

The Effect of Coinsurance on the Health of Adults

Results from the Rand Health Insurance Experiment

Robert H. Brook, John E. Ware, Jr.,
William H. Rogers, Emmett B. Keeler,
Allyson R. Davies, Cathy A. Sherbourne,
George A. Goldberg, Kathleen N. Lohr,
Patricia Camp, Joseph P. Newhouse

Rand

HEALTH INSURANCE EXPERIMENT SERIES

The research described in this report was supported by the U.S. Department of Health and Human Services, under Grant No. 016B-80.

Library of Congress Cataloging in Publication Data
Main entry under title:

The Effect of coinsurance on the health of adults.

"R-3055-HHS."

Bibliography: p.

1. Rand Health Insurance Experiment. 2. Health surveys—United States. 3. Medical care—United States—Utilization. 4. Medical care, Cost of—United States. 5. Insurance, Health—United States—Coinsurance. I. Brook, Robert H. (Robert Henry), 1943- II. United States. Dept. of Health and Human Services. III. Rand Corporation. IV. Rand Health Insurance Experiment. [DNLM: 1. Deductibles and Coinsurance. 2. Health. 3. Health Surveys—United States. 4. Insurance, Health—United States. W 275 AA1 E25] RA410.53.E34 1984 362.1'042 84-24982
ISBN 0-8330-0614-2

The Rand Publication Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

R-3055-HHS

The Effect of Coinsurance on the Health of Adults

Results from the Rand Health Insurance Experiment

Robert H. Brook, John E. Ware, Jr.,
William H. Rogers, Emmett B. Keeler,
Allyson R. Davies, Cathy D. Sherbourne,
George A. Goldberg, Kathleen N. Lohr,
Patricia Camp, Joseph P. Newhouse

December 1984

Prepared under a grant from the
U.S. Department of Health and Human Services



PREFACE

This report presents experimental results from a study of the health status of adults who participated in the fee-for-service portion of the Rand Health Insurance Experiment (HIE). The HIE was a large-scale, controlled trial in health care financing that ran in the field from November 1974 through January 1982. The project, which began in 1971 with a research grant from the Office of Economic Opportunity, has been funded since 1973 under a grant from the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health, Education, and Welfare (now Health and Human Services).

The original design of the experiment was described in "A Design for a Health Insurance Experiment," *Inquiry*, March, 1974, by Joseph P. Newhouse, an article that is also available from Rand as Report R-965-1-OEO. More recently, the experiment's study design was summarized in "Some Interim Results from a Controlled Trial of Cost Sharing in Health Insurance," *The New England Journal of Medicine*, December 17, 1981, by Joseph P. Newhouse and colleagues (also available from Rand as R-2847-HHS).

An earlier version of the present report appeared in the December 8, 1983, issue of *The New England Journal of Medicine* (Vol. 309, pp. 1426-1434). The report contains several appendixes describing the main analyses in more detail and various subsamples of the adult population. In addition, the intervening months have permitted the authors to increase moderately the sample for certain analyses, which has changed some results slightly; these new findings are given here.

SUMMARY

METHODS OF THE EXPERIMENT

Sample

The Rand Health Insurance Experiment was carried out to answer the question (among others), "Does free medical care lead to better health than insurance plans that require the patient to shoulder part of the cost?" It ran from November 1974 through January 1982 in six sites (Dayton, Ohio; Seattle, Washington; Fitchburg and Franklin County, Massachusetts; and Charleston and Georgetown County, South Carolina). Altogether, 3958 individuals between the ages of 14 and 61 were enrolled, they belonged to 2005 families and were free of disability serious enough to have made them eligible for the Medicare program (i.e., public Disability Insurance).

Insurance Plans and Benefits

Families were randomly assigned to one of several insurance plans for three (70 percent) or five years. One plan provided *free* care; the others required that enrollees pay a share of their health bills. On the *individual deductible* plan, the family paid 95 percent of the cost of all outpatient care up to an annual out-of-pocket expenditure of \$150 per person (\$450 per family); all outpatient care beyond that amount, as well as all inpatient care, was free. On the *intermediate coinsurance* plans, the family paid 25 or 50 percent of all health bills each year, inpatient and outpatient, until it had spent 5, 10, or 15 percent of its income or \$1000 (whichever was less). On the *income-related catastrophic* plans, the family paid 95 percent of its health bills up to the same ceiling as in the intermediate plans. All plans covered ambulatory and hospital care, preventive services, most dental services other than orthodontia, psychiatric and psychological services up to 52 visits per person per year, prescription drugs, and nonprescription drugs for certain chronic conditions.

Health Status Measures

We developed or adapted several measures to evaluate the effect of cost-sharing on health in four distinct categories: general health,

health habits, physiologic health, and the risk of dying from any cause. This monograph reports on 11 measures from these categories: physical health, role functioning, mental health, social contacts, and general health ratings (collectively referred to as the general health measures); smoking behavior; weight; cholesterol level; diastolic blood pressure level; visual acuity; and an index of the risk of dying related to specific risk factors (systolic blood pressure, cholesterol, and smoking habits).

Plan of Analysis

For these analyses, we adopted a three-step analytic strategy. We first identified important variables that might explain differences in the dependent health status measures just named, such as the family's experimental insurance plan, family income adjusted for size and composition of the family, and initial state of health. Then we used regression methods to estimate the influence of these explanatory variables on the various measures of health status at the experiment's end. To interpret these results, we employed the results of the regression equations to predict the exit health status of people who had a given set of entry characteristics, including the "average" participant and participants in certain subgroups that differed by income and by level of initial health.

Possible Artifacts and Biases

We also undertook extensive analyses to detect and counteract any problems that might lead to biased estimates or erroneous inferences. For example, we compared health status values at enrollment (a) for participants in each plan and (b) for persons who accepted the offer to enroll in the experiment versus those who did not. By including initial health status values in the regression analyses, we also controlled for any effect of nonrandom composition of the sample. Further, we collected longitudinal information on several measures for persons who had left the experiment prematurely and included information about most of these dropouts in the analyses. Finally, we imputed scores on certain variables to persons with missing enrollment data and included them in the analysis.

RESULTS

Threats to Validity

Steps taken to overcome possible problems of bias or erroneous inference were successful. Our analyses showed that different acceptance rates were unlikely to have affected any conclusions; the same was true of differences among plans in retention of participants until the end of the study. In addition, enrollment health status values in the actual sample used for each analysis did not differ by plan.

Effects on Health Status

For the *average* person enrolled in the experiment, we observed two significant positive effects of free care relative to cost-sharing: corrected far vision (i.e., when the enrollee was wearing his or her usual glasses or contact lenses) was better by 0.1 Snellen lines ($p = 0.001$) and diastolic blood pressure was lower by 0.8 mm Hg ($p = 0.03$). For the remaining measures, confidence limits for the differences between the free and cost-sharing plans were sufficiently narrow to conclude that, for the average participant, any true differences would be clinically and socially negligible.

For the five general health measures, we could detect no significant positive effect of free care for persons who differed by income (high versus low income) and by initial health status (good versus ill health). Because the confidence intervals around differences between free and cost-sharing plans were larger in these subgroup analyses, however, we could be less certain that *clinically* important effects did not occur.

Among participants who were judged to be at elevated risk with respect to smoking habits, cholesterol levels, and weight, free care had no detectable effect.

Among individuals whose uncorrected far vision was worse than 20/20, corrected far vision was 0.2 Snellen lines better ($p = 0.001$) on the free than on the other plans. For persons who were in the upper quartile of the distribution of risk factors included in the risk of dying index, the risk of dying was 10 percent lower on the free than the cost-sharing plans ($p = 0.02$). Improvements in vision and the risk of dying, as well as in blood pressure, were largest among persons with low incomes. Poor people at elevated risk apparently benefited from receiving free care, although we cannot draw any firm conclusion about persons of higher incomes at elevated risk.

CONCLUSIONS

We drew three main conclusions about the influence of free care on the health status of HIE adults. First, free care did not affect the major health habits associated with cardiovascular disease and many types of cancer, even though those habits (especially smoking) were at levels where substantial health benefit from behavior change was possible. Second, free care had at most a small effect on any of five general health measures for the average enrollee. Confidence intervals were wider for subgroups of persons with low income or initially in poor health; therefore, we cannot rule out clinically meaningful changes in particular subgroups. Third, people having specific conditions with well-established diagnostic and therapeutic procedures (myopia, hypertension) benefited from free care, and these improvements appeared to be greater among the poor.

ACKNOWLEDGMENTS

Many individuals participated in the work reported here, and space does not permit separate acknowledgment of them all. We are indebted to the following persons for their unique contributions in particular areas: Carolyn André, Rae Archibald, Marie Brown, Maureen Carney, Lorraine Clasquin, and Kenneth Krug (administration); Anita Stewart (measurement of physical functioning, smoking, and weight); Randi Rubenstein (measurement of vision impairment); Janet Hanley (programming); Darlene Blake, Carol Edwards, Joan Keesey, Bryant Mori, Susan Polich, Martin Seda, David Stewart, and Beatrice Yor-mark (data processing); and Barbara Eubank and Marilyn Martino (secretarial assistance).

The survey data were collected by the National Opinion Research Center and by Mathematica, Inc.; the physiologic data were collected by American Health Profiles and CompuHealth. The sample was maintained by Glen Slaughter and Associates as part of their task of processing claims data.

We are grateful as well to James Schuttinga and Larry Orr for support and guidance in their capacities as project officers from the Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services, and to their superiors through the years whose support made this endeavor possible.

Robert Kane provided a thoughtful and constructive critique of this report and its many appendixes, for which we thank him.

CONTENTS

PREFACE	iii
SUMMARY	v
ACKNOWLEDGMENTS	ix
TABLES	xiii
Chapter	
1. INTRODUCTION	1
2. EXPERIMENTAL DESIGN AND ANALYTIC METHODS .	3
Sample and Sites	3
Insurance Plans and Benefits	4
Health Status Variables	6
Methods of Analysis	7
Possible Artifacts and Biases	12
3. RESULTS OF DESIGN AND HEALTH STATUS	
ANALYSES	14
Threats to Validity	14
Effects on Health Status	18
4. DISCUSSION OF THE HEALTH STATUS RESULTS	25
Appendix	
A. SCHEDULE OF BENEFITS IN THE FAMILY HEALTH	
PROTECTION PLAN	29
B. HEALTH STATUS VARIABLES MEASURED IN THE	
HEALTH INSURANCE EXPERIMENT	34
C. DESCRIPTION OF GENERAL HEALTH MEASURES	43
D. DESCRIPTION OF PHYSIOLOGIC AND HEALTH	
HABITS MEASURES	53
E. DETAILED RESULTS OF REGRESSION EQUATIONS	
FOR GENERAL HEALTH MEASURES AND SMOKING ..	61
F. DETAILED RESULTS OF REGRESSION EQUATIONS	
FOR CHOLESTEROL, WEIGHT, VISION, BLOOD	
PRESSURE, AND RISK OF DYING	68
G. SENSITIVITY OF RESULTS TO THE SIZE OF THE	
ELEVATED-RISK GROUP	78

H. RESULTS OF SPECIAL ANALYSES ON THE EFFECT OF SITE	82
I. SAMPLE LOSS FOR PHYSIOLOGIC MEASURES, CHOLESTEROL, AND WEIGHT	86
J. RESULTS OF SPECIAL ANALYSES ON EFFECTS OF MISSING DATA: PHYSIOLOGIC MEASURES AND CHOLESTEROL	89
K. RESULTS OF SPECIAL ANALYSES ON EFFECTS OF MISSING DATA: HEALTH PERCEPTIONS	95
L. RESULTS FROM CLASSIFYING PEOPLE AS SIMPLY SICK OR WELL	99
M. PHYSICAL HEALTH MEASURES FOR PREGNANT WOMEN	102
REFERENCES	103

