N-1346-HEW

February 1980

NONRESPONSE ISSUES IN PUBLIC POLICY EXPERIMENTS, WITH EMPHASIS ON THE HEALTH INSURANCE STUDY

Carl N. Morris

A Rand Note

prepared for the

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Rand
SANTA MONICA, CA. 90406

The research reported herein was performed pursuant to Grant 016B-7905 from the U.S. Department of Health, Education, and Welfare, 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.

Published by The Rand Corporation

N-1346-HEW

February 1980

NONRESPONSE ISSUES IN PUBLIC POLICY EXPERIMENTS, WITH EMPHASIS ON THE HEALTH INSURANCE STUDY

Carl N. Morris

A Rand Note

prepared for the

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Rand
SANTA MONICA, CA. 90406

PREFACE

This Note was prepared for a Symposium on Incomplete Data, Committee on National Statistics, National Research Council, August 10-11, 1979. Drawing upon his experience with the Rand Health Insurance Study, the author treats nonresponse issues that occur in social experiments. The research was supported by a grant from the Department of Health, Education, and Welfare.

SUMMARY

Missing-data problems, including nonresponse problems, are troublesome because the sampling probabilities are unknown and cannot be determined even with large samples. Judgment, combined with past experience and related data, must be used in the analysis when data are missing; the judgment should be reflected in expressions of uncertainty. Acceptance of the use of scientific judgment is long overdue, as are mathematical and computational methods for formally and explicitly incorporating subjectivity into missing-data formulas. Acceptance will be difficult because data analysts already use complex models and will be reluctant to add a further complication. Some combination of Bayesian theory and sensitivity analysis should provide the needed theoretical basis.

During the design of a study or experiment, one must decide whether to collect more observations or to diminish the number of observations and devote resources to reducing nonresponse or to estimating its effects. Scientists must often design surveys without knowing which analyses, what models, and what variables will be used. With this in mind, the principle of unbiasedness, which suggests that judgment is excluded, plays too important a role in thinking about design. Mean square error is a more useful concept because it permits cost-effective tradeoffs between sample size and bias (Morris et al., 1979).

To make better use of judgment, the scientific community needs to increase its understanding of the effects of nonresponse in different contexts. This understanding would be enhanced if experiments and surveys estimated the effects of nonresponse errors through dual and variable data collection systems. Although individuals have done this, the practice should be universal and the resulting information should be made centrally available.

Some of the nonresponse problems that arise in public policy experiments are discussed in this Note, and have been addressed in the Health Insurance Study. Such experiments are hampered by missing data: human populations may refuse to participate, they are mobile, and they may be affected by respondent burden. On the other hand, public policy

experiments usually have the financial resources needed to secure a considerable degree of participation, to follow up on nonresponse, to effect repeated measures, and to develop dual data collection capabilities. And any experiment or survey, through randomization and its design, can exert considerable control over the distribution of the observed variables.

The missing-data problem can greatly benefit statistics by forcing methodologists to recognize other viewpoints. Substantive knowledge, judgment, Bayesian theory, classical statistical theory, sampling theory, and robust model-fitting all help analysts to understand and deal with missing data. The statistics discipline will benefit from a merging of many viewpoints.

ACKNOWLEDGMENTS

Most of the ideas expressed in this Note were conceived and implemented jointly by Health Insurance Study (HIS) researchers. The work of Dr. Joseph P. Newhouse, the principal investigator, and Dr. Kent H. Marquis was especially useful in developing an approach to nonresponse in the HIS.

CONTENTS

I. INTRODUCTION: COMPARISONS AND CONTRASTS BETWEEN PUBLIC POLICY EXPERIMENTS AND SURVEYS

Analysts of social and public policy experiments encounter problems with nonresponse and measurement errors[1] similar to those of other survey analysts, despite the increased control available to experiments.

Nonresponse errors, whether for persons, forms, or items, may be classified into two categories: (1) those for which the distribution of missing values is known; and (2) those for which the distribution is at least partly unknown. The first category includes cases with data or people missing at random, or with available covariates that fully explain the nonresponse pattern. These cases present theoretical and conceptual problems, and can result in serious loss of precision. But the data contain the information necessary to compute consistent estimates, provided the analyst works hard enough. Accurate estimates can be derived for sufficiently large respondent samples, regardless of the nonresponse rate. Most analytical work and progress on incomplete-data problems probably will be made in this category.

