

**working paper
department
of economics**

MACRO-EXPERIMENTS VERSUS MICRO-EXPERIMENTS
FOR HEALTH POLICY*

by

Jeffrey E. Harris**

Number 286A

July 1981

**massachusetts
institute of
technology**

**50 memorial drive
cambridge, mass. 02139**

MACRO-EXPERIMENTS VERSUS MICRO-EXPERIMENTS
FOR HEALTH POLICY*

by

Jeffrey E. Harris**

Number 286A

July 1981

Department of Economics
Massachusetts Institute of Technology
Cambridge, MA 02139

*Presented at the NBER Conference on Social Experimentation, Hilton Head Island, March 5-7, 1981. To appear in J. Hausman and D. Wise (eds.), Social Experimentation. Chicago: University of Chicago Press.

**Research supported by Public Health Service Research Grant No. DA-02620 and Research Career Development Award No. DA-00072.



Digitized by the Internet Archive
in 2011 with funding from
MIT Libraries

1. INTRODUCTION

In social micro-experiments, the experimenter assigns treatments and gauges responses at the individual level. The response of each individual is assumed to be independent and small in comparison to the market or social system.

In social macro-experiments, treatments are assigned at the group, community, or market level. The responses of entire social units, as well as of individuals within each unit, are the objects of interest. The responses of the individuals within each unit are correlated (Rivlin, 1974; Mosteller and Mosteller, 1979).

Economists and other social scientists, I contend here, have spent disproportionately too much effort on the design and interpretation of micro-experiments. The potential value and limitations of macro-experiments have not been adequately characterized. Accordingly, we need to develop a new science of macro-experimental design, and to articulate more carefully the tradeoff between micro and macro designs as guides to public policy.

My argument is framed within the context of health policy experiments. I concentrate on two policy issues: the effect of changes in health insurance coverage on the demand for medical care; and the effect of life-style intervention on the risk of coronary heart disease (CHD).

In Section 2, I point out several problems in the design, implementation, and interpretation of micro-experiments for health policy. These include subject selection and attrition, anticipatory responses, Hawthorne effects, and ethical constraints on individual randomization. Although the results of micro-experiments may elucidate certain mechanisms of

individual behavior, they may not reveal the total, market equilibrium effects of policy alternatives.

Section 3 considers how macro-experiments may resolve these micro-experimental difficulties. Because macro-experimentation can be less intrusive upon individuals, these experiments may avoid the potential selection and attrition biases, Hawthorne effects, and ethical constraints characteristic of micro-experiments. Most important, macro-experiments can be more useful for evaluating the total market or social system effects of policy options.

In Section 4, I discuss two serious limitations of macro-experimentation. First, intervention at the market or community level reduces the statistical power of the experiment and, in some cases, threatens its external validity. Second, the macro-experimenter may be likely to encounter significant political and administrative obstacles to randomization.

Section 5 considers how these defects of macro-experimentation might be avoided. Decentralization of macro-experiments, along with experimental blocking, is suggested as a means of improving statistical power and overcoming administrative barriers to randomization. Time series experiments, cross-over designs, as well as mixtures of micro and macro designs, are considered. To resolve questions of external validity, I show how the results of different macro-experiments might be combined.

Throughout the analysis, I focus on the experience of two micro-experiments--the Rand Health Insurance Study (Newhouse, 1974) and the Multiple Risk Factor Intervention Trial (Multiple Risk Factor Intervention

Trial Group, 1976a,b)--and one macro-experiment--the Stanford Heart Disease Prevention Program (Farquhar, 1978). Several other macro-experiments in life-style intervention are in progress or under consideration.¹ But no bona fide macro-experiment in health insurance or medical care utilization has been undertaken. One goal of Section 5 is to suggest how such experiments might be executed.

This paper is not a broad endorsement of macro-experimentation for health policy. It does not advocate the abandonment of micro-experiments. Nor do I envisage a strict choice between micro and macro designs. But in many cases, precise micro-estimates of only one or two parameters of a problem do not justify our plunging into full-scale policies. Less precise macro-assessments of the total impact of contemplated policies may then be warranted.

2. PROBLEMS WITH MICRO-EXPERIMENTS

First, I set forth the background of two micro-experiments in health policy.

The Multiple Risk Factor Intervention Trial (MRFIT)

Epidemiologists have repeatedly shown that high blood pressure, elevated

¹The Stanford Five-City Project (Hulley and Fortmann, 1980); the North Karelia Project (Puska et al., 1978); the Minnesota Community Prevention Program; the Pawtucket Heart Health Program; the European Collaborative Heart Disease Prevention Project (WHO European Collaborative Group, 1974; Rose et al., 1980); and the Pennsylvania County Health Improvement Program (Stolley, 1980).

blood cholesterol, and cigarette smoking are independent, powerful predictors of an individual's risk of fatal and non-fatal events of coronary heart disease (Truett et al., 1967). Men and women who spontaneously quit smoking incur a lower risk of subsequent coronary events than continuing smokers (Friedman et al., 1981). These findings have been derived from the natural histories of various study populations (for example, residents of Framingham, Massachusetts). To dispel objections that such predictive relationships are not really causal, it would be logical to attempt to reverse each of the above "risk factors" in a randomized experiment.

Separate clinical trials have been instituted to lower blood cholesterol, to treat hypertension, and to induce smoking cessation (Davis and Havlik, 1977; Hypertension Detection and Follow-up Program Cooperative Group, 1979a, 1979b; Rose and Hamilton, 1978). The difficulty with such single-factor experiments is that participation in the trial is a total experience (Syme, 1978). An experiment may be designed to test the isolated effect to lowering blood pressure. But when subjects are instructed to take antihypertensive medications, and possibly to restrict salt and caloric intake and increase physical activity, they inevitably modify dietary fat intake, smoking, and other aspects of behavior.

The Multiple Risk Factor Intervention trial (Kuller et al., 1980; MRFIT Group, 1976a, 1976b, 1977; Sherwin et al., 1979) recognized this limitation of single-factor trials. The protocol was designed to test the hypothesis that lowering serum cholesterol by diet, reducing high blood pressure by diet and drugs, and cessation of cigarette smoking, in combination, would result in a reduced risk of death from CHD. Men aged 35 to 57, who smoked

cigarettes, had elevated blood pressure and cholesterol, but who displayed no initial evidence of CHD, were to be followed for six years. After initial screening of 361,661 subjects during 1974-76, a total of 12,866 subjects were randomly assigned either to a program of special intervention (SI) directed toward these risk factors or to their usual source of medical care (UC). The experiment is being conducted at MRFIT clinics in 22 sites across the country, and is scheduled for completion in early 1982.

The Rand Health Insurance Study (HIS)

The responsiveness of medical care demand to price is an important factor in the design of health insurance and the control of rising medical expenditures. Price elasticities of demand for medical services have been estimated from a variety of data sources. But the main source of price variation in these non-experimental data is the terms of insurance coverage. Since consumers select their insurance on the basis of health status, income, family composition, and other factors affecting demand, such estimates could be seriously misleading.