TABLES

1.	Operational Definitions and Enrollment Values of Health Status Measures Used To Define Subgroups for Prediction Analyses: Five General Health Measures	8
2.	Operational Definitions and Values of Health Status Variables Used To Define Subgroups for Prediction Analyses: Health Habits and Physiologic Measures	10
3.	Values on Demographic, Study, and Health Status Measures at Enrollment, by Type of Experimental Insurance Plan	15
4.	Numbers of Adult Enrollees, According to Category of Participation in Experiment and Plan	17
5.	Predicted Exit Values and Raw Mean Differences of Health Status Measures for an Average Person, According to Measure and Plan	20
6.	Predicted Exit Values of Five General Health Measures, According to Measure, Plan, Income, and Initial Health Status	21
7.	Predicted Exit Values for Health Habits and Physiologic Variables in Elevated-Risk Groups, According to Variable and Plan	22
8.	Predicted Exit Values of Blood Pressure, Vision, and Risk of Dying in Elevated-Risk Groups, for Low and High Income Groups	24
C.1.	Reliability and Stability Estimates for HIE General Health Status Measures	45
E.1.	Regression Coefficients, t-Test Values, and Other Equation Results: Personal Functioning, Role Limitations	65
E.2.	Regression Coefficients, t-Test Values, and Other Equation Results: Mental Health, Social Contacts	66
E.3.	Regression Coefficients, t-Test Values, and Other Equation Results: Health Perceptions, Smoking	67
F.1.	Regression Equations To Predict Exit Values for Determining Elevated Risk: Functional Far Vision	70
F.2.	Regression Equations To Predict Exit Values for Elevated Risk: Cholesterol	71
F.3.	Regression Equations To Predict Exit Values for Elevated Risk: Diastolic Blood Pressure	73

F.4.	Regression Equations To Predict Exit Values for Elevated Risk: Risk of Dying Index	74
F.5.	Regression Coefficients, t-Test Values, and Other Equation Results: Cholesterol and Weight	75
F.6.	Regression Coefficients, t-Test Values, and Other Equation Results for the Entire Sample: Functional Far Vision, Diastolic Blood Pressure, and Risk of Dying	76
F.7.	Regression Coefficients, t-Test Values, and Other Equation Results with Income and Elevated-Risk Interactions: Functional Far Vision, Diastolic Blood Pressure, and Risk of Dying	77
G.1.	Values of Maximum t, Maximum Absolute t, and Usual t at Selected Points in the Distribution	79
G.2.	Optimal Sample Percentiles and Other Evaluative Measures, by Health Condition	80
H.1.	Predicted Exit Values for Health Perceptions, by Site and Plan	83
H.2.	Predicted Exit Values for Diastolic Blood Pressure, by Site and Plan	84
H.3.	Predicted Exit Values for Corrected Vision, by Site and Plan	85
I.1.	Reasons for Sample Loss for Physiologic Measures, Cholesterol, and Weight, by Plan	87
I.2.	Reasons for Sample Loss and Final Sample Size, by Individual Health Measure and Plan	88
J.1.	Predicted Exit Values for Only Those Persons with Complete Enrollment and Exit Data	90
J.2.	Differences in Enrollment Values Between Noncompleters and Completing Sample, by Plan	92
J.3.	Discrepancies in the Differences Between Free and Cost-Sharing Plans in Predicted Exit Values When Noncompleters Are Included in Analysis and When They Are Not, by Health Status Measure	94
K.1.	Distribution of Sources of Enrollment Values for the General Health Ratings Index, by Source and Plan	96
K.2.	Distribution of Sources of Exit Scores for the General Health Ratings Index, by Source and Plan	97
K.3.	Predicted Exit Values for General Health Ratings Index When Persons with Missing Values Are Eliminated from the Analysis	98

L.1.	Percentage Distribution of Analytic Sample, by Level of Diastolic Blood Pressure at Exit and by Plan	100
L.2.	Percentage Distribution of Analytic Sample, by Level of Functional Far Vision at Exit and by Plan	100
M.1.	Results on Physical Health Measures With and Without Pregnant Women	102

Chapter 1

INTRODUCTION

Spending at least some money on medical care is indisputably worthwhile. But does spending yet more buy still better health? In individual cases, the answer may be an obvious yes or no, but in the population as a whole the point of diminishing (or absent) returns has been difficult to identify.¹

Critics of the existing system have contended that developed countries spend too much on medicine; they argue that this practice increases iatrogenic illness (Carlson, 1975; Illich, 1977). The extreme versions of this argument, a kind of "therapeutic nihilism," have been cogently criticized (Starr, 1976; Rogers and Blendon, 1977), and in this country public policy has proceeded for more than five decades on the assumption that if some medical care is good, more would be better. A main instrument of this policy has been increased insurance coverage, both public and private. While this policy has been in effect, the national outlay on medical care has steadily increased and has now reached a level that causes concern in many quarters. One of the few potential ways to reduce expenditures appears to be to raise the proportion of costs borne by the individuals who are consuming medical care.

What fraction of their costs, if any, patients should be required to pay is thus a central and serious question of policy. Proponents of cost-sharing argue that it curtails frank abuse and restrains the purchase of care that yields little or no benefit. Opponents counter that if people must pay out of pocket for medical care, their access to appropriate levels of care will decrease and they will suffer accordingly. Data in support of either position have been all but nonexistent.

This dearth of information prompted the federal government to support a controlled trial, known as the Rand Health Insurance Experiment. The project randomly assigned a sample of families to a variety of different insurance plans; one group received all their medical care free of charge, whereas others paid some percentage of their health bills up to a stipulated maximum. We have already reported that when

¹See, for example, Benham and Benham (1975); Rice and Wilson (1976); Cochrane et al. (1978); Miller and Stokes (1978); Newhouse and Friedlander (1980); Miller and Miller (1981); and Hadley (1982).

cost-sharing was higher, use of medical care (visits to physicians, adult hospitalizations) and, accordingly, total expenditures, were lower (Newhouse et al., 1981). To illustrate, people enrolled in cost-sharing plans made only about two-thirds as many outpatient visits as those receiving free care (Keeler and Rolph, 1983).

These earlier analyses left an important question unanswered: Were the people who received free medical care, and thus used more of it, *healthier* as a result? Without being able to remove all doubt on this score, we can now reduce the uncertainty. Here we report what happened to health status among a group of adults under age 65 who received free care, as compared with a similar group required to share in the cost of care.

Chapter 2

EXPERIMENTAL DESIGN AND ANALYTIC METHODS

SAMPLE AND SITES

The experiment, which ran from November 1974 through January 1982, enrolled 3958 persons between the ages of 14 and 61 who belonged to 2005 families. Included in the study but not in this analysis (or the above numbers) were children under the age of 14 and families enrolled in a prepaid group practice; they are the subjects of separate analyses. Families who moved stayed in the experiment as long as they remained in the United States, except for those enrolled in a control group in the prepaid group practice. Seventy percent of the sample participated for three years, the remainder for five.

Families lived in six sites: Dayton, Ohio; Seattle, Washington; Fitchburg and Franklin County, Massachusetts; and Charleston and Georgetown County, South Carolina. To select sites, we first calculated the optimal number of sites for our budget, which turned out to be between four and nine. We decided to use six sites, chosen according to several criteria: (a) to represent all four census regions and thus account for any regional variation in the responsiveness of demand for medical care to cost-sharing; (b) to obtain a spectrum of city size, because the complexity of the medical care delivery system (and hence the response to insurance) could vary with city size; (c) to achieve variation in waiting times for an appointment and in the proportion of primary care physicians accepting new patients, because response to demand could vary according to the amount of excess demand for ambulatory services; (d) to include participation from both Northern and Southern rural areas, because these areas differ in economic and racial characteristics; and (e) to ensure that one site had a well-established prepaid group practice.

Enrolled families represented the general populations where they lived except for several intentional differences. The experiment excluded families with annual incomes exceeding \$54,000 (1982 dollars), or about 3 percent of the families initially contacted. Families participating in the Supplemental Security Income (SSI) program, persons who were so disabled that they were eligible for Medicare, and family members over the age of 61 at entry to the study also were

excluded. Low-income families were slightly oversampled. Major demographic characteristics of the HIE sample (averaged across our sites) do not differ markedly from those of the nation as a whole, save for the intentional departures such as those relating to age.

INSURANCE PLANS AND BENEFITS

Families were assigned to one of 14 experimental insurance plans by a random sampling technique, the Finite Selection Model, that made the distribution of over 25 family and individual characteristics as similar as possible across plans (Morris, 1979). We offered families only the plan to which we randomly assigned them because we judged that explaining all the plans before asking a family to join the experiment would be confusing and might make the enrollment process difficult to complete.

For this report, the 14 insurance plans are grouped into four categories (one providing free care, the other three requiring cost-sharing):

- The "free-care" plan, in which the family received all its care without charge (i.e., 0-percent coinsurance).
- The "individual deductible" plan, which imposed a 95-percent coinsurance rate on *outpatient* expenditures only, up to a maximum out-of-pocket expenditure of \$150 per person (\$450 per family) per year; all outpatient care beyond that amount, and all inpatient care, was free.
- The "intermediate coinsurance" plans, consisting of nine plans with either a 25-percent or a 50-percent coinsurance rate. On these plans, the family paid 25 or 50 percent of its medical bills each year up to a maximum dollar expenditure of \$1000 (in 1973 dollars) or 5, 10, or 15 percent of family income, whichever was less. In three of these nine plans, the family paid 50 percent of the cost of its mental and dental services; in some sites and years, the maximum expenditure was held to \$750.
- The "income-related catastrophic" plans, which included three plans with a 95-percent coinsurance rate and the same income-related limitations on maximum dollar expenditure as the intermediate coinsurance plans.

The following example illustrates how the experimental plans operated. A family assigned to a plan with a 25-percent coinsurance rate and a \$1000 maximum dollar expenditure would pay 25 percent of all medical and dental bills in each year until the total of all of these

bills reached \$4000. At that point it would have spent \$1000 out of pocket, after which all further expenditures during that year would be fully paid (or reimbursed) by the plan. At the beginning of the next year, the family would again pay the 25-percent coinsurance until the \$1000 limit was reached.

For all cost-sharing plans, the \$150 individual deductible and the \$1000 (or \$750) cap on family out-of-pocket expenditure remained constant during the entire experiment; these caps were not modified to reflect inflation, which was considerable. At the end of the experiment (1982), nearly \$2000 was required to purchase the amount of medical care that \$1000 had bought at the beginning (1974). Subject to the \$1000 cap, the family's maximum expenditure ceiling would be adjusted each year to reflect any change in family income.

Families assigned to an experimental plan that was less generous than their current insurance were paid an amount (in installments every four weeks) that was sufficient to prevent their ever becoming worse off financially by enrolling (i.e., was equal to their highest possible loss). This money was paid irrespective of actual use of medical services; its intent was to help ensure that acceptance and retention rates would be high and independent of both the plan to which the family was assigned and the family's health status. For example, if the family had been assigned to an experimental plan whose maximum dollar expenditure turned out to be \$425 and the family had had an existing policy with a \$100 deductible and a 20-percent coinsurance above the deductible, it would have been paid \$260 a year, that is $(260 = 425 - 100 - 0.2(425 - 100))$, or about \$20 every four weeks. Such payments had a negligible effect on use (Newhouse et al., 1981).

All plans were administered by the Family Health Protection Plan (FHPP) through a fiscal intermediary with considerable experience in the insurance industry. All plans had an identical, very comprehensive definition of covered services. These included hospital, physician, dental, mental health, vision, and hearing services; drugs (including over-the-counter drugs for certain chronic conditions); and supplies. Services of nonphysician providers, such as audiologists, chiropractors, clinical psychologists, optometrists, physical therapists, and speech therapists, were also covered. The only noteworthy exclusions from coverage were nonpreventive orthodontic services, cosmetic surgery for pre-existing conditions, and outpatient mental health visits exceeding 52 per year. Appendix A gives a detailed accounting of the FHPP benefits; Clasquin and Brown (1977) and Brown (1984) present more information on operations and administration of the FHPP and the HIE.

In many analyses reported in Chapter 3, we grouped the cost-sharing plans and compared them with the free-care plan. In these instances, the total cost-sharing plan represents an equally weighted average of the three types of cost-sharing plans.

HEALTH STATUS VARIABLES

Starting with the World Health Organization (1948) definition of health, we developed or adapted measures to evaluate the effect of cost-sharing on health status. This comprehensive set comprised four distinct categories: general health, health habits, physiologic health, and the risk of dying from any cause related to various risk factors. Appendix B lists the health status variables studied and describes the main data collection procedures.