This Note is concerned mainly with the second nonresponse category, in which case nonrespondents may differ from respondents in unknown ways. Section II treats some nonresponse issues that have been confronted in the design and analysis of the Health Insurance Study (HIS).[2]

---

[1]Although nonresponse problems are emphasized here, they are closely related to measurement errors because nonresponse errors can be converted to measurement errors by making imputations, and measurement errors can be converted to nonresponse errors through deletion of low-quality observations. We emphasize this point because in some cases it is difficult to distinguish between nonresponse and measurement errors, as, for example, when a subject fails to report an event and so a record of zero occurrences is recorded.

[2]Most of my views on nonresponse have been developed since 1972 while participating on the design and analysis of the Health Insurance Study, an HEW project funded at The Rand Corporation, Santa Monica, California, Joseph P. Newhouse, Principal Investigator. The experimental portion of the Health Insurance Study involves enrollment for 3 or 5 years of over 2000 families (experimental units) spread over six geographical sites (Dayton, Ohio; Seattle, Washington; Fitchburg-Leominster, Massachusetts; Franklin County, Massachusetts; Charleston, South Carolina; and Georgetown County, South Carolina). These families

Although public policy experiments face the same nonresponse threats as other surveys, time and size modify the problem. On the positive side,

1. The longitudinal nature of public policy experiments permits measurements over time on each subject, so that losing some measurements on a respondent is less serious than losing all. In Section IV, we consider the advantages of having at least some measurements on the dependent variables, and how this leads to a more satisfactory evaluation of the possible biasing effects of nonresponse.
2. Families have received substantial incentives to participate in public policy experiments, the main benefits being the treatments themselves, e.g., health insurance, income, and housing benefits. In addition, payments to compensate respondents for interview time, for completing forms, etc., have been justified by the overall expense of these experiments. Some respondents may have participated partly because they believed in the value and credibility of these national studies made with federal support for purposes of the U.S. Government. Cooperation of local government and societies (e.g., medical societies in the case of the Health Insurance Study) also has helped to secure respondent participation. (The need for this cooperation is one of the reasons why large public policy experiments have been constrained to a few sites each.) Even with these advantages, 19 percent of the 3469 families offered enrollment in the HIS refused (Morris, Newhouse, and Archibald, 1979, p. 222).

---

were randomly allocated to one of 16 insurance plans (treatments) of varying generosities (coinsurance and deductibles). Their health utilization, expenditures, and levels are monitored throughout their enrollment to determine the effects of the various financing characteristics on these variables. The experiment is still underway; about two-thirds of the family years were completed by August 1979. A further description of the HIS is given in the appendix.

The public policy experiments have also had disadvantages. Their longitudinal structure increases the likelihood of sample losses at some time because of participants' moving, dropping out, etc. And the participation burden extended over the long period provides disincentives to cooperate.

After mentioning some sources of nonresponse in Section II, we discuss methods for reducing nonresponse in Section III. Section IV addresses modeling issues and the analysis of nonresponse. Strong assumptions must be made in making such analyses, for one has no data from nonrespondents.

II. SOME SOURCES OF POSSIBLE NONRESPONSE

The three nonresponse issues mentioned here were chosen because they arose during the sampling and design phase of the HIS, and because they merit further attention. These issues neither exhaust such concerns in the HIS nor are they representative of the HIS. They have occurred in other experiments and panel surveys.

SPECIAL DIFFICULTIES IN LONGITUDINAL SURVEYS CAUSED BY NONSTATIONARY POPULATIONS: SIEVE MODELS, NON-URN MODELS

We have discussed this matter in a previous study (Morris, Newhouse, and Archibald, 1979), but it is worth mentioning again. Before families are enrolled in the HIS, they are interviewed at least twice: first in a baseline interview, and then, about 6 months later, in an enrollment interview. If the latter is successful, the family receives an offer to participate. Families are eligible to participate only if they meet certain requirements: for example, they must live in a certain area, not be in the military or institutionalized, have less than $25,000 annual income, have members under 62 years of age, have verifiable insurance, etc. Some of these requirements vary unpredictably from one interview to another, and so families are nonstationary with respect to eligibility. Further complications are caused by births, deaths, divorces, and other changes in family composition.