The Rand Health Insurance Study (Manning, Morris, Newhouse et al., 1981; Manning, Newhouse, and Ware, 1981; Morris, 1979; Morris, Newhouse, and Archibald, 1980; Newhouse, 1974; Newhouse et al., 1979) was designed to overcome this limitation. A sample of approximately 8000 individuals in 2823 families was enrolled in six sites across the country. Families were enrolled in one of 14 different HIS insurance plans for either three or five years. These plans ranged from free care, to 95 percent coinsurance below a maximum dollar expenditure, to assignment in a prepaid group practice. Low-

income families were oversampled. Persons eligible for Medicare, heads of households 61 years of age and older at the time of enrollment, members of the military, and the institutionalized population were excluded. Enrollment of subjects at the Dayton site was completed in 1975, while enrollment at the Georgetown County, S.C., site was completed in 1979. In addition to analysis of the effects of various insurance plans on medicare demand, the effects of coverage on health status (Brook et al., 1979; Ware et al., 1980), certain administrative aspects of health insurance, and the effects of HMO care are under study.

Both MRFIT and HIS can be legitimately called second generation social experiments. Their designers took advantage of considerable prior experience in clinical trials and social experimentation. Nevertheless, these micro-experiments exhibit important difficulties in design, execution, and interpretation. These difficulties are now considered.

Subject Selection and Other Pre-Experimental Biases

In MRFIT, subjects were initially screened, primarily at work sites, by a series of medical examinations (Kuller et al., 1980). Those eligible at the first screening, on the basis of blood pressure, cholesterol and smoking habits, were invited to a second, more detailed medical screening, at which time the purpose and duration of the study were explained. For those who returned for the third and final screening, informed consent was obtained and then randomization was performed. Since the trial was aimed at men with high CHD risk, and since the experiment could not be blinded, potential subjects

were necessarily informed of their medical status during the screening process.

It is reasonable to suspect that the initial volunteers in this experiment were highly motivated and therefore more susceptible to intervention than the general population. Of those subjects initially eligible by risk factor criteria, about thirty percent declined to participate. Many of them merely refused to consider quitting smoking. It is also hard to imagine that the screening process itself had little effect on subjects' behavior and attitudes. Among those subjects who were ultimately randomized, mean diastolic blood pressure declined by about 10mm Hg from the first to the final screening examination, while the fraction of smokers declined by about five percent. Comparable changes were observed in blood cholesterol. These results may reflect changes in measurement methods between screening exams or statistical regression to the mean. Nevertheless, the evidence suggests that the pre-experimental phase constituted a form of life-style intervention.

The planners of MRFIT screened for subjects with high CHD risks in order to increase the statistical power of the experiment (MRFIT, 1977).² But this practice is not without its problems. Blood pressure, cholesterol, and smoking are undoubtedly influenced by such factors as diet, stress, physical activity, socioeconomic status, family history, occupation and peer pressure,

²Selection was actually based on "modifiable risk," which is not necessarily synonymous with "high risk." This modifiable risk score was based on a multiple logistic model of CHD risk, estimated from the Framingham study data (Truett et al., 1967), in combination with educated guesses about differential success rates in reducing risk factors.

many of which are difficult to measure. These additional, unmeasured variables also affect how subjects' CHD rates respond to experimental intervention. Pre-experimental screening on the basis of blood pressure, cholesterol, and smoking can produce a population of subjects that is highly unrepresentative with respect to the unmeasured variables. Some men who qualify for this study will be former smokers who have backslided into the habit as a result of, say, transient job-related stress. Others will be light smokers who have transient elevations in blood pressure due to, say, excessive salt use or weight gain. Still others will be inveterate heavy smokers. Although the experiment would still yield an unbiased estimate of the effect of special intervention among those patients who qualified, it is not clear how the estimated experimental effect relates to the overall population response. This difficulty applies not only to experimental responses in risk factors, but also to the effect of intervention on CHD incidence. It is compounded further if the additional, unmeasured variables also affect subject attrition during the experiment.

In the Health Insurance Study, the experimenters randomly sampled dwelling units and conducted initial interviews in order to ascertain the occupants' ages, incomes, and other data pertinent to eligibility. A baseline interview was administered to eligible families in order to elicit information about prior insurance status. Following verification of the insurance information, families were selected, assigned to the various plans, and contacted for an enrollment interview (Newhouse, 1974; Morris, 1979; Morris, Newhouse, and Archibald, 1980). If the assigned plan represented less extensive insurance than the subjects had prior to entry, then the

experimenters offered them a compensating incentive payment, in fixed installments, but unconditional upon subsequent medical care consumption. Consent to participate in the study was elicited after these steps had been taken. Of 3863 families who completed baseline interviews and were assigned to treatments, 10 percent refused the enrollment interview. Of those who agreed to the enrollment interview, 19 percent refused the offer to enroll.

The HIS incentive payment scheme was intended to ensure that subjects in all treatment groups were no worse off financially by participating in the experiment. At worst, such payments were supposed to have a small income effect on demand. Nevertheless, with refusal rates on the order of 20 percent, it is worth inquiring whether prior assignment to a plan could have affected the decision to participate in the experiment. Those families assigned to the high coinsurance plans were more likely to receive incentive payments. In these families, the decision to participate should depend more heavily upon attitudes toward risk, expectations about subsequent health care utilization, and other unmeasured variables. In fact, families who expect to make substantial use of medical care will be more likely to refuse to participate in the high coinsurance plans. It is at least arguable that these phenomena will result in an overly optimistic estimate of the effect of cost-sharing on the medical care use.

In both MRFIT and HIS, data have been collected on the characteristics of those subjects who refused to participate at the various pre-experimental stages, at least beyond the initial screening. It may thus be possible to assess some of the determinants of the decision to participate, and to correct for potential non-participation biases. But the determinants of the

decision to participate, it must be recognized, are not easily measured. So long as such intangibles play an important role, potential non-participation biases cannot be completely excluded. Moreover, replenishment of non-participants on the basis of observed characteristics, as suggested by Morris, Newhouse, and Archibald (1980), could be inappropriate.

Subject Attrition Biases

Since MRFIT and HIS are still in progress, little information on attrition rates has been published. In the Health Insurance Study, the 3-year cumulative attrition rates for the free plans and non-free plans have been 4 percent and 8 percent, respectively. In the MRFIT experiment, vital status has thus far been ascertainable for almost all of the participants. But the ascertainment of other morbid endpoints, such as non-fatal heart attacks, has been more difficult. Detection of these morbid events (by self-report or by evidence on periodic electrocardiograms) required subjects' returning for repeated checkups and examinations. At the end of the second year of the study, 6 percent of the Special Intervention group and 7.2 percent of the Usual Care group had missed their annual examinations. These proportions were 8 and 9 percent, respectively, by the fourth year. Among the SI participants, 16.3 percent had missed their triannual interim visits by the fourth year. The extent to which non-reporting subjects experienced a higher incidence of non-fatal morbid events is unclear.