In this report we analyze 11 measures from these four categories; these measures are described in Tables 1 and 2. Appendixes C and D give more detail about the ways these health measures were defined and assessed. A number of other physiologic measures, as well as measures of dental health, remain to be examined. Data on the general health measures (such as physical functioning, role functioning, and health perceptions) and health habits (such as smoking) were collected from a medical history questionnaire (MHQ) that was self-administered at the beginning of the experiment (enrollment) and three or five years later (exit). The reliability, validity, and other psychometric properties of the five general health measures are reported elsewhere and summarized in Appendix C.¹

Serum cholesterol, blood pressure, and visual acuity were measured at medical screening examinations that were given at enrollment to a randomly selected 60 percent of the sample and at exit to the entire sample (see Appendix D for citations to works that describe these measures). The random sample was used to estimate the effect of the screening examination itself on the use of medical services and subsequent health status. Technicians who collected data at the entrance and exit screening examinations did not know the insurance plan to which the family belonged.

Because actual deaths in our experimental population (about 40 adults in all the fee-for-service plans) were too infrequent to allow

¹Physical functioning was measured by the Personal Functioning Index, role functioning by the Role Limitation Index, mental health by the Mental Health Inventory, social contacts by the Social Contacts Scale, and health perceptions by the General Health Ratings Index. See Davies et al. (forthcoming) for rules on scoring these measures, and Appendix C for citations to individual volumes describing their development.

meaningful analysis, we calculated an index that would predict the extent to which longevity would be affected by specified risk factors (systolic blood pressure, smoking behavior, and serum cholesterol level (McGee and Gordon, 1976)). Data for the index came from both the screening examination and the MHQ.

METHODS OF ANALYSIS

To answer the question "Did the free plan improve health more than the cost-sharing plans?" we began by identifying certain variables that could be expected to affect the results and could be used to develop health care policy. We then employed regression methods to estimate the influence of the "explanatory" variables on the "dependent" variable—namely, health status at exit.

Explanatory variables in the regression models included: insurance plan; other experimental variables (whether the subject took the enrollment screening examination; length of enrollment period); site; family income adjusted for size and composition of the family; the subject's initial value for the dependent (health status) variable; and other variables that helped explain health status at exit. Details of these models can be found in Appendixes E and F.

To interpret these effects we then used the results from the regression equations to predict the exit health status of people with any given set of entry characteristics. In particular, using the coefficients for the explanatory variables from the main regression equations, we calculated values for health status at exit for the average participant and for those in certain subgroups with relatively high or low incomes, good or poor health.

Because we especially wanted to know the effect of cost-sharing on persons with poor health or low income, we measured all interactions between health and income and the various insurance plans. A score on each of the five general health measures was determined for a person who was initially "ill" or in "good" health with respect to the specific dependent variable under consideration. We defined "ill" as the mean value of persons in the lowest one-fifth of the distribution at enrollment and "good" as the mean in the highest two-fifths (Table 1). The effect of "low" or "high" income at enrollment was also tested. A "low" income was the mean for the lowest one-fifth of the income distribution (for a family of four, about \$7,300 in 1982 dollars), and "high" the mean for the highest two-fifths (\$40,000). To generate the predic-

Table 1
OPERATIONAL DEFINITIONS AND ENROLLMENT VALUES OF HEALTH STATUS MEASURES USED
TO DEFINE SUBGROUPS FOR PREDICTION ANALYSES: FIVE GENERAL HEALTH MEASURES

Health Measure and Operational Definition	Typical Item	Mean Value of Persons at Enrollment		Interpretation of Effect Size
		"Good" Health ^a	"Ill" Health ^b	
PHYSICAL FUNCTIONING: A standardized (0-100) scale (21 items) that indicates the degree to which the person has limitations in personal self-care, mobility, or physical activities. A high score means greater capacity for physical activity.	"Do you have any trouble either walking one block or climbing one flight of stairs because of your health?"	100	44.8	A 10-point difference = effect of having chronic, mild osteoarthritis. ^{c,d}
ROLE FUNCTIONING: A dichotomous measure (2 items) that indicates whether the person can do work, school, or housework activities free of limitations because of poor health. A high score means a higher probability of role functioning. Mean probabilities are expressed as percentages.	"Does your health keep you from working at a job, doing work around the house, or going to school?"	100	0	A 1-point difference = 1-percentage-point higher probability of being limited in the performance of one's major role.
MENTAL HEALTH: A standardized (0-100) scale (38 items) that measures anxiety, depression, emotional ties, behavioral/emotional control, and psychological well-being during the past month. A high score reflects higher or more positive levels of mental health.	"How much of the time, during the past month, have you felt down-hearted and blue?"	86.4	53.0	A 3-point difference = impact of being fired or laid off from a job.

Table 1 (continued)

Health Measure and Operational Definition	Mean Value of Persons at Enrollment		Interpretation of Effect Size
	Typical Item	"Good" Health ^a	"Ill" Health ^b
SOCIAL CONTACTS: A standardized (0-100) scale (3 items) that measures contacts with friends and relatives during the past month or year. A high score reflects higher levels of social activity.	"About how often have you visited with friends at their homes during the past month?"	94.3	29.1
			A 10-point difference = a 2-percentage-point increase in the probability of being psychiatrically impaired.
HEALTH PERCEPTIONS: A standardized (0-100) scale (22 items) that measures the person's perceptions of past, present, and future health, susceptibility to illness, and worry about health. A high score reflects better perceptions of one's health status.	"My health is excellent."	83.6	47.8
			A 5-point difference = effect of having been diagnosed as hypertensive. ^c

^aMean of the healthiest 40 percent of the distribution.

^bMean of the sickest 20 percent of the distribution.

^cAmong participants in the experiment, adjusted for age and sex.

^dClassification based on the person's responding yes to questions about ever having acute or chronic pain, aching, swelling, or stiffness in fingers, hip, or knee.

^eClassification based on the person's responding yes to a question about ever being diagnosed as having high blood pressure and yes to a question about being so diagnosed more than once or to a question about having been prescribed pills or medicines for high blood pressure.

Table 2

OPERATIONAL DEFINITIONS AND VALUES OF HEALTH STATUS VARIABLES
USED TO DEFINE SUBGROUPS FOR PREDICTION ANALYSES:
HEALTH HABITS AND PHYSIOLOGIC MEASURES

Health Variable and Operational Definition	Mean Value of Persons at Elevated Risk ^a	Specific Scoring
SMOKING: A six-level measure of the risk of death from smoking compared to not smoking.	1.89	Never smoked/ex-smoker: 1.00 Pipe/cigar smoker only: 1.06 Cigarette smoker: <1 pack/day 1.57 1 pack/day 1.79 2 packs/day 2.07 >2 packs/day 2.20
WEIGHT (kg): ^b	88.5	Standardized for height (in meters) by multiplying by (1.75/height ²) for men and by (1.65/height ^{1.5}) for women. Standardized for sex by summing 0.5 (average value for men) and 0.5 (average value for women).
SERUM CHOLESTEROL (mg/dl):	242	
DIASTOLIC BLOOD PRESSURE (mm Hg):	87	
FUNCTIONAL FAR VISION: Measured in Snellen lines. Functional means with whatever correction (if any) was used by the individual to improve vision.	2.95 ^c	Line 2 = 20/20 Line 3 = 20/25 Line 4 = 20/30 Line 5 = 20/40
RISK OF DYING: The risk of dying from any cause compared to persons with average values of major risk factors: $100 \exp(\text{Index}) / (1 + \exp(\text{Index}))$. Index = 1.28 smoking scale + 0.0023 cholesterol + 0.023 systolic blood pressure - 4.92.	2.02	The coefficients of the risk factors are median values of the coefficients in the logistic regressions for death from any cause in five studies of heart disease in middle-aged men.

^aMean of the sickest 25 percent of the distribution except for functional far vision. For smoking and weight, these are enrollment values. For cholesterol, blood pressure, vision, and risk of dying, these are predicted exit values.

^bExcludes those 14-17 at enrollment and pregnant women.

^cMean corrected vision of those whose uncorrected vision in the better eye was worse than 20/20; i.e., mean of the worst 53 percent of the distribution.

tions, we used mean population values for all remaining explanatory variables.

Medical care could be expected to benefit most those people who have a health problem, but the effects of cost-sharing might be obscured were data on this subsample pooled with the whole. Accordingly, for each indicator of physiologic health (blood pressure, vision), health habits (smoking, weight, and cholesterol), and risk of dying (Table 2) we divided our sample into those likely, by exit, to have abnormal or normal values. The division was based on data from the initial examination and responses to pertinent items on the MHQ for each measure separately. Appendix F describes how the elevated-risk groups were defined in more detail, and the possible effects of alternative definitions on our conclusions can be found in Appendix G.

We detected few important effects of plan on values for the group expected to exit with normal values, so we focused the analysis on the group expected to be least healthy or at elevated risk of dying. The least healthy one-quarter of the sample for five of these six measures was arbitrarily designated a priori as the elevated-risk group. For visual problems, we defined "high risk" as an acuity at exit worse than 20/20 in the better eye without glasses (roughly half the sample). Because the "best" results might have been obtained had the elevated-risk groups been more (or less) inclusive, we performed additional analyses to clarify how size of the group might have influenced our findings. This work is discussed in Appendix G.

We had no prior expectations of positive or negative effects of cost-sharing, so we used two-tailed tests of significance throughout. All statistical tests were corrected for correlation of the error term within each family and for the nonconstant variance of the error term (Huber, 1967). The corrections were generally small.

We followed the convention of labeling a result "significant" were it likely to occur by chance no more often than one time in 20 (i.e., $p \leq 0.05$). Nonetheless, results falling short of this criterion should not be dismissed. In some cases, the confidence intervals may indicate that the result's actual value could plausibly have some clinical importance; that is, the range of values having 95-percent likelihood of bracketing the real one could include some that are medically important. For this reason, we have generally given a probability value if that value is close to the 5-percent level.

POSSIBLE ARTIFACTS AND BIASES

We anticipated three problems that may have led to biased estimates or erroneous inferences. First, the various plan offerings might have been accepted by different kinds of people whose health or other characteristics would have biased the outcome. Second, participants may have dropped out of the various plans at different rates as a function of their current health. Either factor could distort our picture of the actual effects of a particular plan. Third, certain data were missing: Some gaps were "unplanned" (for example, participants occasionally did not complete all questions on the exit questionnaire), and some were "planned" (certain participants, for instance, were not asked to take an enrollment screening examination). Only the unplanned loss of data carried the potential for bias, because the planned gaps were distributed randomly.

We adopted several strategies to counter the potential for bias. First, we compared health status values at enrollment for participants in each plan, and we compared selected characteristics of people who refused the offer with those of people who accepted. If these groups had similar values, we would have little reason to suspect bias.

Second, our regression models included initial values of the health status variables and of other variables known to influence the response variable under study. (For example, high blood pressure at entry predicts high blood pressure at exit.) We thereby controlled statistically for any effect of nonrandom composition of the sample with respect to these explanatory variables.

Third, we obtained longitudinal information on the general health measures and smoking for people who voluntarily withdrew from the experiment and for those who did not complete the experiment for other reasons. These data came from an abbreviated medical history questionnaire completed after they left the experiment or from an annual health questionnaire filled out during the experiment. Thus, the analyses included information on many of the dropouts. We did not attempt to recover information on physiologic measures from participants who left the sample prematurely; results for these measures are based only on those who completed the experiment.

Missing data from unplanned nonresponses never amounted to more than 2 percent for any one question, so bias from this source should be negligible. Nevertheless, to include people with missing data in the analysis, we imputed scores to them. For the questionnaire-based measures we estimated a value from that individual's responses to related questions or used values from questionnaires completed in the previous year. For the physiologic measures and risk of dying, persons

who by design did not receive the initial screening examination were imputed an enrollment value from a prediction equation based on age, sex, and MHQ data. They were included in the analysis but given less weight (Dagenais, 1971).