The usual model for cross-sectional (one-time) sampling is the urn model. The "urn" is equivalent to the eligibility criteria, the balls in the urn are the eligible subjects. Between sampling periods, the urn composition may be changed, as in a Polya urn scheme, although possibly in unknown ways. Repeated interviews therefore show that the urn is more like a "sieve." Families leave and return to eligibility between interviews. Field procedures designed to locate those returning to eligibility are prohibitively expensive (they involve returning to previously ineligible households, so there is some sample loss over time. Those families who are likely to lose eligibility are the less stable ones, and therefore differ characteristically from the more stable members of the sample population. A nonresponse bias will occur if stability correlates with the dependent variables being measured.

The HIS has partially controlled nonstationary problems by enrolling new families in dwellings vacated by a selected family that has left the area, even though this practice gives some new families a second chance of being surveyed. We did this because we believe that it is more important to maintain the stability of variables (in this case, the proportion of movers) than of individuals. The HIS thereby shifted second and third interview sampling criteria away from the rows of the usual data matrix to its columns, attempting to preserve the relationship of the variables rather than the sampling probabilities of individuals. This seems appropriate for a sampling theory based on sieve models.

A related rule applies to situations in which a married couple divorces and each remarries between surveys. The HIS retained one member of the original couple, at random, and included the new spouse. The other member and new spouse were abandoned. This also violates the usual sampling assumption, but seems justified because HIS treatments (insurance plans) apply to families as a unit, i.e., total family expenditure determines when the deductible is reached. A probability sample of individuals could be achieved by enrolling the two halves of the old families and ignoring the new spouses. But subsets of families could respond differently from entire families to anticipated forms of health insurance. This will cause bias of estimates produced by the experimental data from the probability sample. It is therefore better to make one original spouse a "nonrespondent," in the manner described.

These modifications of standard sampling rules might be uncontroversial, but the designers of the HIS felt it wise not to deviate from the accepted urn model norm. The preceding examples illustrate the inadequacy of standard sampling rules for longitudinal surveys. For a further discussion of this issue, see Section II of Morris, Newhouse, and Archibald, 1979.

Longitudinal sampling issues are included here because they lead to nonresponse through inadequate coverage, and this problem seems best treated by preserving certain characteristics of the population: mobility, family structure, etc. Work on theory, rules for substitution, and justification for sampling schemes seem much needed and long overdue. Generally accepted guidelines on these matters would have been

most welcome during the planning stages of the HIS. Lacking such guidelines, progress in this area will have to rely more on mathematical models than sampling statisticians historically have favored. But this seems inevitable if one is to treat nonresponse problems, whether they arise from nonstationary populations or otherwise.

NONRESPONSE PROBLEMS CAUSED UNINTENTIONALLY IN DEVELOPING THE DESIGN

Although the Incomplete Data Panel of the National Research Council is concerned mainly with nonresponse directly attributable to the surveyed population, nonresponse can be introduced unintentionally in the experimental design. Some examples follow.

A design based on imperfectly measured or stochastic variables can result in nonresponse. A simple example occurs in the HIS because only families with incomes under $25,000 (in 1973 dollars) are eligible for the experiment. Income is stochastic and imperfectly measured. The rule excluding families earning over $25,000 in the year before the experiment therefore excludes those families with high preexperimental incomes that drop below $25,000 during the experiment. The HIS suffered little from this exclusion because only a small fraction of potentially eligible families had higher incomes. But what happens in the experiments and surveys in which only the low-income population is under study? In the negative income tax and income maintenance experiments, for example, regression of incomes to the mean income (which is above the participation cutoff) occurred because selection was based on low income in one year. Since permanent income (the nonrandom component of income) differs from that used to determine the sample, the low-permanent-income families who happen to have randomly high incomes in the baseline year are excluded (and hence are "nonrespondents"). These families could differ in important ways from those surveyed. How should we adjust for this unknown distortion? Some efforts have been made to model the censored data arising from such selection procedures, but this problem is often ignored. It may be handled better in the design phase than in the analysis.