It must be emphasized that subject attrition does not merely erode the statistical power of an experiment. Those who drop out may be least susceptible to the contemplated intervention. Certain imperfect covariates

of the decision to drop out can be measured. But any attempt to correct for unmeasured determinants requires a model of the distribution of these determinants. The interpretation of the experimental effect may then be very sensitive to unverifiable assumptions about the parametric form of such a model (Harris, 1981; Hausman and Wise, this volume). In micro-experiments, the only foolproof remedy for attrition bias is to keep subjects from dropping out altogether.

Hawthorne Effects and Anticipatory Responses

The subject's knowledge of his treatment assignment raises some serious problems for the MRFIT experiment. Although the Usual Care subject does not receive the benefits of group sessions, counseling, behavioral therapy and dietary instruction, he and his physician are informed of his risk status. Moreover, subjects in the UC group are asked, as in the SI group, to return for periodic visits and examinations. Highly motivated subjects who consent to randomization, but who end up in the UC group, may nevertheless alter their behavior. At the very least, this phenomenon will reduce the contrast between UC and SI interventions and diminish the power of the experiment.

Preliminary reports from MRFIT (Sherwin et al., 1979; Kuller et al., 1980; Schoenberger, 1981) in fact show improvements in risk factor scores for both SI and UC groups. After four years, SI men exhibited an 11 mm Hg drop in diastolic blood pressure, a 19 mg/dl drop in serum cholesterol, and a 41 percent smoking cessation rate. UC men showed a 6 mm Hg drop in diastolic blood pressure, an 11 mg/dl drop in serum cholesterol, and a 23 percent smoking cessation rate. Among SI men, 56 percent were being treated with

antihypertensive drugs, compared to 41 percent in the UC group. These improvements could reflect further regression toward the mean, or trends in behavior independent of the experiment.³ But the motivating effect of the experiment itself can hardly be excluded.

MRFIT experimenters recognize that many years may be required before the observed changes in risk factors are manifested in reduced CHD rates. In that case, the long-term mortality results will hinge critically on subjects' behavior after the termination of formal life-style intervention. Perhaps the UC men, who received dramatic attention only in the pre-experimental period and who were forced to take responsibility for their behavior from the start, will display greater long-run improvements. By contrast, if SI subjects become dependent upon the experiment itself, then discontinuation of formal intervention could lead to higher relapse rates (Syme, 1978).

The planners of the HIS have made special efforts to detect instrumentation artifacts and anticipatory responses (Newhouse et al., 1979). Participants' incentives to file insurance claims might depend on the amount of reimbursement. Hence, the plan assignment could affect subjects' reporting of medical care utilization. To avoid this interaction between treatment and measurement of response, a system of weekly reminders to file claims was used. But the reminders themselves were also found to affect reporting. Therefore, a subexperiment involving biweekly probes was instituted. Since intrusive questionnaires and health reports could also affect subject desires

³Initial cholesterol levels among all MRFIT randomized subjects, as well as dietary intake of cholesterol, total fat, and saturated fat, were already considerably below those observed in previous diet-heart studies.

to seek medical care, the sequence of examinations was similarly varied in a subexperiment. For the prepaid care group, moreover, a set of "controls on controls" was employed, with no instrumentation at all. To ascertain whether certain subjects would earmark the incentive payments solely for medical care, the schedule of incentive payments and bonuses was also varied. In order to detect possible anticipatory responses to the beginning and end of the study, the experimenters plan to follow the three-year intervention group for an additional two years. They also plan to be watchful of initial declines in price elasticity after the onset of the experiment, followed by increases in price sensitivity as the end of the experiment approaches, followed by post-experimental responses to intra-experimental price changes (Arrow, 1975).

It is difficult at this stage to see how all these instrumentation and anticipation artifacts can be estimated precisely. The issue here is not so much the separate, main effect of each form of instrumentation, but its interaction with treatment effects. There are too many interactions of instrumentation, treatment, and subject anticipations to test all of them satisfactorily. It is not completely clear how information on such artifacts can be easily incorporated into the final results.

Ethical Constraints

In the Multiple Risk Factor Intervention Trial, ethical considerations dictated that subjects with initial diastolic blood pressures above 114mm Hg be excluded from the study. Unfortunately, this form of sample truncation leads to difficulties similar to those encountered at the other end of the

risk factor scale. Thus, those individuals with previously undetected, severe hypertension may be derived from a population least motivated to seek routine care. These persons may have life styles or other unmeasured characteristics that counteract or reduce any salutary effects of risk factor reduction.

Even if a high-risk subject is eligible by screening criteria, ethical considerations dictate that treatment cannot be completely withheld. Hence, MRFIT does not compare treatment and nontreatment, but intensive intervention with "Usual Care." The Usual Care is not even average care, since the men randomized to the UC group have already undergone pre-experimental "treatment." Moreover, the planners of the experiment felt compelled to tell UC subjects that they were at high risk, including which risk factors were implicated (Kuller et al., 1980).

Interpretation of Treatment Effects

The design of MRFIT explicitly recognizes that people do not change their CHD risk factors one at a time. But its interpretation is still complicated by concomitant changes in dimensions of behavior other than the three risk factors. Subjects who are asked to change the saturated fat content of their diet may also be influenced to increase their physical activity, which may in turn affect cardiac status. Men involved in a smoking cessation group may alter their responses to stress, which could in turn affect cholesterol levels. Among SI subjects, in fact, nonsmokers and men who had quit smoking had the greatest improvements in serum cholesterol (Kuller et al., 1980, Table 8). This makes it difficult to assess whether

the effect of intervention resulted from changes in diet, serum cholesterol levels or other factors (Syme, 1978). Furthermore, the methods of life-style intervention may vary considerably across the 22 clinical centers in MRFIT. Within a specific MRFIT clinic, treatments are further adapted to the idiosyncracies of the experimental subject. Even if we regard Special Intervention as a homogeneous entity, Usual Care remains ill-defined. In the final analysis, if CHD rates improve with intervention in MRFIT, it may be difficult to know exactly what was responsible.

To be sure, one might attempt to elucidate the details of the experimental effect by specifying a response model. Thus, the Health Insurance Study was designed to estimate contrasts between the effects of different plans (e.g., the 95 percent coinsurance group versus the free care group, or the prepaid care group versus the remaining fee-for-service groups). But as early HIS data came in, the experimenters found the distribution of health care expenditures to be highly asymmetric, with a discrete atom at zero expenditures and a fat right-hand tail (Manning, Morris, Newhouse et al., 1981; Manning, Newhouse, and Ware, 1981). To perform statistical tests of treatment effects, they therefore proposed a multiple stage response model, involving the decision to seek care and expenditures conditional upon that decision. In addition to expenditures, health status was considered an important outcome measure. But health status could be both a determinant and a consequence of medical care utilization (Brook et al., 1979; Ware et al., 1979). These considerations led the experimenters to some interesting, but even more complicated structural models of the experimental response. No doubt with further structural

specifications, price elasticities and the parameters of response to deductibles and exclusions might also be estimated. I do not wish to denigrate these sophisticated efforts. But it should be pointed out that the conclusions derived from detailed response surface models may be very sensitive to the structural specification assumed by the analyst. As discussed in several other papers in this volume, such models are far removed from the classical ideal of the one-way analysis of variance.