To determine whether various factors such as site, sample loss, or missing data biased our results, we carried out several special analyses. Appendix H examines whether findings on the general health ratings index, diastolic blood pressure, and vision differed by study site. Appendix I gives additional information on sample loss, and Appendices J and K deal with the effects of including or excluding from the analytic sample persons who were missing enrollment or exit data. Appendix J in particular shows that, for the health measures based on screening examination information, not including persons who were missing such data because they left the study prematurely had only negligible effects on our findings.

Most of our measures are continuous, not categorical, but often the medical profession views people as being in one of two categories, sick or well. We investigated whether our main results on the effects of the plan would persist when enrollees were classified as "sick" (diastolic blood pressure 90 mm Hg or higher or corrected vision 20/40 or worse) or "well" (all others). Findings based on the proportion hypertensive or myopic by plan are discussed in Appendix L.

Chapter 3

RESULTS OF DESIGN AND HEALTH STATUS ANALYSES

THREATS TO VALIDITY

Acceptance of the Enrollment Offer

Acceptance rates varied by plan: 92 percent of the families accepted the offer to join the free plan, 83 percent the individual deductible plan, 89 percent the intermediate plans, and 75 percent the catastrophic plans. To determine whether these different acceptance rates might have biased our results, we examined the health status of all enrollees at the start of the experiment and detected no significant differences among plans in any enrollment health measure or in family income, education, or age (Table 3). Only the proportion of females differed slightly by plan, and one significant difference would be expected by chance alone among the 20 comparisons made.

We also compared people who refused the enrollment offer with those who accepted. Results of this comparison (Rogers and Camp, forthcoming) established that the different acceptance rates were unlikely to have affected our conclusions.

Retention in the Experiment

During the experiment, each plan lost some of its participants owing to voluntary withdrawal (including joining the military), involuntary factors (such as incarceration), health reasons (mainly by becoming eligible for disability Medicare), or death. The latter two health-related factors did not differ materially by plan (Table 4). In all, 95 percent of those on the free plan completed the experiment and exited normally by completing the MHQ and going through the final screening examination, as did 88 percent of those on the individual deductible plan, 90 percent on the intermediate plans, and 85 percent on the catastrophic plans.

To test whether these differences affected our results, we collected data on general health measures and smoking behavior of people who had terminated for various reasons. Our findings were not altered by

Table 3

VALUES ON DEMOGRAPHIC, STUDY, AND HEALTH STATUS MEASURES AT ENROLLMENT, BY TYPE OF EXPERIMENTAL INSURANCE PLAN^a

Variable and Brief Description ^b	Cost-Sharing Plans				Free Plan	t-Test Value ^d for Difference Between Free and Cost-Sharing Plans
	Catastrophic	Inter-med.	Ind. Deduct.	Total Cost-Sharing ^c		
No. of enrollees ≥14 years of age	759	1024	881	2664	1294	
Mean age (yr)	32.8	33.8	33.6	33.4	33.3	-0.0
Sex (% female)	56.1	53.5	53.8	54.4	52.2	-2.1
Race (% nonwhite)	20.8	17.4	18.3	18.9	16.6	-0.5
Mean family income adjusted for family size (\$ 1982 thousands)	21.5	22.8	23.3	22.5	22.1	-0.5
% Hospitalized in year before enrollment	11.5	11.2	12.0	11.6	11.7	0.1
Mean no. of physician visits in year before enrollment	4.49	4.23	4.80	4.51	4.55	0.2
Mean education (yr)	11.9	12.0	12.0	12.0	11.8	-1.4
% Taking enrollment screening examination	59.1	57.8	58.6	58.5	62.5	1.6
% Enrolled for 3 years	69.8	67.4	71.3	69.5	68.9	-0.3
Physical functioning (mean score, 0-100)						
Enrollees	89.6	88.7	89.1	89.1	88.9	-0.2
Analytic sample	89.6	89.0	89.6	89.4	89.0	-0.5
Role functioning (mean score, %)						
Enrollees	94.8	91.9	91.8	92.8	93.1	0.3
Analytic sample	94.8	92.1	92.5	93.1	93.0	-0.2
Mental health (mean score, 0-100)						
Enrollees	73.8	75.0	73.7	74.2	74.7	0.9
Analytic sample	73.8	75.1	73.9	74.3	74.7	0.8
Social contacts (mean score, 0-100)						
Enrollees	72.8	72.1	72.3	72.4	72.5	0.1
Analytic sample	72.6	72.2	72.0	72.2	72.5	0.3

Table 3 (continued)

Variable and Brief Description ^b	Cost-Sharing Plans				Free Plan	t-Test Value ^d for Difference Between Free and Cost-Sharing Plans
	Catastrophic	Inter-med.	Ind. Deduct.	Total Cost-Sharing ^c		
Health perceptions (mean score, 0–100)						
Enrollees	70.5	71.1	69.4	70.4	69.7	–1.2
Analytic sample	70.4	71.2	69.7	70.4	69.8	–1.2
Smoking scale (mean score, 1–2.20)						
Enrollees	1.29	1.30	1.32	1.30	1.29	–0.7
Analytic sample	1.28	1.29	1.30	1.29	1.29	–0.3
Mean standardized weight (kg)						
Enrollees	71.5	71.3	71.0	71.3	71.3	0.0
Analytic sample	71.6	71.3	71.6	71.5	71.6	0.2
Mean cholesterol level (mg/dl)						
Enrollees	207	205	206	206	202	–1.9
Analytic sample	208	205	207	207	204	–1.5
Mean diastolic blood pressure (mm Hg)						
Enrollees	75.2	75.3	75.4	75.3	74.6	–1.4
Analytic sample	76.0	75.4	75.7	75.7	74.7	–1.9
Functional far vision (mean no. of lines)						
Enrollees	2.28	2.39	2.42	2.37	2.33	–0.9
Analytic sample	2.28	2.37	2.41	2.35	2.32	–0.9
Risk of dying (mean score)						
Enrollees	0.99	1.05	1.12	1.05	1.04	–0.6
Analytic sample	1.00	1.06	1.13	1.06	1.03	–0.8

^aValues are adjusted for differences according to site.

^bFor demographic data, table entries include everyone with valid enrollment data. For health measures, the mean score for enrollees excludes persons who did not have valid enrollment data owing to the study design (e.g., they were not assigned to an initial screening examination) or to missing data, and the mean score for analytic samples excludes the same persons plus those who did not have valid exit data.

^cValues represent equally weighted averages of the three types of cost-sharing plans.

^dValue shown is for the difference between free and total cost-sharing plans.

Table 4
NUMBERS OF ADULT ENROLLEES, ACCORDING TO CATEGORY OF
PARTICIPATION IN EXPERIMENT AND PLAN

Category of Participation	Cost-Sharing Plans									
	Cata- strophic		Inter- med.		Ind. Deduct.		Total		Free Plan	
	No.	%	No.	%	No.	%	No.	%	No.	%
Total enrolled	759	100.0	1024	100.0	881	100.0	2664	100.0	1294	100.0
Completed enrollment and exited normally	642	84.6	926	90.4	772	87.6	2340	87.8	1225	94.7
Left experiment voluntarily	83	10.9	43	4.2	53	6.0	179	6.7	5	0.4
Terminated for health reasons ^b	3	0.4	13	1.3	11	1.2	27	1.0	15	1.2
Terminated for nonhealth reasons ^b	24	3.2	31	3.0	34	3.9	89	3.3	38	2.9
Died	7	0.9	11	1.1	11	1.3	29	1.1	11	0.9
Recovered for analysis ^c	94	80.3 ^d	84	85.7	69	63.3	247	76.2	54	78.3
									301	76.6

^aThe actual analyses were based on a slightly smaller sample, because forms were not available for under 1 percent of this sample.

^bParticipation ended because the person no longer fulfilled criteria for eligibility. Health reasons included becoming eligible for disability Medicare and being institutionalized; nonhealth reasons included joining the military and failure to complete data-collection forms.

^cForm nonresponse not included. The number analyzed equals the number completed plus the number recovered minus the number of nonresponses.

^dPercentages in this row are based on the number of enrollees in each plan who did not complete enrollment.

including or excluding these data, which were obtained from 73 percent of those who withdrew voluntarily, 83 percent of those terminated for health reasons, 82 percent of those terminated for nonhealth reasons, and 78 percent of those who were reported to have died. Thus, reported results include data from these individuals, and the final sample for the questionnaire-based analyses comprises 99 percent of the participants on the free and intermediate plans, 97 percent of those on the catastrophic plan, and 95 percent of those on the individual deductible plan. The percentages with complete data on physiologic measures (as well as weight) are lower because no post-enrollment screening examination was administered to the participants who left the experiment early.

As a further check for possible bias, we examined the values for health status at enrollment in the actual sample used for each analysis. We detected no significant differences by plan (Table 3).

EFFECTS ON HEALTH STATUS

Exit Values, by Plan

For the *average* person enrolled in the experiment, the only significant positive effects of free care ($p < 0.05$) were for corrected far vision and diastolic blood pressure (Table 5). Corrected vision of enrollees in the free plan was better (2.4 versus 2.5 Snellen lines, or an acuity of about 20/22 versus 20/22.5). Blood pressure was lower in the free plan by 0.8 mm Hg.¹

For any individual, of course, a change of 0.1 Snellen line seems small. As discussed in Appendix L, for the dichotomous analyses we classified these same enrollees as "having impaired vision even with correction" when their measured vision was 20/40 or worse (Snellen line 5 or higher) at exit. With this definition of impairment, 9.6 per-

¹This figure differs from those we reported earlier (Brook et al., 1983), which did not show a significant difference for the average enrollee ($p = 0.06$ in that analysis). The new results come from analyses that differed in three respects from the earlier one. First, we were able to include an additional 207 people from the South Carolina site on whom data had not been previously available. Second, we include the 58 people with incomplete MHQ enrollment data but complete exit data, after imputing to them an enrollment value based on site, sex, and age. They had been included in the other physiologic analyses but inadvertently omitted in blood pressure studies. Finally, in the earlier analyses, we had mistakenly used the first of two exit blood pressure readings in much of the sample; for this report, we used the second (and typically lower) reading in all cases. Appendix D explains when the two blood pressure measurements were taken during the screening examinations.

cent of persons on the free plan had a deficit in their corrected vision, contrasted with 12.0 percent on the cost-sharing plans.

A drop in blood pressure of 0.8 mm Hg translates into about a 2-percent decrease in the risk of dying, about the amount of decrease shown in Table 5. When the same sample is classified as hypertensive or not (i.e., at or above 90 mm Hg diastolic blood pressure), 14.1 percent of persons in the free plan and 16.9 percent of those on the cost-sharing plans would be hypertensive at exit (see Appendix L.)

No other health measure showed a significant difference between the free and cost-sharing plans. Furthermore, only for the risk of dying and role functioning did the direction of the overall (main) effect appear to favor the free plan (see the column for predicted mean differences in Table 5).² For the remaining measures, the direction of the main effect favored the cost-sharing plans or was neutral.³

Confidence limits for the differences between the free and the other plans were relatively narrow in all cases; thus, it is unlikely that our conclusion of little or no effect is far off the mark. To verify that this conclusion does not depend on our method of prediction, we compared the predicted differences with the differences between the raw means of the two groups. The predicted differences and the differences in the raw means scarcely diverged (see the two rightmost columns of Table 5), although precision is better for the predicted values.

Outcomes were more similar among the cost-sharing plans than between the free care plan and the cost-sharing plans. Such a finding is not surprising because utilization differences were greater between the free plan and the cost-sharing plans than within the cost-sharing group (Newhouse et al., 1981).

The Influence of Income and Health Status on General Health

In addition to detecting no significant effect on five general measures of health for the average individual, we were unable to detect any significant differences among subgroups who differed in income and initial health status (Table 6). Confidence intervals for subgroup

²In the earlier analyses (Brook et al., 1983), the first of two exit systolic blood pressure readings (rather than the second) was used to calculate the risk of dying index. Analyses were revised to use the second measurement in all cases, but because blood pressure is only one of three components in this index, changes in results were negligible. For example, the predicted mean difference fell from -0.018 to -0.014. The difference for the elevated-risk group (Table 7) fell from -0.21 to -0.19. Consequently, the results shown here are those already published.

