditures being pooled for purposes of the expenditure limit.

Rules must also be defined to handle changes in family units through departures or additions. Partly because a national plan could work well that way, partly because it simplifies analysis, we will recognize only departures at the beginning of a new accounting period. At that time, those who have left the unit will be placed in a new unit and kept in the study. Coverage will in general not be extended to additions to units because observations for a short period of time are not as useful (the problem of transitory demand), although newborn or adopted children and new spouses of heads will be covered.¹

Income is defined quite comprehensively to reflect the resources available to a family unit for contingencies. Measurement of a number of components of income will permit us to simulate the consequences of using alternative definitions.

In any plan that includes a deductible (or more generally that makes price a function of total expenditure), the accounting period for expenditures must be defined. We will use a fixed twelve-month period to minimize the effect of seasonality and concentration of discretionary expenditure. The accounting period is fixed, and not a moving average, because a moving average is difficult to analyze. We have rejected the notion of a carry-over of unreimbursed expenses near the end of one accounting period to another because of the difficulties it creates for analysis.

These are the issues that an income-related national plan would have to resolve. But many other issues are peculiar to the experiment.

¹Permitting additions other than these also raises the possibility of adverse selection, sick individuals becoming members of units with generous plans. Incentives for such additions in the experiment (which covers only a small fraction of the population) are much greater than the incentives would be in a national plan with universal coverage.
and would not be relevant to a national plan. Of these, two deserve discussion here, since they bear on the usefulness of the results and the ethics of the experiment. For both scientific and ethical reasons, families will be paid an amount of money sufficient to ensure that they will never be worse off financially from participating. This money will be independent of the family's utilization of medical care. It will therefore not alter the price of services and should be treated like other income by the family (see below). The formula for calculating this amount to be paid is

\[ P = B + \max_{E} \left[ 0, \max_{E} \left\{ I(E) - X(E) - R \right\} \right] \quad (4) \]

where \( P \) is the sum paid to the family, \( B \) is the amount paid for interviewing fees and returning monthly forms, \( I \) is the amount paid by the family's existing insurance plan(s) conditional on expenditures \( E \), \( X \) is the amount paid by the experimental plan, and \( R \) is the family's out-of-pocket insurance premium.

This formula may be explained as follows. The family heads are each paid $5 per interview (interviews are expected to be quarterly) and $10 per month for returning forms indicating family composition changes, job changes, or insurance changes. These payments are represented by \( B \). To guarantee that the family will never be worse off from participating, the largest possible difference between what it would receive under its old policy and what it would receive under its new policy is computed. This is \( I(E) - X(E) \) in the formula. For example, if a family has hospitalization insurance covering the first 30 days at a semi-private room rate, and we were to assign it to a plan with a 25 percent coinsurance rate subject to a maximum out-of-pocket payment of $1,000, we would pay the family $1,000. We would assume that a member of the family could spend 30 days in the hospital and could incur total expenditures of $4,000 or more. In this case the family would have to pay $1,000 with its experimental coverage; but because the bill might have been paid in full under its old plan, the family exposes itself to a potential $1,000 loss. From this "worst case"
payment is deducted any out-of-pocket premium payments the family would have made on its old policy, since it will no longer have to make these payments.

If the family were not protected against its worst case, families who expected high expenditures would differentially refuse plans requiring large out-of-pocket payments. This would have the effect of overstating the responsiveness of demand to insurance when comparing results across plans. In addition, if worst case payments were not made, the family would be in effect participating in a lottery in which it could, under some circumstances, lose. Some consider it unethical for the government to make such a lottery available.

The experiment will reimburse the family's out-of-pocket premiums if it keeps its prior insurance in force, for two reasons. First, doing so guarantees that no family will become uninsurable by participating in the experiment. If the policy were not kept in force, a person who became sick might find that he could not obtain insurance afterward. Second, keeping the policy in force guarantees that the family can withdraw from the experiment, if it chooses, and not be uncovered for a period of time.

In return for paying the family an amount equal to the most it could "lose" and reimbursing the family's out-of-pocket premiums, the families are asked to assign the benefits of their existing policies to us for services that we cover.¹ There are two reasons for this. By obtaining an assignment, we can be certain that the families are not being double-covered, thereby changing the price of medical care and defeating the purpose of the experiment. Second, by reclaiming against the family's old policy, it will be possible to recoup a portion of the funds paid to maintain the family's insurance in force.²

¹If we do not cover a service--such as dental services for adults in some plans--and the family's existing insurance covers such a service, the family will continue to be covered under the terms of its existing policy.