Relevance of the Results to Policy Options

Even if MRFIT clearly demonstrates a reduction in CHD risk, its Special Intervention does not necessarily correspond to a viable policy option. For one thing, widespread intervention at the individual level is expensive. Although employment-based health and fitness programs have become more prevalent, they may be quite different from the specialized research environments of the MRFIT clinical centers. Moreover, changes in life style are likely to involve social learning, the diffusion of information, the changing of norms, and other phenomena that render individuals' responses interdependent. It is not clear that MRFIT captures these phenomena (Farquhar, 1978; Kasl, 1978; Syme, 1978). Finally, such micro-experiments reveal little about the effects of mobilizing voluntary health agencies, public restrictions on smoking, or the use of the mass media. Thus, MRFIT may reveal that CHD rates can be reversed. It may also offer some confirmation of the causal effects of risk factors. But it will offer much less information on the magnitudes of treatment effects in the general population. We could still be far from an operational public policy for

preventing coronary heart disease.

The Health Insurance Study was designed primarily to be a demand experiment. Except for comparative analysis of responses at sites with different supply conditions, no attempt was made to assess the supply response to an insurance-induced increase in demand. Nor were the market equilibrium effects of changes in coverage at issue. Yet the supply response to changes in insurance coverage is a critical factor in the recent rapid rise of health care expenditures in this country (Feldstein, 1977; Harris, 1979, 1980; Newhouse, 1978). Even after the HIS results are complete, policy-makers contemplating changes in insurance coverage will still be uncertain about the effects of reimbursement on hospital behavior, the consequences of insurance subsidy for technological change, or the effect of extensive insurance on competitive market discipline.

The HIS, to be sure, focuses to a great extent on ambulatory care demand. If the supply of ambulatory care were relatively elastic, and if the supply response of the ambulatory care sector were independent of the remainder of the health care sector, then the results of the experiment may offer a more complete picture of the ambulatory care market response. Even so, the behavior of the elderly population, who consume a substantial and growing fraction of health care costs, is not assessed in HIS. The decision to exclude the Medicare-eligible population from HIS was based on practical concerns about pre-experimental and experimental logistics. And a case can be made that an experiment on elderly responses to insurance ought to be designed very differently. But if young and old demand from the same suppliers, then changes in the coverage of the under-65 population could

affect the price and access to care of the elderly. What is more, the redistributive effects of changes in insurance may be quite different in the market than within the confines of the micro-experiment. At the very least, the proper application of the Health Insurance Study results to policy decisions necessitates the use of other non-experimental data.

3. POSSIBLE MACRO-EXPERIMENTAL REMEDIES

I now set forth the background of an illustrative macro-experiment.

The Stanford Heart Disease Prevention Program (SHDPP)

From 1972 to 1975, the Stanford Heart Disease Prevention Program (Farquhar, 1978; Farquhar et al., 1977; Meyer, Nash, McAlister et al., 1980; Stern et al., 1976) conducted a field experiment in three California communities, each with a population of approximately 15,000. The objective was to develop methods for modifying CHD risk that would be generally applicable to other community settings. Previous research had suggested that mass media campaigns directed at large populations could effectively transmit information, alter some attitudes, and produce small shifts in behavior, such as influencing consumer product choice. But the effect of the media on more complex behavior was poorly characterized.

The planners of SHDPP therefore attempted a factorial experiment, in which the combined effect of mass media and individualized intervention was assessed. From pre-experimental surveys in all three towns, they drew a subsample of men and women, aged 35 to 59, at high risk for CHD on the basis of cigarette smoking, blood pressure, and cholesterol level. In two towns

(Watsonville and Gilroy), an extensive media campaign was conducted. In Watsonville only, two-thirds of the high-risk subjects were randomly assigned to individualized intervention, while the remaining third served as the media-only control. In the third town (Tracy), no intervention was performed. Most of the reported results of this experiment have been derived from annual follow-up surveys of the sampled high-risk individuals in the three towns.

Since the trial was to be coordinated from a single research center, intervention was restricted only to three towns. Although the assignment to individualized intervention in Watsonville was performed randomly, the allocation of media-based treatments was non-random. Although the three towns were geographically isolated, the overlapping television signals of Watsonville and Tracy dictated that these two towns be assigned to media intervention.

Longitudinal versus Cross-Sectional Sampling

The planners of the SHDPP experiment, as I see it, made a serious but avoidable error of instrumentation: they relied upon longitudinal observations from a cohort of pre-experimentally screened, high-risk subjects. To be sure, changes in CHD mortality statistics in each community over four years might have been too small to distinguish a treatment effect. A longitudinal sample may have appeared most appropriate to ascertain non-fatal coronary events, as well as changes over time in behavior and knowledge of risk factors. Because media intervention was not randomly assigned, it may have seemed logical to use serial observations on many variables to

bolster the claim that an observed effect was causal. But reliance on a cohort of pre-experimentally screened subjects leaves the experimental results wide open to many of the criticisms of micro-experimentation, including selection artifacts, attrition biases, and Hawthorne effects.

Of the entire pre-experimental sample of 2151 subjects in the three towns, only 1204 actually completed all three follow-up surveys. The great fraction of those who failed to complete the study actively refused to participate or later dropped out (Stern et al., 1976, Table 1; Maccoby et al., 1977, Table 1). Among the 381 high-risk subjects who completed the baseline survey and who had not moved or died, 75 had dropped out after two years (Maccoby et al., 1977, Table 2). By three years, the attrition rates among eligible high-risk subjects varied from 22 to 33 percent of eligible subjects across towns (Meyer, Nash, McAlister et al., 1980, Table 2). The average dietary cholesterol and saturated fat intake, smoking prevalence and intensity, and systolic and diastolic blood pressures generally showed improvements over time in both experimental and control groups (Meyer et al., 1980, Table 4). After three years, the only striking finding was that the subjects given both media exposure and individualized instruction had quit smoking at a higher rate than the other groups. Relative weight and blood pressure showed no difference, while the differential changes in cholesterol were only suggestive. In view of these results, it is not unreasonable to suspect that the ultimate participants in SHDPP were highly motivated, that pre-experimental screening for high risks yielded an unrepresentative population, that subject attrition was biased, favoring a positive treatment effect, and that many subjects were aware of the presence of an experiment.

These difficulties, I maintain, should not be inherent to macro-experiments. Since the treatments are applied at the market or community level, there is no compelling reason why the responses in each unit should be obtained from a cohort. Sufficiently large, independent cross-section samples could be used to assess endpoints within each macro-unit. Since all of the residents in a community are subject to the same treatment, it matters little if different residents are sampled pre- and post-experimentally. Even in the case of certain morbid events of CHD, repeated cross-section samples of health care providers could serve as a reasonable substitute for longitudinal samples. To be sure, these procedures sacrifice precision. But they avoid the biases engendered by subjects' decisions to participate and remain in a cohort, as well as their awareness of participation in an experiment.⁴

It is arguable that this tradeoff between bias and precision does not differ from that encountered in micro-experimentation. Thus, the experimenter who does not screen on risk factors or other dependent variables sacrifices statistical power. Overcoming this loss of precision requires more subjects, which in turn increases the cost of the experiment. However, the cost of increasing the size of repeated cross-sectional surveys within communities may be far less than the cost of including additional subjects in a longitudinal micro-experiment, with all its follow-up interviews, diaries and logs.