A related problem occurs with an inadequate sampling frame, as in oversampling certain neighborhoods, or sampling from subpopulations

thought to have desired characteristics. In seeking low-income families, perhaps by restricting search to poor neighborhoods, one builds in nonresponse from the poor in high-income neighborhoods. This will cause biased estimates if the "nonresponding" group differs from those surveyed.

An "optimal design" for an assumed model excludes those elements of the population that are inefficient for estimating the model. This practice can be disastrous if the model is incorrect, or if quite different models must also be estimated. Most of the best-known public policy experiments have avoided use of such models. One was used in the New Jersey Negative Income Tax Experiment (Conlisk and Watts, 1969) to assign higher-income families to generous plans and lower-income families to less generous plans, thereby cutting costs. This confounding of income and treatment continues to cause difficulties for analysts of that experiment (Keeley and Robins, 1978; Cogan, 1978).

NONRANDOM ATTRITION CAUSED BY UNEQUAL RESPONDENT BURDEN

Families suffering from severe health problems are required to file more forms; because this task is hardest when one is sick, missing data may be associated with health status.

An example of this phenomenon occurred early in the Dayton experiment. The HIS attempted to induce some families to participate in the data collection while continuing to use the health insurance plan (or none) they had before the experiment. These families were compensated for participating and filling out forms in the same way as other families, but the HIS did not reimburse them for their health expenditures. The enrollment rate for families not offered HIS insurance was similar to those offered an insurance plan, but this group suffered higher attrition. Disproportionately many families may have terminated participation following a high incidence of physician and hospital visits. The HIS could not pay more to those incurring this burden, for that would have been an incentive to overreport. A much larger fixed incentive was possible, but was not considered worth the cost. Ultimately, this group of Dayton families was terminated by the HIS.

III. METHODS FOR REDUCING AND DETECTING NONRESPONSE

REDUCING NONRESPONSE IN THE HIS

Families receive substantial incentives in the HIS to participate, complete questionnaires, and report their health activities. Financial incentives are an accepted way to reduce nonresponse. Obviously care must be taken to make certain that the incentive is uncorrelated with the response level.

Public relations played an important role in getting families to participate in the HIS. Before beginning enrollment in each site, the HIS sought and obtained cooperation from local government and from medical and dental societies. Interviewers assured prospective participants (a) that the study was being conducted by the U.S. Government for legitimate purposes, with cooperation of local authorities; (b) that they could not be worse off for participating; and (c) that they could terminate whenever they chose to do so.

Repeated attempts were made, when necessary, to locate individuals contacted in the preenrollment surveys, and to convince refusers of earlier surveys to participants in later ones. Incentives to participate increased throughout the preenrollment period. (Families had no idea, initially, that they ultimately would be offered free health insurance.)

Policy interest in health insurance includes plans with a 100-percent coinsurance rate (coinsurance is the fraction that the family pays until a deductible is met, after which health care is free). Since families on 100-percent coinsured plans have less incentive to report their health activities than those being reimbursed, some noncomparability exists among the data taken from different insurance plans. To counteract this problem, the HIS substituted a 95-percent coinsurance treatment for the 100-percent treatment. The extra 5 percent on covered expenses provides a financial incentive for filing forms and a clear signal that all expenditures must be reported. Thus, a slight modification of the treatment has inhibited nonresponse.

METHODS FOR DETECTING NONRESPONSE

When a missing entry appears on a form, a nonresponse has occurred. A nonresponse also occurs in the HIS if a family fails to file a claim after a medical event. Since no claim is expected unless a medical event occurs, nonresponse can go undetected. How then does one detect nonresponse? The HIS has used several approaches, some of which are discussed in Newhouse et al., 1979.

The most expensive methods for detecting nonresponse and estimating its magnitude depend on dual data collection systems to provide independent comparisons. For example, families are asked to complete "health reports" (self-reporting questionnaire forms) every 2 weeks, in which medical events are recorded. These reports are compared with claims filed by the families or their physicians. Discrepancies result in follow-up procedures. Any discrepancies remaining after follow-up indicate the magnitude of nonresponse. Physician-claim underreporting was about 2 percent in the first year of the first site.