³As explained in Appendix M, we could detect no meaningful effect on our results of including or excluding pregnant women in the physical and role functioning analyses, so they were left in these reported findings.

Table 5
PREDICTED EXIT VALUES AND RAW MEAN DIFFERENCES OF HEALTH STATUS MEASURES FOR AN
AVERAGE PERSON, ACCORDING TO MEASURE AND PLAN

Health Status Measures	No. ^a	Cost-Sharing Plans				Predicted Mean Difference (Free Minus Cost-Sharing) ^b	Raw Mean Difference (Free Minus Cost-Sharing)
		Cata-strophic	Inter-med.	Ind. Deduct.	Total Plan		
General Health (score, 1-100)							
Physical functioning	3862	86.0	85.0	84.9	85.3	0.0(-1.6,1.5)	-0.3(-2.3,1.7)
Role functioning	3861	95.5	95.0	94.7	95.4	0.3(-0.6,1.2)	-0.3(-2.2,1.6)
Mental health	3862	75.6	75.5	75.8	75.6	-0.2(-1.1,0.8)	-0.1(-1.1,1.0)
Social contacts	3827	69.3	70.2	69.8	69.8	-0.3(-2.3,1.6)	-0.2(-2.4,2.0)
Health perceptions	3843	68.1	68.0	67.9	68.0	-0.6(-1.5,0.3)	-0.9(-2.1,0.3)
Health Habits							
Smoking (scale, 1-2.20)	3758	1.28	1.29	1.29	1.29	0.0(-0.02,0.02)	-0.00(-0.03,0.03)
Weight (kg)	2804	72.8	72.6	73.1	72.8	0.0(-0.5,0.5)	0.0(-1.0,1.0)
Cholesterol level (mg/dl)	3381	202	200	204	202	1.0(-1.0,3.0)	-1.3(-4.5,1.9)
Physiologic Health							
Diastolic blood pressure (mm Hg)	3495	79.0	78.5	78.8	78.8	-0.8(-1.5,-0.1) ^c	-0.9(-1.8,-0.1) ^d
Functional far vision (no. of Snellen lines)	3477	2.55	2.50	2.51	2.52	-0.1(-0.16,-0.04) ^e	-0.13(-0.21,-0.05) ^f
Risk of dying (score)	3317	1.01	0.98	1.03	1.01	-0.02(-0.05,0.02)	-0.03(-0.08,0.02)

^aNumbers of persons in various parts of the analysis are dissimilar because noncompleters were not included for physiologic health, weight, or cholesterol level and because of differences among measures in the number of persons with valid enrollment or exit data. Teenagers 14-17 at enrollment and pregnant women were excluded from analyses of weight.

^bNumbers in parentheses are 95-percent confidence intervals; an approximate confidence interval is given for role functioning.

^ct = 2.21; p = 0.03.

^dt = 2.20; p = 0.03.

^et = 3.28; p = 0.001.

^ft = 3.18; p = 0.001.

Persons with normal vision were included and given a value of 2.0.

Table 6
PREDICTED EXIT VALUES OF FIVE GENERAL HEALTH MEASURES,
ACCORDING TO MEASURE, PLAN, INCOME, AND
INITIAL HEALTH STATUS^a

General Health Status Measure	Total Cost- Sharing	Free Plan	Free Minus Cost-Sharing ^b	Total Cost Sharing	Free Plan	Free Minus Cost-Sharing ^b
	<i>Low Inc. and Initial Ill Health</i>			<i>Low Inc. and Initial Good Health</i>		
Physical functioning	60.3	65.9	5.6(−2.9,14.0)	89.8	91.2	1.4(−1.6,4.4)
Role functioning	69.0	46.3	−22.7(−53.2,7.8)	95.0	96.1	1.1(−1.8,4.0)
Mental health	65.6	67.0	1.4(−1.8,4.7)	81.1	79.3	−1.8(−4.1,0.6)
Social contacts	51.8	55.3	3.5(−5.2,12.2)	77.7	77.9	0.2(−4.1,4.5)
Health perceptions	54.2	54.6	0.3(−3.0,3.7)	74.7	72.4	−2.3(−4.8,0.1)
	<i>High Inc. and Initial Ill Health</i>			<i>High Inc. and Initial Good Health</i>		
Physical functioning	59.9	55.6	−4.3(−9.8,1.2)	92.6	91.9	−0.6(−2.8,1.6)
Role functioning	60.3	56.0	−4.3(−24.1,15.5)	96.3	96.3	0.0(−2.0,2.0)
Mental health	63.3	64.5	1.3(−1.6,4.1)	82.7	82.1	−0.6(−1.9,0.7)
Social contacts	47.3	47.6	−0.3(−5.0,5.5)	82.2	80.1	−2.1(−5.1,1.0)
Health perceptions	52.8	52.1	−0.7(−3.1,1.7)	77.7	77.8	0.1(−1.4,1.6)

^aInitial health status is defined with respect to the individual health measure denoted in each row.

^bNumbers in parentheses are 95-percent confidence intervals; approximate confidence intervals are given for role functioning.

analyses were, of course, wider than for the sample as a whole. Hence, we cannot be as certain as with the entire sample that clinically important effects did not occur in these subgroups.

The Elevated-Risk Groups

At the end of the experiment, smoking, cholesterol, weight, and blood pressure did not differ significantly as a function of plan among participants judged to be at elevated risk on these measures at the study's outset (Table 7). Among those whose uncorrected far vision was worse than 20/20, corrected vision was, collectively, about 0.2

Table 7

PREDICTED EXIT VALUES FOR HEALTH HABITS AND PHYSIOLOGIC VARIABLES
IN ELEVATED-RISK GROUPS, ACCORDING TO VARIABLE AND PLAN

Health Habits and Physiologic Variables	Definition of Elevated-Risk Group ^a	Total Cost- Sharing	Free Plan	Free Minus Cost-Sharing ^b
Smoking	≥1.79 (≥1 pack per day)	1.75	1.73	-0.02(-0.06,0.03)
Weight	20% over ideal weight (kg)	89.1	89.4	0.3(-1.1,1.7)
Cholesterol level	≥220 mg/dl	243	244	2(-3,7)
Diastolic blood pressure	>83 mm Hg or taking hypertension drugs at enrollment	88.4	87.6	-0.7(-2.2,0.8)
Functional far vision	Line 3 (20/25) or worse for better eye	2.98	2.78	-0.2(-0.3,-0.1) ^c
Risk of dying	Risk >1.42	2.11	1.90	-0.21(-0.39,-0.04) ^d

^aElevated-risk groups are the least healthy 25 percent of the people as defined with respect to the individual health measure denoted in each row. For functional far vision, all persons with uncorrected natural vision worse than 20/20 are included.

^bNumbers in parentheses are 95-percent confidence intervals.

^ct = 3.29; p = 0.001.

^dt = 2.41; p = 0.02.

Snellen lines better, an improvement in visual acuity from 20/25 to 20/24 (p = 0.001).

For the average person at exit, the risk of dying from any cause (based on smoking habits, cholesterol level, and systolic blood pressure) was set arbitrarily at 1.0. By comparison, the relative risk of dying for someone in the elevated-risk group at enrollment (generally the upper quartile of the distribution of risk factors) was, on average, 2.02; a member of this group would be twice as likely to die during the next year as the average person of the same age and sex. For high-risk members on the free-care plan, the relative risk of dying was 1.90 at exit, as contrasted with 2.11 for those in the cost-sharing plans (Table 7). This 10-percent difference favoring free care was significant (p =

0.02) and was principally attributable to the improved control of high blood pressure among those on the free plan.⁴

Improvements in vision, blood pressure, and risk of dying were largest in the group with low income and elevated risk (see the first column of Table 8). For them, the differences between the free and cost-sharing plans were substantial for blood pressure and significant for the risk of dying; neither difference was significant in the higher income group. For instance, diastolic blood pressure for persons of low income judged initially to be at increased risk of hypertension was 2.3 mm Hg lower in the free plan ($p = 0.08$); for such persons of high income it was 0.1 mm Hg higher in the free plan.⁵

At this point, it is tempting to infer that free care improved the health of the sick poor but not the sick rich. Unfortunately, our data do not permit quite such a firm conclusion. If we begin with the hypothesis that free care makes *no* difference to the poor who are at elevated risk, our findings permit us to reject it: Free care *did* make a difference especially for the risk of dying measure, as shown by the values in Table 8. On the other hand, we did not demonstrate that free care benefited people with a high income and high risk; here we cannot reject the equivalent null hypothesis. Given the conditions of our experiment, free care made no *detectable* difference to this group. Now, however, a paradox emerges. If we start with another null hypothesis—that the two income groups responded in the same way to the various plans—we would expect to see it rejected, but because the differences between the two groups are not significant, we cannot reject this hypothesis.

Apparently, poor people at elevated risk benefited from receiving free care, but we cannot draw any conclusion about the higher-income group. We cannot say that they benefited from receiving free care, but we also cannot show that they responded differently from the lower-income group, which did benefit.

⁴Qualitatively, the systolic blood pressure results were very similar to those for diastolic blood pressure.

⁵Again, the findings for blood pressure differ somewhat from those reported in Brook et al. (1983). In particular, the results for the low income group had been significant at the $p < 0.05$ level in favor of the free-care plan.

Table 8
 PREDICTED EXIT VALUES OF BLOOD PRESSURE, VISION,
 AND RISK OF DYING IN ELEVATED-RISK GROUPS,
 FOR LOW AND HIGH INCOME GROUPS

Physiologic Measure	Total Cost-Sharing	Free Plan	Free Minus Cost-Sharing
<i>Elevated Risk and Low Income^a</i>			
Diastolic blood pressure	89.3	87.0	-2.3 (-4.9,+0.3) ^{b,c}
Functional far vision	3.61	3.30	-0.3 (-0.6,+0.02)
Risk of dying	2.13	1.83	-0.30 ^d (-0.6,-0.04)
<i>Elevated Risk and High Income</i>			
Diastolic blood pressure	88.0	88.1	+0.1 (-2.0,+2.2)
Functional far vision	3.21	3.14	-0.07 (-0.4,+0.2)
Risk of dying	2.09	1.96	-0.13 (-0.4,+0.1)

^aFor definitions of elevated risk for diastolic blood pressure and risk of dying, see Table 7. For functional far vision, elevated risk in this table refers *only* to the upper one-quarter of the distribution of values for uncorrected natural vision. Predictions in these two columns were made with use of the mean value of the elevated-risk group.

^bNumbers in parentheses are 95-percent confidence intervals.

^ct = 1.72; p = 0.08.

^dt = 2.23; p = 0.03.

Chapter 4

DISCUSSION OF THE HEALTH STATUS RESULTS

A central purpose of the HIE was to learn whether varying the amount of cost-sharing in medical care, including imposing no cost-sharing, affects the health of a general, nonaged population. Cost-sharing, or for that matter providing free care, is not a targeted or disease-specific approach to curtailing or expanding the use of medical services, in contrast to a program such as hypertension detection and follow-up.

Participants in the experiment received one of a graded set of insurance plans; for some, medical care was absolutely free, whereas for others the annual cost could range up to the lesser of 15 percent of family income or \$1000. The experiment was designed to be as "realistic" as possible. The sample was typical of a general population of adults with two major exceptions: It excluded severely disabled individuals eligible for public Disability Insurance (and hence Medicare) and those over age 65. (The disabled eligible for Medicare constitute 1 percent of the population.) Moreover, the study was conducted at sites representing a cross-section of American medicine; participants could, and did, choose their own physicians.

We found that under these circumstances the more people had to pay for medical care, the less they used. Adults who had to share the cost of care made about a third fewer ambulatory visits and were hospitalized about a third less often (Newhouse et al., 1981). Such reductions involve decisions both on the part of the patient to seek care and on the part of the physician to order services; they might well be expected to affect health status.

From our data, we can draw three conclusions about how such large differences in use affected health. We can therefore narrow the range of speculation about the relationship between cost-sharing and health status.

First, free care had *no* effect on the major health habits associated with cardiovascular disease and many types of cancer. More generous insurance, which prompted an average of one to two more encounters with a physician each year for several years, had no impact on smoking, on the weight of either the average or the overweight enrollee, or

on cholesterol levels (either average or elevated). Moreover, these habits, especially smoking, were at levels where substantial health benefit from behavior change was possible.