⁴In the Stanford Five-City Project, the Pawtucket Heart Health Program, and the Minnesota Community Prevention Program, a mixture of cohort and cross-section sampling has apparently been used.

The advantage of repeated cross-section samples in macro-experiments is that individual subjects are less likely to be aware of the experiment. In fact it may be possible to perform blinded experiments, or at least blinded controls.⁵ Even if some subjects became aware of experimentation, their incentives to avoid or anticipate the treatment may be weaker than in a micro-experiment, where subjects can make decisions to participate separately from other economic choices. Thus, in a macro-experiment, an individual will have less incentive to leave a community merely to avoid certain media messages. So long as a different cross-section is sampled on each round, refusals to respond are much less severe a problem. Of course, it is possible for an entire community to be aware of the presence of the experiment. But it is hardly clear that this is so undesirable. If the institution of an experimental policy causes anticipatory emigration, or compensatory changes in local laws, or mass protests, that would appear to be a result worth knowing.

Repeated cross-section sampling in macro-experiments may further avoid ethical problems inherent in individual randomization. This is because the controls in a macro-experiment are "faceless," and the lives at stake are not specifically identified. To be sure, any subject found during sampling to be at high risk must still be informed of his condition and referred appropriately. However, so long as the experimenter samples from independent cross-sections, and so long as the samples are not large in comparison to the

⁵A blind control community is planned for the Pawtucket Heart Health Program.

population of the community, these ethical obligations should not materially affect the results. It is arguable that imposing involuntary participation on the citizens of a community is itself unethical (Hulley and Fortmann, 1980). But I do not see this objection as insurmountable.

Costs of Macro-Experimentation

Macro-experiments may incur lower costs of instrumentation. But the more difficult question is the costs of treatment. In a micro-experiment, only those individuals who are recruited and sampled undergo treatment. In a macro-experiment, everyone in a community receives the treatment, even if his experimental response is not measured.

Certain types of macro-experiments, such as those involving price subsidies in large communities, are undoubtedly very expensive. But in many instances macro-experimental intervention may exhibit significant economies of scale. This applies especially to the use of mass media in SHDPP and related experiments, where the marginal cost of exposing an additional person to a health message is near zero.

Relevance of Macro-Experimentation

Despite its flaws of instrumentation, the SHDPP media experiment had one salient advantage over clinical trials such as MRFIT. The experimental treatment--that is, the use of mass media to transmit health information, to alter preferences, and possibly to change behavior--corresponded to a genuine policy option. The micro-experiment may have revealed little about the social and behavioral mechanisms underlying the response to media

intervention (Leventhal et al., 1980). But the elucidation of mechanisms, I contend, should not be the objective of macro-experimentation. The main idea is to observe the effect of a contemplated policy in an experimental setting that closely approximates the environment in which the policy is to be applied.

The logical response, of course, is to ask whether the "black box" results of a macro-experiment are really relevant to the policy under consideration. Even if SHDPP and its progeny experiments should demonstrate an effect of media intervention on coronary risk factors and rates, how do we know that media intervention will succeed in other communities? To this and related questions I now turn.

4. MORE PROBLEMS WITH MACRO-EXPERIMENTS

The Confounding of Treatment Effects and Site Effects

The most serious difficulty with the Stanford three-city trial is the experimenters' misconception about the number of independent observations in their sample. In virtually every scientific report on this study, the authors assumed that the number of independent observations equalled the total number of sampled subjects in the three communities. This assumption would be valid if applied only to the Watsonville micro-experiment in which subjects were individually randomized. But for the mass media macro-experiments, there were really only three independent observations.

Confusion over the number of degrees of freedom in macro-experiments has been widespread. In fact, the issue appears to have been resolved, broached

all over, and then settled several times in the literature. Yet biostatisticians continue to propose formulas for appropriate sample size in community trials as if the individual were the unit of randomization (Gillum et al., 1980).

The confusion derives in part from the view that outcome measurement in community prevention trials is merely a form of cluster sampling (Cornfield, 1978; Gillum et al., 1980). If the experimenter wishes to estimate, say, CHD death rates, then sampling by community, rather than by individuals, will increase the variance of estimated population rates. The increase in variance would be inversely related to the degree of homogeneity of death rates within communities and directly related to the extent of heterogeneity between communities. Hence, if the experimenter could select relatively homogeneous intervention sites, the loss of efficiency would appear to be minimal. But this view ignores the fact that an experiment has been conducted and must be interpreted. The real issue is that in the interpretation of the results, the "site effects" are confounded with the "treatment effects."

Consider the following example. Suppose that community A is chosen for a media campaign and community B is selected as control. Suppose further that we could randomly allocate N subjects each to live in these two towns. Each subject, it is assumed, belongs to a homogeneous population with respect to pre-experimental risk of CHD. How should we interpret the results of the media campaign? If we believed that the two communities were merely artificial vessels for separating experimental from control groups, and that within each community there was no intercorrelation of subject responses,

then we have $2N$ observations on the two treatments. But if the billboard density in a community affects the frequency of messages, or the ideology of the local television station owner affects the prominence of health-related commercials, or if the configuration of voluntary agencies affects opinion leadership, or if social networks permit greater diffusion of information, or if subjects' responses depend on their conformity with others, or if subjects' changes in dietary habits depend on food prices in a community, then we no longer have $2N$ independent observations. Even if we could randomly assign subjects to communities A and B, the results could be quite different if town B were instead chosen for intervention and town A were instead chosen for the control. Moreover, it would not help to assess the pre-experimental variance of death rates between and within communities. By construction, these variances would all be zero. The issue is not pre-experimental death rates, but the responses of death rates to the intervention.

To be sure, site effects are common in micro-experiments, such as MRFIT, where the size of the experiment dictates the deployment of multiple clinical centers. But the situation in micro-experiments is considerably different because randomization of subjects takes place within each site. Hence, site effects can be distinguished from treatment effects and site-treatment interactions can be tested.

The literature on clinical trials is replete with tests of site effects and site-treatment interactions (e.g., hospital effects in the National Halothane Study, clinical center effects in the University Group Diabetes trial of insulin versus oral hypoglycemic agents). Hopefully, in the

analysis of the final results of MRFIT, treatment successes at particular clinical centers will receive scrutiny. But in pure macro-experiments, there is no cross-over of treatments within a community. The site effects are fully nested within the treatments. Sampling more subjects at each site will diminish the variance of the estimated death rate within each site. But it will not affect the precision of these site-treatment interactions. In fact, if we have only two treatments and two sites, there are no degrees of freedom to disentangle these treatment-site interactions. Only more sites will solve this difficulty.