A second dual reporting system is being fielded in the Massachusetts and South Carolina sites. Physicians in these sites will be contacted and presented with lists of HIS participants whom they may have served. The physicians (most likely their office staffs) will record the listed individuals who are their patients, the number of visits, and the charges for those visits. These results will be matched with the participants' claims to estimate the existence and extent of family nonresponse and misreporting. The payment compensating physicians for time spent by their staffs makes this procedure expensive. In a pilot test of the procedure, 80 percent of the physicians complied with the request, a higher rate than expected and sufficient to justify proceeding.

Nonresponse is most easily detected by varying a critical parameter and observing the results. (This permits detection of aggregate, but not of individual, nonresponse.) In different sites and years, the HIS has varied on (a) whether or not health reports must be filed; (b) health report filing frequency; and (c) whether or not families are reminded to file health reports (Newhouse, 1974). Reporting rates are then related to these variations. Presumably, more frequent filing and

prompting will lead to more complete health records, but the procedure is expensive. This increased contact can also have undesirable side effects; it might, for example, stimulate families to use medical services.

Longitudinal studies offer some hope for testing for nonresponse effects, providing earlier observations on the dependent variable. For example, HIS families were asked at the baseline stage about health expenditures and the number of physician visits in the previous year. These are two of the main dependent variables in the economic portion of the experiment. Thus, it is possible to compare families that enrolled with those that did not in terms of past values of the dependent variables. The procedure used was as follows:

All families were asked to respond to baseline questionnaires (for a fee) months before enrollment. Few refused. Later, certain baselined families were not enrolled: some because they were not selected for an offer; some because they had moved, or could not be found to receive the enrollment offer, etc.; and others because they refused the enrollment interview or enrollment offer. In Dayton (the only site for which this analysis has been completed), this process resulted in 1822 persons enrolled in experimental status and 1046 individuals eligible for enrollment, but not enrolled.

These two Dayton groups were compared in terms of the aforementioned baseline dependent variables, and were found to be in better than random agreement. This is more reassuring than comparing covariates associated with the dependent variables, because important covariates may not have appeared in the questionnaire (R^2 values using observed variables to explain expenditure tend to be low). Use of the preexperimental value of the dependent variable incorporates the effects of all variables, latent and measured, and so avoids this defect.

The lesson here is that longitudinal studies should be designed to collect versions of the dependent variable at the earliest possible opportunity, so that subsequent effects of sample losses, refusal, and attrition can be evaluated. Baseline variables can also be used to produce a highly balanced design. These have been the most important uses of the HIS baseline data.

IV. ANALYSIS OF NONRESPONSE

Let y and x be the vectors of dependent and independent variables. Let $f_T(y,x)$ be the true joint density of (y,x), p(y,x) the probability of observing (y,x), and $f_R(y,x)$ the joint density of respondent observations. Only the case of both (y,x) or neither (y,x) being observed is treated here. Then

$$\pi f_R(y,x) = p(y,x)f(y,x) \quad ,$$

with $\pi \equiv \iint p(y,x)f(y,x)\,dydx$ being the marginal probability of making an observation. In most studies, analysts can infer π and f_R from observed data. They would like to know f_T, or equivalently, p(y,x). The nonresponse problem, the problem of estimating f_T or p, is possible if any of the following approaches are used:

(a) p(y,x) = constant = π. This is the most frequently made assumption. Data are assumed missing at random, so $f_R = f_T$.

(b) p(y,x) = k(x) depends in a known way on observed covariates. Methods that weight observations, or adjust estimates so that the distribution of covariates matches a given population, fall in this category. So do hot-deck procedures (Bailar and Bailar, 1979).

(c) If bounds can be placed on y, and π is close to unity, worst-case assumptions about the unobserved values provide adequate upper and lower bounds for estimates (e.g., Birnbaum and Sirken, 1950). Birnbaum and Sirken's example is equivalent to choosing p(y,x) = p(y) to be two extremal values, thereby defining two extremal densities f_T.

(d) Outgrowths of (c) include sensitivity analyses, and the use of expert judgment to derive the effects on results of moderate upper and lower bounds for y, and of moderate variation in p(y,x).