²Note that if there were no reclaiming, there would be a windfall to insurance companies, since they would be collecting premiums on a policy that would not be used.
Payments to the family for participating will, in general, be made monthly. This is done to maximize the possibility that the family will treat the money paid it as it would any other income; and therefore, when making decisions on medical care, the family will act as if it faced the price specified in the policy.\(^1\) Although it is unlikely, if the money were paid at the beginning of the accounting period, the family might "put it in a cookie jar," use it for medical care, and treat its medical care as free. In this case, the family's behavior would differ from its behavior under a national plan, where it would receive no payments. However, if the family does not put the money in a cookie jar, a different problem may arise. The family might spend its payment, then be hospitalized and not be able to meet its bills. In this case, the family will be referred to local lending institutions. (Note that over the succeeding twelve months the family will be paid an amount equal to its largest possible bill.) During the course of the experiment, we will attempt to find out how families finance their medical-care expenditures. If plans that call for substantial out-of-pocket payments cause the families financing difficulties, legislation that envisions such payments must give some attention to the financing question.

The monthly payment scheme will be changed in the final year; the money will be held in an escrow account for the family. The family may draw down its escrow account to finance medical expenditures, and any balance in the account at the end of the year will be paid to the family. The purpose of the escrow account is to prevent a situation where it is not in the family's interest to continue in the experiment, one that could occur if payments continued to be made monthly. For example, if an illness occurred in the middle of the last accounting period, a payment of $500 might be required from the family, but continued participation in the experiment might net the family only $250

\(^1\)To test whether the family treated its payments as other income, we will enter the amount of the payment as an explanatory variable. There will be some variation independent of income and plan because of the truncation factor and variation in out-of-pocket premiums on pre-existing coverage; additional variation, created by paying something extra to a few families, is being considered.
(if $250 had already been paid), while dropping out of the experiment and returning to its old insurance might save the family the entire $500. Since this would tend to decrease reported expenditures in plans with high coinsurance rates, it would lead to overstated responsiveness of demand.

A second issue peculiar to the experiment is the definition of eligibility. Those over 65 were deemed ineligible, since a national insurance program could well leave the Medicare program intact and since the different kinds of health problems among the aged would require a larger sample. Those with incomes greater than $25,000 were also eliminated. A fairly high income eligibility figure was chosen to permit generalization to most of the population. An income limit of $25,000 includes over 90 percent of the population, and it avoids the problem of using tax money to make large payments to very-high-income families. Those with access to the military medical-care system were also excluded, because it was felt that they might continue to use this system, making observations of their behavior non-comparable to the other observations. Foster children were excluded, on the assumption that the state would continue to pay for their medical care in full, even if a national plan were enacted.

SCHEDULE AND COSTS

In late November and early December 1973 fifty families were enrolled as a pilot sample. Fifty-five offers were made and there were five refusals. The purpose of the pilot sample is primarily to test the operating systems to eliminate flaws before enrollment of the regular sample. These fifty families will continue in the experiment for three years and will serve as a group upon which interview instruments can be pretested. The pilot sample also permitted the testing of enrollment techniques.

The regular sample of 500 families will be enrolled in Dayton in late summer 1974; enrollment of 500 additional families in the second site will take place around March 1975; enrollment is scheduled in the
third and fourth sites in September and December of 1975.\(^1\) Table 4 shows the cumulative number of family years by site by fiscal year if this schedule is followed. The last column of the table gives the multiple of the final coefficient of variation through time. The calculation of the coefficient of variation assumes that data from the first year are usable and that observations across time are independent.\(^2\)

As might be expected, precision improves rapidly at the beginning of the experiment. By the end of fiscal year 1977, confidence intervals are only about 40 percent as large as they will be at the end of the experiment. Thus, useful data should be available fairly early.

The total projected cost of the experiment (1973 dollars) is $32 million. Of this figure, $12 million will be paid to families either as claims payments or as worst case payments. (This assumes that on the average about $1,550 per year will be paid to families and that $250 per year can be reclaimed from existing insurance.) Some $6 million will be paid in local field costs, primarily for interviewers and quality-control operations. About $9 million of the cost is for design, analysis, and administration (exclusive of data processing), and $5 million for data processing. Administrative and data-processing costs are high because of the volume of data being collected and the need to collect it in ways that make it amenable to different types of analysis. For example, one may wish to look at a subset of families over time or all families at one point in time. Similarly, we wish to analyze the data both by illness episode and by annual expenditure, as described above. Also, one must be able to identify individuals with family units, which may change through the course of the experiment.

Although the cost estimates have been thoroughly reviewed, their uncertainty must be emphasized. The better one is able to predict

---

\(^1\) About one year is required from the date of beginning operations in a site to the date of completing enrollment. This allows time for randomly choosing areas of the city, listing dwelling units in those areas, conducting screening and baseline interviews, verifying insurance coverage of those in the sample, contacting local providers, and enrolling the families.