External Validity

When the experimenter tests for site-treatment interactions, he is asking whether any specific characteristic of a market or community could be uniquely responsible for, say, an observed effect of media campaigns. If he samples enough communities, he can distinguish between a general media effect, applicable to all sites, and media effects that are merely idiosyncratic for certain communities. But then how does the experimenter know that the selected sites constitute a representative sample of these idiosyncracies? What would be the effect of media intervention in communities where a single, large employer also started his own employee health program, or where a national manufacturer test-marketed a new, low-cholesterol product? If relatively small towns were selected, as in SHDPP, what would the results tell us about the effects of intervention in large cities? Would they be relevant to macro-experiments on work groups or domiciliary institutions (Rose et al., 1980; Sherwin, 1978; WHO European Collaborative Group, 1974)?

So long as the site-treatment interactions are regarded as random effects, the experimenter is obligated to choose judiciously experimental sites that are representative of the environment in which the policy is to be instituted. I recognize that even in macro-experiments, one ought to select sites that are not wholly unrepresentative. It is thus worth inquiring whether the communities selected for HIS possess doctors, hospitals, medical standards and institutions that are typical of the United States. And I have already inquired whether the clinical centers in MRFIT are representative of programs of individualized intervention throughout the country. But it seems to me that the burden on macro-experiments is much greater.

Randomization of Macro-Units

Many of the proponents of community-based intervention trials regard randomization as an impractical ideal. There are just too many administrative and political obstacles. Unfortunately, I see virtually no way out of the requirement that experimental sites, once selected, must be allocated randomly to treatments. I acknowledge numerous instances where evidence from nonrandomized studies has proved convincing. But in those cases, the analysis has hinged on a paucity of plausible rival explanations for the observed difference between treatment and control groups (Campbell and Stanley, 1966). But in macro-experimentation, there are likely to be an abundance of rival explanations. It is not hard to imagine that a town with its own television station or health-conscious opinion leaders will be more willing to undergo a media campaign. Such a community may be more susceptible to the effects of such an intervention.

5. TOWARD A SCIENCE OF MACRO-EXPERIMENTATION

Despite substantial advances in design, execution and interpretation, micro-experiments still have serious and possibly inherent difficulties. Individuals make non-random decisions to participate or drop out of the experiment. They may be influenced by the instrumentation process. Even in the absence of these difficulties, micro-experiments do not necessarily test real policy options. Macro-experimentation, on the other hand, may avoid some of these problems. But convincing macro-experiments require many observations at the community or market level. Moreover, political and administrative factors may dictate non-random selection of communities, with its attendant difficulties. And there is always uncertainty whether the observed effect of treatment in a sample of communities was not due to idiosyncratic, unrepresentative characteristics of the experimental sites.

We are thus faced with a serious dilemma. Should we perform a micro-experiment, optimistic that instrumentation artifacts will not arise, and thankful to learn something about one aspect of a complicated policy problem? Or should we plunge ahead with a "sloppy" macro-experiment, with all of its difficulties of interpretation and generalization?

Decentralized Macro-Experiments

Because SHDPP was to be coordinated by a single research center, the experiment was restricted to only three towns. Once these three were selected, random assignment to media exposure was made impossible by overlapping television signals. But it is worth speculating what experimental design might have arisen from a multi-center trial. If the

Stanford group had been one of many research centers, couldn't they have selected a pair of towns, both of which had non-overlapping television signals? Why couldn't treatment be randomly assigned between the two towns? Why couldn't the Stanford city-pair be one block in a larger matched pair experiment?

My point here is that many of the most serious difficulties of macro-experiments may result from over-centralization. So long as we could allocate pairs of comparable sites (or perhaps larger subsets) to individual experimental blocks, the execution of each block could be the responsibility of a separate research center. Within each block, randomization may be more feasible. Increasing the statistical power of the experiment, and perhaps its external validity, means increasing the number of blocks.

Such a design is not entirely speculative. In fact, the WHO European Collaborative Group (1974; Rose et al., 1980) has been conducting a macro-experiment in CHD prevention in 12 pairs of factories in various cities. These factories (or in some cases occupational units within factories) were recruited into the trial before random assignment to treatment or control. The factory pairs were matched as far as possible by age, geographical area and the nature of the industry. The subjects include all male employees aged 40 to 59 years, and not merely those at high risk. This design unfortunately involves longitudinal follow-up of cohorts. Hence, it may be susceptible to participation biases, selective employee turnover, and Hawthorne effects. But it illustrates the possibility of randomization within blocked pairs of macro-units.

One might object that only small units, such as factories and

domiciliary institutions, are susceptible to randomization (Sherwin, 1978). Larger political entities will merely balk at the uncertain prospect of receiving the less desirable assignment. But it is hardly clear to me that this state of affairs is inevitable. For one thing, the possibility of randomization among matched pairs may be more palatable politically than random drawings from a larger population of sites. In some cases where the eligible sites are political subdivisions under the governance of a higher authority, the possibility of site self-selection may not be so serious. In fact, several macro-experiments in cancer screening, in which census tracts, townships, or counties are the relevant sites, have already been proposed (Apostolides and Henderson, 1977). Moreover, in cases where communities or organizations have already received some type of government grant or benefit, the continued receipt of that benefit could be made the incentive for participation in the experiment. In cases where various communities apply for grants to become demonstration sites for a particular innovation, the awards process could be broken down into two stages. A subset of deserving, eligible sites would first be chosen. Among eligible sites, treatment and control assignments could then be made. It is remarkable to me how often government agencies and other grantors first make the awards to the most deserving sites and then ponder how a comparable set of control sites is to be chosen from the losers for the purpose of project evaluation.

When intervention at a large number of sites is managed by one research or administrative group, the inevitable consequence is a rationing of limited intervention effort to a few sites. In extreme cases, many of the so-called intervention sites do not receive any intervention because the research team

has merely lost control of the project. Administrative decentralization of macro-experiments could allay some of these problems. Moreover, some degree of blinding may be possible. At the least, a research team responsible for intervention in one block of sites need not know the progress of the experiment in other blocks.

Time-Series Experiments and Cross-Over Designs

The possibility that communities or other macro-units could serve as their own controls has not been adequately explored. Admittedly, any comparison over time is susceptible to confounding interpretations. Experimental responses take some time to be completed. What appears to be the effect of a cross-over may actually be a transient from earlier intervention (Morris, Newhouse, and Archibald, 1980). If the macro-experiment is not blinded, then the effects of cross-over could be confused with anticipatory responses or other Hawthorne effects. Nevertheless, there is a variety of familiar devices for detecting time-varying responses. Although these devices have been derived from micro-experiments, they could at least be tried in the macro-setting.

For example, in the case of a matched pair design, the treatment and control communities could reverse their assignments later in the experiment. The timing of this reversal need not be scheduled in advance, or at least known to the experimental units. Stopping short of complete cross-over, I could also envisage folding back designs. We could begin by a series of observations on communities in which no intervention is instituted. Thereafter, one or more of the communities becomes a treatment site. In

sequence, the remaining communities receive the intervention. Again, the sequence and schedule of assignment could be random and unknown to the experimental units. If all of the units are destined ultimately to receive the intervention, randomization with respect to the sequence and timing of the intervention may not present so many political or administrative obstacles.