Second, we detected no effects of free care for the average enrollee on any of five general health measures (physical health, role functioning, mental health, social contacts, or health perceptions), and the confidence intervals rule out the possibility of anything beyond a modest effect. With respect to subgroups differing in income and initial health status, we can be less certain that this interpretation is correct, because the smaller samples yield wider intervals. Nonetheless, persons who scored low on these measures at enrollment exhibited substantial impairment in their health, and a more noticeable benefit from free care might have been anticipated for them.

Third, people who have certain conditions that are easy to diagnose and have well-established treatments (myopia, hypertension) benefited from free care. At the end of the experiment, enrollees with free care had better visual acuity and lower blood pressure. From the latter improvement, we infer that their risk of early death had been diminished. Although differences between income groups were insignificant, the improvements appear greater among the poor.

To illustrate the magnitude of the gains we did observe, consider the following examples. An average 50-year-old man in the late 1970s had approximately a 5-percent chance of dying within five years (U.S. Public Health Service, 1980). A 50-year-old man at elevated risk had approximately double that chance of dying. If 1000 50-year-old men at elevated risk were enrolled on a free rather than a cost-sharing plan, then we would anticipate that about 11 of them, who would otherwise have died, would be alive five years later ($1000 \times 0.05 \times (2.11 - 1.90) = 10.5$). An average 39-year-old woman, on the other hand, had a 1-percent chance of dying over the five-year period (U.S. Public Health Service, 1980); free care given to 1000 high-risk women of that age could be expected to keep only two more women alive than would care provided under cost-sharing arrangements.

These mortality reductions, in and of themselves, are not sufficient to justify free care for all adults. Investment in more targeted programs, such as hypertension detection and screening, represents a more cost-effective method for saving lives (Berwick et al., 1980). If free care yielded other life-saving benefits, for example a reduction in cancer deaths because of increased or more appropriate screening, such a conclusion could change.

Precisely how the various health benefits occurred is not yet known. We can comfortably conclude that free care did not prevent the occurrence of hypertension or visual impairments but rather

ameliorated their consequences. Future analyses based on data collected during the experiment, which examine both the use of services and the quality of medical care provided to patients with these conditions, will shed light on how the benefits—or lack thereof—came about.

Certainly, our results should prompt further examination of how, in the context of the fee-for-service system, people use services and physicians render care, and what the medical profession sees as its responsibility for improving the health of the nation. The questions prompted by our results are myriad: Why, when the levels of ambulatory care and hospital admissions were on average one-quarter to one-half again as high on the free as on the cost-sharing plans, were there so few positive effects on health status? Why was there no effect on health habits? Why was more physician contact seemingly of little or no benefit for many people, especially for those who reported themselves in poor health at the beginning of the study?

In view of other ways for improving health, we conclude that providing all medical services free to the nonaged public at large is not justified by the health benefits realized. A case for free care could be made, however, on other grounds. Cost-sharing plans could be inequitable to the poor or to those who have or will develop a chronic problem that demands continuing treatment. The latter group will clearly bear less of the medical bill, the less the cost-sharing. In addition, some may feel that free medical care, at least for some services, is necessary for a just society.

Our results must be used with caution to derive policies for special groups in the population. Poor families were protected by an income-related ceiling on their out-of-pocket medical expenses. The aged and those too disabled to work were not included in the experiment. Additional medical care for such persons may well provide benefits that a young, relatively healthy population does not experience.

The findings presented here may well disappoint those physicians who believe that reducing financial barriers to their care will dramatically improve patients' health. Some of these physicians might now contend that their apparent inability to influence patients' health habits argues for improved medical training and postgraduate education. Others may say it is evidence that such concerns should not fall within the purview of medicine, but rather are the obligation of lay groups or individuals themselves. One can only speculate whether professional training or the nature of the help that society seeks from physicians and other health care professionals could be sufficiently altered so that, if this experiment were to be repeated in the coming decades, its outcomes would be appreciably different.

Future studies will evaluate the benefits of free care already observed, as well as other possible benefits, relative to their costs. At this juncture, we conclude that although free care did not improve health status across the entire range of measures or income groups examined, it did confer demonstrable benefits for patients with selected conditions that physicians are trained to manage.

Appendix A

SCHEDULE OF BENEFITS IN THE FAMILY HEALTH PROTECTION PLAN

INTRODUCTION

This appendix outlines the main medical, dental, and mental health benefits provided by the Health Insurance Experiment (HIE) insurance plans administered through the Family Health Protection Plan (FHPP). The FHPP was the insurance carrier established solely for the HIE, for which the fiscal intermediary was Glen Slaughter and Associates, located in Oakland, California.

The main benefits covered are given first; certain important exclusions are also listed below. "Prior authorization" refers to an action of the FHPP Administrator; "per year" refers to an accounting, not a calendar, year. More complete documentation of these operational details can be found in Clasquin and Brown (1977). Brown (1984) discusses lessons learned regarding administration of the HIE.

BENEFITS

Inpatient Hospital Care

The following items were covered for an unlimited number of days for treatment of illness, injury, or pregnancy while the enrollee was an inpatient in a general, maternity, or mental hospital: (a) standard charges for other than a private room or for a private room, intensive care unit, or isolation room if certified as medically necessary or if no semi-private room was available; and (b) standard charges for medically necessary or prescribed supplies and services, to include use of operating room, anesthetic supplies, surgical supplies, dressings, and cast materials, physical therapy, drugs, x-ray and laboratory services, and private duty nursing services if medically necessary.

Outpatient Hospital Care

Standard charges were covered for services performed and supplies provided in the emergency room or outpatient department in conjunction with diagnosis or treatment of any illness or injury.

Skilled Nursing Facility

The following were covered for an unlimited number of days while the enrollee was an inpatient in a skilled nursing facility upon recommendation of a physician that such occupancy was medically necessary: (a) standard charges for other than a private room or for a private room if medically necessary or when no semi-private room was available; and (b) standard charges for medically necessary or prescribed supplies and services.

Physician Services

Reasonable or standard charges were covered for professional services provided by a physician on an inpatient or outpatient basis. (Acupuncture was covered only when performed by a physician.)

Surgical Benefits

Reasonable or standard charges for medically necessary surgical services in or out of a hospital were covered, including those provided by surgeons, assistant surgeons, anesthetists, and consultants during or after an operation.

Maternity Benefits

Reasonable or standard charges were covered for services related to maternity for the mother and child(ren) on either an inpatient or outpatient basis and including both hospital and physician services.

Psychiatric Services

Reasonable or standard charges were covered for psychiatric services rendered by psychiatrists, clinical psychologists, and mental health teams, up to a maximum of 52 visits per year.

Prescription Drugs

Reasonable or standard charges were covered for drugs that require a prescription.

Reasonable or standard charges were covered for certain nonprescription drugs (i.e., those that could be purchased without a

prescription) for selected conditions (including chronic allergic conditions, arthritis and rheumatism, chronic lower or upper gastrointestinal disease, pregnancy, chronic respiratory disease, and chronic skin conditions). A physician had to complete an insurance claim form certifying that the condition existed before payment was made, and purchases of nonprescription drugs so covered were limited to \$100 per person for each condition per year.

Dental Services

The following were covered when provided by a dentist and staff: diagnostic procedures, preventive services, restorative services, oral surgery, endodontics, periodontics, *preventive* orthodontia, and prosthodontics (except for any fixed bridge of seven or more units and replacement of any satisfactory denture or fixed bridge). Prior authorization was needed for any treatment plan exceeding \$500 (except emergency care) and replacement of crowns, bridges, or dentures.

Vision Care

Reasonable or standard charges were covered for services and supplies related to vision care, with the following limitations: (a) only one eye examination for refractions by an ophthalmologist or optometrist per year; (b) only one pair of corrective lenses per year; (c) additional costs for contact lenses, or light-sensitive, tinted, or shaded eyeglass lenses only when certified by an ophthalmologist or optometrist as medically necessary and given prior authorization; and (d) only one pair of eyeglass frames every two years, with a maximum payment based on the normal price of standard frames in the area.

Hearing Care

Reasonable or standard charges were covered for one hearing examination by a certified audiologist per year upon recommendation by a physician.

Reasonable or standard charges were covered for one hearing aid device prescribed by a certified audiologist every two years, with prior authorization and maximum payment based on the normal price of hearing aids in the area.

Care by Other Practitioners

Reasonable charges were covered for health care services provided by the following practitioners: (a) chiropractors; (b) podiatrists or chiropodists, when medically necessary; (c) physical, occupational, or speech therapists, when medically necessary and as part of a treatment plan given prior authorization; (d) private duty nurses (RN or LPN) when medically necessary (and with prior authorization if such services extended beyond 30 days in any one year); (e) Christian Science nurse, practitioner, or sanatorium.

Home Health Care

Reasonable or standard charges were covered for medically necessary services and supplies provided by a home health agency or others under a treatment plan established and periodically reviewed by a physician and provided on a visiting basis at home or at a hospital, skilled nursing facility, or rehabilitation center. Such services and supplies included: (a) part-time or intermittent nursing care; (b) physical, occupational, or speech therapy; (c) medical social services; (d) part-time or intermittent services of a home health aide; and (e) selected medically necessary equipment, appliances, and supplies (other than drugs) allowed by Title XVIII of the Social Security Amendments of 1965 (Medicare) (subject to some limitations and prior authorization).

Other Health Services and Supplies

Reasonable or standard charges were covered for the following: (a) x-ray and laboratory tests authorized by a physician (including x-ray, radium, and radioactive isotope therapy); (b) ambulance services when medically necessary; (c) family planning services provided by an agency licensed by the state;¹ (d) alcoholism treatment (from licensed or federally approved centers); (e) drug rehabilitation (from physician-directed or federally approved centers); (f) prosthetic devices including braces, limbs, eyes, and replacements; (g) other medically necessary equipment, appliances, and medical supplies allowed by Medicare (subject to some limitations and prior authorization).

¹Sterilization procedures for persons over 21 required prior authorization, and such procedures for persons under 21 or who were judicially declared mentally incompetent were not reimbursed at all. After 1977 abortion was not a covered service.

EXCLUSIONS

Apart from the few limitations noted for the service areas described above, the FHPP did not cover, or covered only with broad restrictions, the following types of services: (a) cosmetic surgery, except for repair of physical damage arising from accidental injury while the plan was in effect; (b) all orthodontics except those procedures considered interceptive or preventive (such as space maintainers and appliances needed to forestall later orthodontic treatment); (c) cosmetic dental services except for repair of physical damage arising from accidental injury while the plan was in effect; (d) supplies or services for injuries or conditions that were compensable under workmen's compensation, employers' liability laws, or automobile accident insurance policies, until those resources had been exhausted; (e) any custodial, personal, or other care that was not medically necessary; and (f) any care rendered outside the United States that could have been deferred until the enrollee returned to the United States (i.e., emergency medical care outside the country was covered).

Appendix B

HEALTH STATUS VARIABLES MEASURED IN THE HEALTH INSURANCE EXPERIMENT

HEALTH STATUS MEASURES FROM THE MEDICAL HISTORY QUESTIONNAIRE

The variables outlined below are constructed from information gathered on a self-administered MHQ that all adults (14 years and older) were requested to complete at enrollment and again at exit three or five years later. Some of these measures reflect states of health that medical care might well be expected to influence (such as physical health). Other measures, such as stressful life events, are only indirectly related to health status per se; they were evaluated as part of the comprehensive set of MHQ batteries and are noted here for completeness. Procedures for administering the MHQ are described later in this appendix.

The subheadings below denote the main dimensions of health status; the entries under each subheading describe discrete variables belonging to that dimension of health. The list is illustrative, not inclusive. Combining some variables listed under the subheadings yielded more aggregate indexes such as those used in this report.