\(^2\) If there is substantial serial correlation within families across time, these multiples are too large.
claims payments, the less compelling the rationale for doing the experiment. In addition, many aspects of the project are novel, making the prediction of cost hazardous. Because of uncertainties with regard to both cost and feasibility, the Department of Health, Education and Welfare has, to date, authorized enrollment only in Dayton. A review of progress will be made before enrollment takes place in each remaining site. If the project is not fulfilling its objectives, it will be curtailed.

Table 4

CUMULATIVE NUMBER OF FAMILY YEARS BY SITE, BY FISCAL YEAR

<table>
<thead>
<tr>
<th>End of Fiscal Year</th>
<th>Dayton</th>
<th>Site 2</th>
<th>Site 3</th>
<th>Site 4</th>
<th>All Sites</th>
<th>Multiple of Final Coefficient of Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY 75</td>
<td>417</td>
<td>125</td>
<td>--</td>
<td>--</td>
<td>542</td>
<td>3.86</td>
</tr>
<tr>
<td>FY 76</td>
<td>917</td>
<td>625</td>
<td>375</td>
<td>250</td>
<td>2167</td>
<td>1.92</td>
</tr>
<tr>
<td>FY 77</td>
<td>1417</td>
<td>1125</td>
<td>875</td>
<td>750</td>
<td>4167</td>
<td>1.39</td>
</tr>
<tr>
<td>FY 78</td>
<td>1717</td>
<td>1565</td>
<td>1375</td>
<td>1250</td>
<td>5907</td>
<td>1.17</td>
</tr>
<tr>
<td>FY 79</td>
<td>1977</td>
<td>1825</td>
<td>1695</td>
<td>1630</td>
<td>7127</td>
<td>1.06</td>
</tr>
<tr>
<td>FY 80</td>
<td>2020</td>
<td>2020</td>
<td>1955</td>
<td>1890</td>
<td>7885</td>
<td>1.01</td>
</tr>
<tr>
<td>12/31/80</td>
<td>2020</td>
<td>2020</td>
<td>2020</td>
<td>2020</td>
<td>8080</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Assume enrollment complete in Dayton August 31, 1974; Site 2, March 31, 1975; Site 3, September 30, 1975; Site 4, December 31, 1975. Assume 500 families per site. The multiple of the final coefficient of variation is \((8080/x)^{1/2}\), where \(x\) is the cumulative number of family years at that time.
IV. THE SIGNIFICANCE OF THE RESULTS

The experiment must be appraised in the context of the debate over health care financing. In this debate, neither the costs nor the benefits of proposed legislation can be well predicted. One factor in estimating cost is the responsiveness of demand to insurance (price elasticity); yet existing estimates vary widely both in the popular press and in technical journals. For example, estimates of the effect of price on demand made in the economics literature differ by more than an order of magnitude from those made by the Social Security Administration.\(^1\) Answers to more specific questions concerning the effect of price—such as how large a deductible would affect the demand for ambulatory services, or how much supplementation of a plan requiring out-of-pocket payments would occur—are much less certain. Reliable estimates of the effect of tax-law changes are impossible to make.

The benefits of insurance are even less well understood than the costs. The evidence suggests that additional medical care consumption has little effect on measured indices of health status.\(^2\) Many observers find this statement counterintuitive and suspect that if different aspects of health status were measured, significant results would be found. By obtaining objective and subjective measures of health status over time, the experiment should make a considerable advance in the data available to examine the question of the effect of insurance on health status.

Some have commented that the experiment is likely to be too late—that it would be valuable if we had the information now, but by the time the experiment ends, a national plan will have been enacted. Apart from the obvious comment that this might not be true, two points might be made. Reasonably precise results will be available

\(^1\) Social Security Administration (1971); J. Feldstein (1971); Rosett and Huang (1973); Davis and Russell (1972).

\(^2\) Haggerty (1972).
early on; within two years confidence intervals will be about double their final level, and in four years confidence intervals will be only 20 percent higher than their final level (see Table 4). Furthermore, whatever plan is adopted is unlikely to remain immutable; the current debate over altering the Medicaid and Medicare programs attests to that.

Although the experiment has been designed to provide better answers to some of the principal questions in the present debate over health care financing, it is also a basic research project concerning health care financing. The project should generate not only results of immediate usefulness and policy relevance, but also a data base that should be a valuable asset for health service researchers for decades to come.

The principal dimensions of the Health Insurance Study are briefly summarized in Table 5.

Table 5
SUMMARY OF THE EXPERIMENT'S PRINCIPAL DIMENSIONS

- Four sites (projected)
- 2,000 experimental families, each randomly assigned to one of the 16 plans
- Families nearly evenly divided between three-year and five-year periods of participation
- 1,200 control-group families
- Plans vary coinsurance rate from zero to 100 percent
- All plans requiring out-of-pocket payments limit such payments to 5 or 15 percent of family income
- Some families assigned to a Health Maintenance Organization
- Opportunity for purchase of additional insurance in the final year
REFERENCES