Mixed Macro and Micro Designs

In some cases, a mixture of micro and macro designs might enhance the power of the experiment. Such cases arise when the interventions at the individual and site levels are qualitatively similar.

In the SHDPP trial, a subexperiment of individual intervention was performed within Watsonville, a town receiving media intervention. This subexperiment was designed to test the interaction between the two types of experimental treatments. Unfortunately, the investigators failed to conduct an identical subexperiment in Tracy, the town receiving no media intervention. But even if a full factorial design had been undertaken, the two types of treatment were so qualitatively different that only their crude interaction could be profitably investigated.

But in other cases, both interventions could be close enough to conform to a simple response model. Suppose, for example, that the experimenter wishes to investigate the effects of varying employer contributions to employee health insurance premiums. Since changes in employee benefits are typically performed at the level of the firm, a macro-experiment would be appropriate, with various firms corresponding to different macro sites. But

within each firm, employer contributions could be further varied among employees. Such an experiment could offer considerable insight into firm-specific and employee-specific responses to changes in employee premium subsidies.

Combining Macro-Experiments

A potential significant advantage of macro-experimental blocking is its ability to enhance the external validity of the experiment. Within each block, experimental sites might possess similar characteristics. But between blocks the site characteristics could vary considerably. In community-based life-style intervention, it would be especially informative for blocks to vary with respect to the size, climate, age structure, sex, racial and ethnic composition of their member communities.

A number of independent community-based life-style intervention trials are already in progress in this country. Taken together, these trials might be considered a single macro-experiment with multiple blocks. The difficulty with this interpretation, however, is that the method of intervention may vary considerably from one block to the next. We thus cannot easily distinguish between a block effect and a block-treatment interaction. If some community trials show significant effects of life-style intervention and others do not, it will be unclear whether the discrepancies resulted from differences in the type of media intervention across trials, or differences in the susceptibility of communities to media messages. The results of different trials could be combined only if we had some prior information on the relationship between types of media intervention employed.

Some recent theoretical work on combining diverse experiments might be usefully applied to this problem (DuMouchel and Harris, 1981). A complete exposition is necessarily beyond the scope of the present paper. But the main idea is to specify formally a structural relationship between the treatment effects in each community trial. For example, the magnitude of the effect on CHD rates might depend on the extent of electronic media intervention, the duration of intervention, or the recruitment of voluntary agencies. A model of the treatment effect that relates these characteristics is then superimposed upon the results of each trial. The main issue in the application of such a technique is the degree to which life-style intervention in each trial was independent of the characteristics of the communities under observation. For example, if the experimenters in a particular trial resorted to scientifically-oriented media messages because the target communities were highly educated, it may be impossible to distinguish between the treatment effect of media content and the role of educational background in a community's response.

Competition Experiments, Regulation Experiments and Deregulation Experiments

Reduction of the tax subsidy on health insurance coverage, elimination of barriers to entry for prepaid health care providers, and enhancement of consumer choice of health insurance plans have been proposed to control rising health care expenditures. Virtually all of the evidence supporting the efficacy of these interventions is non-experimental. Our policy-makers could, of course, take the available data as sufficient cause to plunge ahead with a full-scale policy. But the correct course, it seems to me, is to

assess some of these innovations experimentally before taking such drastic action. I have already hinted how several large employers in a number of different cities might serve as sites for experimental changes in employee health insurance benefits. Perhaps several distinct divisions of the same large corporation could form an experimental block. Community-based experiments, in which the effects on market competition are observed, are also conceivable.

Regulatory controls on health care expenditures have also been suggested. Although various innovative forms of hospital reimbursement have been tried, most of the so-called reimbursement experiments have really been uncontrolled demonstration projects. In view of the substantial likelihood that hospitals subject to those novel controls have been selected in a biased manner, it is hard to know exactly what significance these projects should have for future policy decisions. It is difficult for me to see why the experimenters have not blocked participating hospitals according to, say, size, teaching status, or range of facilities, and then randomly assigned the novel form of reimbursement within each block.

One variant of the fold-back design discussed above is the deregulation experiment. In this case, the experimental treatment is the removal of an intervention already in place. The sequence and timing of deregulation at various sites is the critical control variable. This type of design may be particularly useful when the value of a regulatory program is in question. Even if our policy-makers deem that physician peer review schemes or health planning agencies are to be discontinued, it would be valuable to learn something about the effects of these policies during their demise.

6. CONCLUSIONS

This paper can be easily criticized for its lack of balance. I have sought out the most subtle crack in micro-experiments. Yet I am willing to cover large faults in macro-experiments with hopeful speculation.

The plain truth is that macro-experiments in public policy--or at least corrupted versions of macro-experiments--are far more prevalent than the micro-experiments to which social scientists have devoted so much attention. It is not too soon to develop some meaningful strategies for effective macro-experimentation.

REFERENCES

- Arrow, Kenneth J. (1975). Two notes on inferring long-run behavior from social experiments. Rand Report P-5546. Santa Monica: Rand Corporation.
- Apostolides, Aristide and Henderson, Maureen (1977). Evaluation of cancer screening programs: parallels with clinical trials. Cancer 39: 1179-1785.
- Brook, Robert H., Ware, John E., Jr., Davies-Avery, Allyson, et al. (1979). Conceptualization and measurement of health for adults in the health insurance study: vol. VII, overview. Report R-1987/8-HEW. Santa Monica: Rand Corporation.
- Campbell, Donald T. and Stanley, Julian C. (1966). Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally Publishing Co.
- Cornfield, Jerome (1978). Randomization by group: a formal analysis. American Journal of Epidemiology 108: 100-102.
- Davis, C.E. and Havlik, R.J. (1977). Clinical trials of lipid lowering and coronary artery disease prevention. In B.M. Rifkind and R.I. Levy, eds., Hyperlipidemia: Diagnosis and Therapy. New York: Grune and Stratton.
- DuMouchel, William H. and Harris, Jeffrey E. (1981). Bayes and empirical Bayes methods for combining cancer experiments in man and other species (with Discussion). Journal of the American Statistical Association, in press.

- Farquhar, John W. (1978). The community-based model of life style intervention trials. American Journal of Epidemiology 108: 103-111.
- Farquhar, John W., Maccoby, Nathan, Wood, Peter D., et al. (1977). Community education for cardiovascular health. Lancet 1: 1192-1195.
- Feldstein, Martin S. (1977). Quality change and the demand for hospital care. Econometrica 45: 1681-1702.
- Friedman, Gary D., Petitti, Diana B., Bawol, Richard D., and Siegelau, A.B. (1981). Mortality in cigarette smokers and quitters. New England Journal of Medicine 304: 1407-1410.
- Gillum, Richard F., Williams, Paul T., and Sondik, Edward (1980). Some considerations for the planning of total-community prevention trials--when is sample size adequate? Journal of Community Health 5: 270-278.
- Harris, Jeffrey E. (1979). The aggregate coinsurance rate and the supply of innovations in the hospital sector. Cambridge, Mass.: Massachusetts Institute of Technology, Department of Economics Working Paper.
- Harris, Jeffrey E. (1980). Commentary. In M. Pauly, ed., National Health Insurance: What Now? What Later? What Never? Washington: American Enterprise Institute.
- Harris, Jeffrey E. (1981). Prenatal medical care and infant mortality. In V. Fuchs, ed., Economic Aspects of Health. Chicago: University of Chicago Press, forthcoming.
- Hausman, Jerry A. and Wise, David A. (1982). Technical problems in social experimentation: cost versus ease of analysis. In J.A. Hausman and D.A. Wise,, eds., Social Experimentation. Chicago: University of Chicago Press, forthcoming.