Physical Health

- Person limited in self-care, mobility, and physical activities
(acute or chronic)
- Person limited in role activities (acute or chronic)
- Count of the number of disease conditions
- Person's ability to perform physical activities (i.e., no limitation in self-care, mobility, or physical activities)
- Number of physical symptoms in previous month
- Proportion of symptoms for which medical care was sought
- Bed disability days
- Work or school loss days
- Reduced activity days

Mental Health

- Presence of feelings of anxiety
- Presence of feelings of depression
- Absence of control over behavior
- Satisfaction with personal life
- Feelings of positive well-being (e.g., hopeful future, happy, full life)

Social Health

- Acquaintance with families in neighborhood
- Close friends and relatives to visit or talk with
- Getting along with other people
- Activity in voluntary groups
- Attendance at religious services

General Health Perceptions

- Self-ratings of health in the absolute and in comparison to that of others, in the past, present, and future
- Resistance and susceptibility to illness
- Pain due to health
- Worry and concern due to health
- Rating of health in terms of "excellent, good, fair, poor"

Patient Role

- Attitudes about going to doctor and avoiding the sick/patient role

Weight, Overweight, and Eating Habits

- Self-reported weight and height
- Self-evaluation as to being overweight
- Types of efforts undertaken to lose weight
- Eating habits (e.g., eat breakfast, between meals)
- Impact of weight problem (e.g., pain, worry, activity restriction)

Weight loss

Under doctor's care to lose weight

Smoking

Classification as to smoker, ex-smoker, or nonsmoker of cigarettes

For smokers and ex-smokers: length of time smoked and amount smoked (pack-years), type of tobacco smoked

Advice from physician to stop smoking

Drinking

Classification as to ever a drinker (of beer, wine, liquor)

Quantity of beer, wine, liquor (i.e., ethanol) consumed per day

Advice from physician to cut down on drinking

Types of efforts undertaken to cut down on drinking

Impact of drinking problems (worry, activity restriction, days in bed)

Other problems associated with drinking (DTs, cirrhosis of liver, accidents)

Exercise

Levels of physical activity (strenuous work or leisure activities, medium work or leisure activities)

Measures of metabolic cost of work or leisure activities and level of cardiac conditioning

Reasons for exercise

Health Habits

Behaviors related to smoking, drinking, maintaining low weight-for-height, being physically active, sleeping, eating breakfast, eating between meals

Diets

Type of special diet(s) individual might be on, and reasons for being on it

How carefully the diet is followed and how much bother it is

Medications and Psychotropic Drugs

Whether take heart medications (digitalis, etc.)

Frequency of use of sleeping pills, sedatives

Impact of use of sleeping pills, sedatives (worry, activity restriction, days in bed)

Types of Elective Surgery

Ever had cholecystectomy, hysterectomy, tonsillo-adenoidectomy

Symptoms in Recent Past

Presence of any of a list of 27 "trivial" to "serious" symptoms such as cough without fever, upset stomach, backache, shortness of breath, headache, bleeding not due to accident, joint stiffness, bladder trouble, loss of consciousness, feelings of nervousness or depression

Whether sought care for any symptom present

Chronic Illness

Presence of any of a list of 10 "other" illnesses in past year such as arteriosclerosis, allergies, chronic hepatitis, phlebitis, disease of uterus

Whether sought care for these illnesses in past year

Stressful Life Events

Personal circumstances related to jobs, marriage, etc.

Whether any of a number of positive or negative events have occurred within, say, the previous six months, such as change in responsibilities at work or at home, relations with spouse and in-laws (or significant other), death of relation or

close friend, change in residence, accident or victim of assault or robbery, debt, legal problems

Patient Satisfaction

Attitudes about availability and continuity of care, facilities, financing and costs of care, interpersonal aspects of doctor-patient relationship, quality and competence of physicians, general satisfaction with medical care

Other Measures

One or more questions was asked about:

missing limbs

cancer

tuberculosis

glaucoma

use of various appliances or aids (glasses, hearing aid)

HEALTH STATUS MEASURES BASED ON SELF-REPORT OR PHYSIOLOGIC TESTS

About 20 "tracer" conditions were assessed for adults on the basis of information from the MHQ and the screening examination:

Acne vulgaris

Anemia (mainly iron-deficiency anemia)

Angina pectoris

Chronic obstructive pulmonary disease—chronic bronchitis and emphysema without or without chronic bronchitis

Congestive heart failure

Diabetes mellitus

Hay fever

Hearing loss

Hemorrhoids

Hernia

Hypercholesterolemia

Hypertension

Hyperthyroidism and hypothyroidism

Joint disorders—acute joint problems, arthritis, gout, and rheumatoid arthritis

Peptic ulcer disease and dyspepsia

Urinary tract infection

Varicose veins

Vision impairment

In general, the self-reported information involved a question about past diagnosis made by a physician (or past advice from a physician or nurse that a diagnosis-specific sign, such as "sugar in the urine," was present) or a question or set of questions about diagnosis-specific symptoms. For individuals who indicated that such a diagnosis had been made or that such symptoms were present, the self-reported information included the following: presence or severity of related symptoms; whether medical care was obtained for the condition; whether medications or other therapies were prescribed by a physician; whether medications or other therapies were used, whether prescribed or not; and whether the individual experienced negative impacts of the illness (pain or distress, worry or concern, restrictions in usual activities, or days spent in bed).

For most of these conditions, laboratory or other diagnostic tests or procedures were carried out as part of the enrollment and exit screening examinations. Certain tests (marked with an asterisk in the list below) were not given to all examinees; age, sex, and reported symptoms determined who received the test. (Procedures are described more fully later in this appendix.) For adults, tests and procedures included the following:

Audiometry (pure-tone)

Blood tests—complete blood count (white blood cell, red blood cell, hemoglobin, hematocrit, mean corpuscular volume)

—bilirubin (enrollment only)

—blood urea nitrogen

—blood alcohol (enrollment only)

—glucose test (2-hour post load (enrollment only); random blood sugar (exit only))

—rheumatoid factor test*

—serum cholesterol

—serum glutamic-oxalacetic transaminase (SGOT) (enrollment only)

- serum iron if low hemoglobin (exit only)
- serum thyroxine (T4, resin T3 uptake, T7 or free thyroxine index)
- thyroid stimulating hormone if low T7 (exit only)
- uric acid

Blood pressure

Chest x-ray (enrollment only)

Electrocardiogram*

Height and weight

Joint measures*—joint size, 50-foot walk, grip strength
—hand/wrist x-ray (enrollment only)

Spirometry—forced vital capacity (FVC), forced expiratory volume in 1 second (FEV₁), FEV₁ percent.

Urinalysis—dipstick blood, protein, glucose, microscopic urinalysis; urine culture

Varicose vein examination*

Vision testing—near and far acuity with and without corrective aids, pinhole acuity; tonometry*

A physician was present at the exit screening examinations to obtain a medical history and conduct a physical evaluation on those enrollees who were known (from diagnoses on insurance claim forms) to have certain chronic conditions such as hypertension or thyroid disease.

GENERAL PROVISIONS OF THE MEDICAL HISTORY QUESTIONNAIRE

The MHQ was a two-part, self-administered instrument that enrollees were asked to complete at the time of enrollment and again at the time of exit. The adult version was given to all persons 14 years of age or older; assistance was provided for individuals who could not complete it by themselves because of eyesight problems or illiteracy. The MHQ was essentially the same for all sites at both enrollment and exit; the exception was the Dayton enrollment MHQ, which had early versions of some batteries. The individual volumes of the R-2262-HHS and R-1987-HEW Rand Report series give further information on the Dayton MHQ.

The first part of the MHQ, "Form A," contained (among other things) the batteries pertaining to self assessment of health status. Separate batteries pertained to physical health, mental health, social

health, and general health perceptions. Form A also included separate items on height and weight and a battery on smoking behavior. The second part of the MHQ, "Form B," contained the disease-specific batteries pertaining to blood pressure, cholesterol, and vision.

All individuals in all sites at both enrollment and exit received Form A. Because of study design and scheduling considerations, only a random sample of Dayton participants received Form B at enrollment; all participants in other sites at enrollment and in all sites at exit received Form B.

ATTRITION QUESTIONNAIRES

For individuals who left the study prematurely, we administered an "attrition" MHQ by mail. This data collection effort was carried out in 1982-1983, once field operations had ended. The sample included persons who left the experiment and persons who had been terminated for reasons of noncooperation or for health or other reasons (such as military service or incarceration).

The mailed questionnaire contained all the self-assessment and disease-specific batteries reported on in this monograph (i.e., both Forms A and B of the MHQ). Persons who did not complete the mailed questionnaire were asked to respond to certain key questions over the telephone. These pertained mainly to physical, mental, and social health, general health perceptions, and smoking. We did not ask about the other health conditions or weight over the telephone.

SCREENING EXAMINATION

Over the course of the experiment, two different subcontractors conducted the multiphasic screening examination. It was given in each site at a screening center consisting of seven testing stations plus reception and waiting areas. The medical portion was performed by trained paramedical personnel at enrollment (typically four medical assistants, three technicians, and a registered nurse supervisor); at exit, the staff included a physician (to evaluate selected patients with specific chronic conditions) as well as four registered nurses, a medical assistant, and an administrative supervisor. Extensive quality control measures (e.g., calibrating equipment, monitoring lighting and other environmental factors, and reviewing confidentiality procedures) were maintained. In addition, selected tests were repeated on a random basis. Smith et al. (1978) give further details on the screening examination.

For adults, the typical order of steps during the screening examination for persons eligible for the specific test was the following: obtain identification data, authorize release of findings, complete a medications sheet, complete a medical history abstract sheet (to ascertain indications or contraindications for later tests), height and weight, urine collection, glucose load, blood pressure, audiometry, chest x-ray, hand and wrist x-ray, electrocardiogram, spirometry, tympanometry, varicose vein examination, tests of joint problems (e.g., grip strength), acne photograph, vision testing, tonometry, venipuncture, and dental evaluation. Some tests were given only at enrollment (see above list) and others only at exit. At exit, glucose loads were not given, so blood specimens were drawn from participants when convenient.

Electrocardiograms, x-rays, and acne photographs were sent to board-certified physician consultants for interpretation. All test results were reviewed by a physician member of the screening examination subcontractor's staff.

At enrollment, this physician determined what type of notification letter would be sent to the participant and the participant's regular physician. Disregarding information on dental status, the main types of letters pertinent to the medical tests were the following: normal (all test results were within normal ranges); abnormal (at least one test fell outside normal ranges but no test was in the alert range); and alert/abnormal (at least one test in the alert range). The alert range was essentially any test result sufficiently outside of normal to suggest that the participant's health might be in immediate danger (e.g., a systolic blood pressure over 250 mm Hg); in this event, the enrollee's physician was telephoned and a follow-up letter was sent. The enrollee received a letter corresponding to one of the categories listed above; the physician received the letter and a copy of all test results. Similar procedures were followed at exit.

A random sample of participants was assigned to receive the screening examination at enrollment; all participants were assigned to receive it at exit. No standardized screening examination was given for persons who withdrew from the study prematurely; however, out-of-area participants were requested to go to their own physician at our expense, at which time many of the same physiologic data were collected.

Appendix C

DESCRIPTION OF GENERAL HEALTH MEASURES

PHYSICAL HEALTH AND ROLE FUNCTIONING

Among the more widely used measures of physical health are those that assess limitations in functioning in everyday life owing to poor health, including personal and role functioning (Stewart et al., 1977, 1978). For our analyses of the effects of cost-sharing on health status for adults, we chose two summary measures: (1) a 21-item measure of personal function that includes self-care, mobility, physical activities, and capacities (the Personal Functioning Index), and (2) a two-item measure of performance and capacity to perform one's usual role activities (the Role Functioning Index). These batteries were administered annually to all adults.

Content

The Health Insurance Experiment functional status items were derived from work by Patrick, Bush, and Chen (Patrick et al., 1973), the National Center for Health Statistics (NCHS, 1974), and Hulka and Cassel (1973). HIE items assessed limitations due to poor health regardless of duration. Although most limitations were present for more than three months, both acute and chronic limitations were captured.

Scoring

Associations among subscales measuring self-care, mobility, and physical activities were very strong and indicated a cumulative relationship between limitations in the various categories. In other words, persons unable to perform self-care tasks were also limited in mobility and more strenuous physical activities. Thus, from the perspectives of both the content of the measures and empirical studies of their associations, a summary index was justified. However, many persons with role limitations do not have limitations in personal functioning; others

are so limited. Hence, physical and role limitations could not be summarized in a single index without considerable loss of information.