(e) Method (d) evolves naturally into a Bayesian approach, parameterizing f_T directly, and making subjective assumptions about these parameters (e.g., Rubin, 1977). This is a useful approach because it formally acknowledges and incorporates the additional uncertainty (width of interval estimates) due to nonresponse effects. To illustrate

a Bayesian analysis of nonrandom nonresponse, let μ_T be the mean response of a population, to be estimated, and write $\mu_T = \pi\mu_R + (1 - \pi)\mu_N$, with π the proportion of respondents, and μ_R and μ_N the mean response for respondents and nonrespondents. Let $\hat{\mu}_R$ be the sample mean for n_R respondents. Make the Bayesian assumption that $\mu_N \sim N(\mu_R, \tau^2)$. Then a Bayesian calculation, such as Rubin's, gives

$$\text{var}\ (\mu_T|\hat{\mu}_R) = \sigma^2/n_R + (1 - \pi)^2\tau^2 \geq \text{var}\ (\mu_R|\hat{\mu}_R)\ ,$$

with σ the standard deviation of observations. If $\tau = 0$, the data are missing at random. If $\tau > 0$, the variance, relative to σ^2/n_R, which is the variance assuming data are missing at random, is controlled by keeping the response rate high so that $1 - \pi$ is small. Still, $\tau > 0$ increases the variance above the naive value. By letting the unknown τ vary, the sensitivity of this variance can be investigated.

Suppose that $\pi = 0.5$ is observed and f_R is a half-normal. Do we assume that f_T is a half-normal, i.e., that the missing data are missing at random, or that f_T is normal, in which case the missing values are all larger than those observed? The econometrics papers just mentioned provide appropriate theory and methods for the case of f_T normal, but the observed data, no matter how abundant, cannot help one choose among the alternatives for f_T. The results of such an analysis will be extremely sensitive to the normality assumption for f_T (actually, the conditional distribution of y given x is assumed normal), and so the user of such methods must rely on solid information from other sources about the assumed distribution for f_T.

Econometricians have recently fitted complicated models into multiple and multivariate regression frameworks that asume the unobserved f_T to be normal (Heckman, 1979; Hausman and Wise, 1979).

All missing-data procedures depend on judgment: either the form of $f_T(y,x)$ or of $p(y,x)$ must be postulated without data to verify the postulate. In this sense, all missing data procedures are Bayesian. Since subjective information must be used, no valid grounds exist for avoiding Bayesian methods in analyses of nonresponse data.

Authors proposing missing-data methods necessarily base their analyses on an unverifiable assumption, an "heroic assumption." They must

clearly indicate this assumption. In most analyses it is at level (a), the most "heroic," or (b), as in many of the papers of the National Research Council (1979). This seems appropriate, since analysts must often fill in missing values, derive estimates, or determine precision with minimal effort. But while (a) and (b) are appropriate conventions, they are not neutral assumptions. Rather, they are highly favorable in that they assume minimal uncertainty. Nonrespondents may often be unlike respondents, and the nonresponse bias would contribute error components that are omitted in (a) and (b), as in the τ^2 term in the example (e). Conservative statisticians try first to minimize the amount of nonresponse, and then to estimate the effects of nonresponse through supplemental data. Accounting for nonresponse through statistical analysis is the least preferred, however necessary, method.

Appendix
A BRIEF DESCRIPTION OF THE HEALTH INSURANCE STUDY

The HIS has several objectives, including the following: (1) to measure the insurance elasticity of demand for medical care services (i.e., the response to varying the portion of the expenditure that the participant must pay out of pocket); (2) to determine if the insurance elasticity of demand interacts with permanent income; (3) to determine what effects on health, if any, are observed when variations occur in the consumption of medical care services because of differences in amount paid out of pocket. To achieve these ends, nearly 2800 families have been enrolled in the experiment; these families are located in six geographic locations (Dayton, Ohio; Seattle, Washington; Fitchburg-Leominster, Massachusetts; Franklin County, Massachusetts; Charleston, South Carolina; Georgetown County, South Carolina).