- Hulley, Stephen B. and Fortman, Stephen F. (1980). Clinical trials of changing behavior to prevent cardiovascular disease. In S.M. Weiss, ed., Perspectives in Behavioral Medicine. New York: Academic Press.
- Hypertension Detection and Follow-up Program Cooperative Group (1979a). Five-year findings of the hypertension detection follow-up program: I. Reduction in mortality of persons with high blood pressure, including mild hypertension. Journal of the American Medical Association 242: 2562-2571.
- Hypertension Detection and Follow-up Program Cooperative Group (1979b). Five-year findings of the hypertension detection follow-up program. II. Mortality by race, sex and age. Journal of the American Medical Association 242: 2572-2577.
- Kasl, Stanislav V. (1978). A social-psychological perspective on successful community control of high blood pressure: a review. Journal of Behavioral Medicine 1: 347-381.
- Kasl, Stanislav V. (1980). Cardiovascular risk reduction in a community setting: some comments. Journal of Consulting and Clinical Psychology 48: 143-149.
- Kuller, Lewis, Neaton, James, Caggiula, Arlene, and Falvo-Gerard, Lorita (1980). Primary prevention of heart attacks: the multiple risk factor intervention trial. American Journal of Epidemiology 112: 185-199.
- Leventhal, Howard, Safer, Martin A., Cleary, Paul D., and Gutman, Mary (1980). Cardiovascular risk modification by community-based programs for life style change: comments on the Stanford study. Journal of Consulting and Clinical Psychology 48: 150-158.

Maccoby, Nathan, Farquhar, John W., Wood, Peter D., and Alexander, Janet (1977). Reducing the risk of cardiovascular disease. Effects of a community-based campaign on knowledge and behavior. Journal of Community Health 3: 100-114.

Manning, Willard G., Jr., Morris, Carl N., Newhouse, Joseph P., et al. (1981). A two part model of the demand for medical care: preliminary results from the Health Insurance Study. Proceedings of the World Congress on Health Economics, forthcoming.

Manning, Willard G., Jr., Newhouse, Joseph P., and Ware, John E., Jr. (1981). The status of health in demand estimation: beyond excellent, good, fair, and poor. In V. Fuchs, ed., Economic Aspects of Health. Chicago: University of Chicago Press, forthcoming.

Meyer, Anthony J., Nash, Joyce D., McAlister, Alfred L., Maccoby, Nathan, and Farquhar, John W. (1980). Skills training in a cardiovascular health education campaign. Journal of Consulting and Clinical Psychology 48: 129-142.

Morris, Carl (1979). A finite selection model for experimental design of the Health Insurance Study. Journal of Econometrics 11: 43-61.

Morris, Carl N., Newhouse, Joseph P., and Archibald, Rae W. (1980). On the theory and practice of obtaining unbiased and efficient samples in social surveys. Rand R-2173-HEW. Santa Monica: Rand Corporation.

Mosteller, Fred, and Mosteller, Gail (1979). New statistical methods in public policy. Part I: experimentation. Journal of Contemporary Business 8: 79-92.

- Multiple Risk Factor Intervention Trial Group (1976a). The multiple risk factor intervention trial (MRFIT). Journal of the American Medical Association 235: 825-827.
- Multiple Risk Factor Intervention Trial Group (1976b). The multiple risk factor intervention trial. Annals of the New York Academy of Medicine 304: 293-308.
- Multiple Risk Factor Intervention Trial Group (1977). Statistical design considerations in the NHLBI multiple risk factor intervention trial. Journal of Chronic Diseases 30: 261-275.
- Newhouse, Joseph J. (1974). A design for a health insurance experiment. Inquiry 2: 5-27.
- Newhouse, Joseph P. (1978). The erosion of the medical marketplace. Rand Report R-2141. Santa Monica: Rand Corporation.
- Newhouse, Joseph P., Marquis, Kent H., Morris, Carl N., Phelps, Charles E., and Rogers, William H. (1979). Measurement issues in the second generation of social experiments: the Health Insurance Study. Journal of Econometrics 11: 117-129.
- Puska, P., Tuomilehto, J., Nissinen, A. et al. (1978). Changing the cardiovascular risk in an entire community: the North Karelia project. Paper presented at the International Symposium on Primary Prevention in Early Childhood of Atherosclerotic and Hypertensive Diseases. Chicago.
- Rivlin, Alice (1974). Allocating resources for policy research. How can experiments be more useful? American Economic Review Papers and Proceedings 64: 346-354.

- Rose, Geoffrey, and Hamilton, R.J.S. (1978). A randomised controlled trial of the effect of middle-aged men of advice to stop smoking. Journal of Epidemiology and Community Health 32: 275-281.
- Rose, Geoffrey, Heller, R.F., Pedoe, Hugh T., and D.G.S. Christie (1980). Heart disease prevention project: an randomised controlled trial in industry. British Medical Journal 280: 747-751.
- Schoenberger, James A. (1981). The Multiple Risk Factor Intervention Trial. Presentation at American Heart Association meetings, Washington, D.C.
- Sherwin, Roger (1978). Controlled trials of the diet-heart hypothesis: some comments on the experimental unit. American Journal of Epidemiology 108: 92-99.
- Sherwin, Roger, Sexton, Mary, and Dischinger, Patricia (1979). The Multiple Risk Factor Intervention Trial of the primary prevention of coronary heart disease: risk factor changes after two years. Paper presented at the VII Asian Pacific Congress of Cardiology, Bangkok.
- Stern, Michael P., Farquhar, John W., Maccoby, Nathan, and Russell, Susan H. (1976). Results of a two-year health education campaign on dietary behavior: the Stanford three community study. Circulation 54: 826-833.
- Stolley, Paul D. (1980). Epidemiologic studies of coronary heart disease: two approaches. American Journal of Epidemiology 112: 217-224.
- Syme, S. Leonard (1978). Life style intervention in clinic-based trials. American Journal of Epidemiology 108: 87-91.
- Truett, J., Cornfield J., and Kannel, W. (1967). Multivariate analysis of the risk of coronary heart disease in Framingham. Journal of Chronic Diseases 20: 511-524.

Ware, John E., Jr., Brook, Robert H., Davies-Avery, Allyson et al. (1980).

Conceptualization and measurement of health for adults in the health insurance study: vol. I, model of health and methodology. Santa Monica: Rand Corporation Report R-1987/1-HEW.

WHO European Collaborative Group (1974). An international controlled trial in the multifactorial prevention of coronary heart disease.

International Journal of Epidemiology 3: 219-224.