The families are enrolled in one of 16 health insurance plans that vary the fraction of the total expenditure that the participant must pay; that fraction is either 0, 25, 50, or 95 percent. In addition, the family's financial exposure is limited to a certain amount in any one year; this amount is called the Maximum Dollar Expenditure (MDE). Generally, the MDE is set as a fraction of income, but in one plan it is $150 per person. Some families are also assigned to a Health Maintenance Organization (HMO) (prepaid group practice), and their care is free as long as it is received at the HMO. Families participate for either 3 years (70 percent) or 5 years (30 percent) in order to measure transitory behavior at the beginning and end of the experiment. Several years were needed to allow for transitory demand to disappear (i.e., rates of consumption that do not reflect steady-state behavior, such as, restorative dentistry, which would be done on a one-time basis) and for health status effects to appear.

During the period of participation in the experiment, families do not use their own health insurance; rather, they assign the benefits of that insurance to the experiment. They are paid lump sums (not based on utilization) to ensure that they will not be worse off financially

by participating in the experiment. They do not have a choice of insurance plan within the experiment; instead, they are made an all-or-nothing offer to participate in the plan to which they have been assigned.

Families were enrolled using the following procedure: (1) A screening interview was administered to determine eligibility (the aged and certain other populations are not eligible). (2) A baseline interview was administered to the eligible families to elicit certain information--in particular, information about health insurance policies. This information was verified with the employer or insurance company and used as the basis for the guarantee to the families that they will not be worse off by participating. (3) Following verification of the insurance information, families were selected, assigned to insurance plans (experimental treatments), and made an offer to enroll.

The experiment is well underway. All families are enrolled, and approximately 75 percent of the ultimate number of person-years have been completed (as of August 1979).

REFERENCES

1. Bailar, Barbara A., and John C. Bailar III, "Comparison of the Biases of the 'Hot Deck' Imputation Procedure with an 'Equal-Weights' Imputation Procedure," published by the National Research Council in *Symposium on Incomplete Data*, Washington, D.C., The Council, 1979.

2. Birnbaum, Z. W., and Monroe G. Sirken, "On the Total Error Due to Non-interview and to Random Sampling," *International Journal of Opinion and Attitude Research*, Vol. 4, No. 2, 1950, pp. 179-191.

3. Cogan, John F., "Negative Income Taxation and Labor Supply: New Evidence from the New Jersey-Pennsylvania Experiment," R-2155-HEW, The Rand Corporation, February 1978.

4. Conlisk, John, and Harold Watts, "A Model for Optimizing Experimental Designs for Estimating Response Surfaces," *Proceedings of the Social Statistics Section*, American Statistical Association, 1969, pp. 150-156.

5. Hausman, Jerry A., and David A. Wise, "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," *Econometrica*, Vol. 47, No. 2, March 1979, pp. 455-473.

6. Heckman, James J., "Sample Selection Bias as a Specification Error," *Econometrica*, Vol. 47, No. 1, January 1979, pp. 153-161.

7. Keeley, Michael C., and Philip K. Robins, "The Design of Social Experiments: A Critique of the Conlisk-Watts Assignment Model," Center for the Study of Welfare Policy, SRI International, Research Memorandum 57, November 1978.

8. Morris, Carl N., Joseph P. Newhouse, and Rae Archibald, "On the Theory and Practice of Obtaining Unbiased and Efficient Samples in Social Surveys and Experiments," in Vernon L. Smith (ed.), *Research in Experimental Economics*, Vol. 1, JAI Press Inc., Greenwich, Conn., 1979, pp. 199-230.

9. National Research Council, *Symposium on Incomplete Data*, Washington, D.C., The Council, 1979.

10. Newhouse, Joseph P., "A Design for a Health Insurance Experiment," *Inquiry*, Vol. 11, No. 1, March 1974, pp. 5-27.

11. Newhouse, Joseph P., Kent H. Marquis, Carl N. Morris, Charles E. Phelps, and William H. Rogers, "Measurement Issues in the Second Generation of Social Experiments," *Journal of Econometrics*, Vol. 11, No. 11, September 1979, pp. 117-129.

12. Rubin, Donald B., "Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys," *Journal of the American Statistical Association*, Vol. 72, No. 359, September 1977, pp. 538-543.

RAND/N-1346-HEW