Initially, separate scales were constructed to measure self-care, mobility, physical activity, role limitations, and physical capacities. Tests of scalability indicated that a Personal Functioning Index (PFI) could be constructed according to Guttman criteria by aggregating measures of self-care, mobility, physical limitations, and physical capacities (Stewart et al., 1981, 1982a, 1982b). The PFI defines six levels, transformed to a 0-100 scale to reflect differences in impact on health.¹ A score of 100 indicates complete freedom from limitations in self-care, mobility, and physical activity because of poor health.

The Role Functioning Index was actually developed from a negatively oriented Role Limitations Index. For this report, the Index was reversed and scored dichotomously. Thus, persons who were completely free of limitation in their usual role activities were assigned a score of unity on the Role Functioning Index. Those reporting one or more limitations in role performance because of poor health were assigned a score of zero. For results reported here, these scores were converted to percentages.

Reliability

In all studies of HIE physical health and role functioning measures, reliability coefficients were high, based on internal-consistency (reproducibility) and test-retest estimates (see Table C.1) (Stewart et al., 1978, 1982a). Substantial stability of scores across repeated measurements a year apart has also been observed (0.59 for the PFI and 0.50 for the Role Functioning Index across all sites).

Validity

The Personal and Role Functioning Indexes had good content validity. They well represented the content of functional status measures identified in the published literature (Stewart et al., 1978, 1982a). Their validity has also been well supported in empirical terms. Multivariate studies of correlations among the HIE functional status measures and other health and health-related measures strongly supported their discriminant validity as measures of physical functioning (Ware et al., 1980). They have also been shown to be predictive of general medical expenditures and mental health expenditures in models that

¹Scale level values were determined by estimating mean General Health Rating Index (GHRI) scores (described below) for persons at each of the six levels of PFI. The highest and lowest values were transformed to 100 and zero, respectively, and values in between were expressed as a proportion of the observed range in GHRI scores.

Table C.1

RELIABILITY AND STABILITY ESTIMATES FOR HIE GENERAL HEALTH STATUS MEASURES

Measure	k ^a	Reliability ^b	One-Year Stability ^c
Personal Functioning Index	21	0.97 ^d	0.59
Role Limitations Index	2	0.92 ^d	0.50
Mental Health Inventory	38	0.96	0.64
Social Contacts Scale	3	0.72	0.55
General Health Ratings Index	22	0.89	0.68

^aNumber of items.^bInternal-consistency reliability estimated by Cronbach's (1951) alpha, unless otherwise noted.^cProduct-moment correlation between scores obtained approximately one year apart.^dCoefficient of reproducibility.

control for a wide range of health-related variables (Manning et al., 1982; Ware et al., 1984a). Nonetheless, the majority of persons in a general nonaged population do not have measurable functional limitations; hence, the precision of these measures for testing hypotheses about effects of insurance plan on health status is somewhat restricted.

MENTAL HEALTH

The concept of mental health as surveyed in general population studies has evolved from rather confounded measures to ones that focus more exclusively on symptoms of psychological distress (Ware et al., 1979). A typical mental health measure fielded before 1970 aggregated physical symptoms, physical functioning, general health perceptions, health habits, and some symptoms of psychological distress. Consistent with a new generation of general population mental health survey instruments, the HIE Mental Health Inventory (MHI) measured psychological distress and psychological well-being and was purged of other health status concepts that were scored and interpreted separately. As with the other general, self-reported measures, the MHI was administered to all adults annually.

Content

In its first mental health survey during enrollment at the Dayton site, the HIE fielded a 22-item adaptation of Dupuy's General Well-Being Schedule (Veit and Ware, 1983), 33 experimental items, and 17 criterion items measuring history of emotional problems and treatment (Ware et al., 1979, 1980). Studies of the scalability of these items and tests of reliability and validity led to the development of the 38-item MHI.

The MHI retained 15 items from the Dayton enrollment instrument and added other items selected to increase the comprehensiveness of the instrument and the precision of some MHI subscales. These additional items were very similar to items in instruments developed by Zung (1965), Beck (1967), Costello and Comrey (1967), Comrey (1970), and Dohrenwend et al. (1980). Some items were selected from Dupuy's experimental items; others were constructed anew (Ware et al., 1984b).

The 38 MHI items well represented the most prevalent symptoms of psychological distress (such as anxiety and depression), loss of behavioral/emotional control, and a general positive affect. Each item asked about a psychological construct during the past month and was associated with a six-choice response scale that varied the frequency and/or intensity of that construct (e.g., "How much of the time, during the past month, have you felt downhearted and blue?" with six choices ranging from "all of the time" to "none of the time").

Scoring

The MHI was constructed to yield an overall Mental Health Index score, two summary scales (psychological distress, psychological well-being), and five subscales: anxiety, depression, general positive affect, self-control, and emotional ties. The Index is a summated ratings measure transformed to a 0-100 scale; better mental health has a higher score.² Confirmatory factor analyses performed independently for each HIE site provided a strong psychometric basis for the MHI summary score (Viet and Ware, 1983).

²Continuous scales were transformed to a 0-100 range using a linear transformation algorithm that defines the highest possible score as 100 and the lowest possible score as zero. The algorithm is as follows:

$$\text{Transformed score} = (\text{Actual raw score} - (\text{minimum possible raw score})) / (\text{Maximum possible raw score} - (\text{minimum possible raw score})) * 100.$$

The transformed score was then rounded to the nearest decimal.

Reliability and Stability

The reliability of the MHI was high, 0.96 in the combined-sites HIE sample (N = 5089), when estimated using internal-consistency methods (see Table C.1). These coefficients varied across HIE sites, with the lowest reliability being 0.93 in site-specific analyses.

The stability of the MHI summary scale has been estimated for intervals of one, two, and three years between administrations. One-year stability coefficients ranged from 0.60 to 0.76 for adults across sites and different one-year intervals in the HIE. Two-year stability coefficients for adults (across sites) ranged from 0.54 to 0.69 and three-year estimates from 0.54 to 0.58.

Validity

Several strategies have been employed to evaluate the validity of the HIE mental health measures. The first studies focused on the relationships among the mental health subscales and their associations with other conceptually related variables also measured by self-administered questionnaire. The subscales were substantially intercorrelated but the intercorrelations were well below their reliability coefficients. Thus, it is meaningful to aggregate the subscales to define the MHI. Subscales can differ in interpretation, as is reflected in differing correlations with other variables (Ware et al., 1980).

Factor analytic studies of the mental health subscales and numerous other health and health-related variables demonstrated that they are distinct from physical and social factors (Ware et al., 1980). Further, as would be expected for valid measures, they have been linked to life events, social contacts and resources, chronic diseases, acute physical symptoms, and general health perceptions (Davies and Ware, 1981; Donald and Ware, 1982; Manning et al., 1982; and Williams et al., 1984).

More recent HIE validity studies have focused on data gathered from sources other than the questionnaire containing the mental health items. For example, controlling for other health and attitudinal variables, we showed that the mental health subscales significantly predicted subsequent general medical expenditures (Manning et al., 1982) and outpatient mental health expenditures (Ware et al., 1984a).

SOCIAL CONTACTS

Measures of social contacts extend the concept of health beyond the physical and mental status of the individual to include the quantity of visits and other contacts with friends and relatives. Measures of these visits and contacts reflect social functioning (Donald et al., 1978; Ware et al., 1981; Donald and Ware, 1982, 1984). For the HIE, we developed an 11-item battery that was administered annually; from this was excerpted a 3-item Social Contacts Scale used in this report.

Content

The social well-being items were adapted from measures used by Myers and his colleagues (Myers et al., 1972) in their studies of social activity, life events, and mental status and by Dohrenwend and his colleagues (Dohrenwend et al., 1973) in their studies of role functioning of psychiatric patients. Three items measuring social contacts were selected from an 11-item battery to represent the social well-being concept. These items measured the frequency with which study participants got together with friends or relatives, visited with friends at their homes, and visited with friends at their friends' homes. Each item was accompanied by six or seven standardized response categories defining frequency of occurrence.

Scoring

Using the criterion approach, we estimated the level of health and general well-being associated with responses to each of the three social contacts items (Donald and Ware, 1982). Responses for each item were compared with criterion measures of self-rated health, emotional ties, and general positive affect to determine scoring rules. After developing scoring rules, we standardized items to take into account differing item variances. The standardized items were then summed to score a Social Contacts Scale. For HIE purposes, we transformed the final Social Contacts score to a 0-100 scale to facilitate comparisons of scores across scales; this transformation used the same algorithm as for the MHI (see footnote 2).³

³A different approach was used at enrollment at the Dayton site, because the 11 social well-being items were not fielded at enrollment in Dayton. To improve the precision of our analyses, we developed an algorithm for predicting Social Contacts at enrollment in Dayton. Using the Dayton Social Contacts score at the end of the experiment's first year, we used regression analyses to choose the best set of predictor variables. The final set of predictors included three sociodemographic variables (age, sex, race) as well as four items from the questionnaires self-administered at enrollment, namely, trouble getting

Reliability and Stability

Internal-consistency estimates indicated satisfactory reliability for the Social Contacts Scale (see Table C.1). Coefficients ranged from 0.70 to 0.72 for adults ($N = 4565$) across HIE sites. As expected, stability coefficients (based on a one-year interval between repeated administrations) for the Social Contacts measure were substantially lower than internal-consistency estimates and ranged from 0.44 to 0.57 across sites.

Validity

Items in the Social Contacts Scale satisfied our discriminant validity criteria: The items correlated more highly with the specific social well-being construct they were intended to measure than with other health constructs measured in the same way. Correlations between Social Contacts and other HIE social well-being items and measures were significant and positive, supporting the validity of the scale. The correlations were low, however, suggesting that social well-being constructs were only weakly related. The magnitude of positive associations between social contacts and psychological well-being suggested that the scale may be more predictive of positive emotional states and general views about life than of negative emotional states or general states of health. Findings from a longitudinal analysis of HIE data suggested a direct positive effect of social contacts on mental health (Williams et al., 1981). This effect has been demonstrated for men and women separately, as well as for employed and unemployed men and women (Williams et al., 1984).

GENERAL HEALTH PERCEPTIONS

Self-ratings of general health are among the more commonly fielded measures of health status (Ware, 1976).⁴ An example is the rating of health in terms of "excellent," "good," "fair," or "poor." These ratings are considered measures of *general* health because they do not focus on a specific health status attribute and because they have been linked

along with others at work; under stress or tension at home; felt cheerful, lighthearted; married at least part of the time during the last six months. Diagnostic tests of the model's performance showed that it was an improvement over other, more crude methods to estimate an initial score (e.g., the use of an overall mean score). Predicted values were computed for all adults at Dayton enrollment who had actual values for the predictor variables.

⁴An edited version of this summary will appear in Ware (1984).

empirically to a wide range of physical and mental health concepts as well as health and illness behaviors. They are considered *ratings* as opposed to reports because they reflect individual differences in how people evaluate the information they have about their health.

Content

The HIE administered 29 general health rating items to all study participants annually (Davies and Ware, 1981); 26 were taken from the Health Perceptions Questionnaire (HPQ) developed by Ware and Karmos in work supported by the National Center for Health Services Research (Ware and Karmos, 1976; Ware et al., 1978). HPQ items were worded as complete statements of opinion about personal health (e.g., "I have been feeling bad lately" and "I expect to have a very healthy life") and were accompanied by five standardized response categories defining a true-false continuum. These HPQ items were used to score six subscales that assessed the following dimensions of health perceptions: three time-bounded perceptions of health (past, present, and future), health-related worry and concern, resistance vs. susceptibility to illness, and the tendency to view illness as a part of life.

Scoring

Twenty-two HPQ items were used to compute the General Health Rating Index (GHRI), a favorably scored summary of health perceptions (Davies and Ware, 1981). The GHRI was scored using a simple summated ratings method. For this report, the GHRI was transformed to a 0-100 scale, with higher scores representing a better rating of one's health. (See footnote 2 above for the transformation formula.) In a general population, scores are roughly symmetrically distributed; at enrollment, for example, one or more adults was observed at 80 of the 89 GHRI levels.

Reliability and Stability

Internal-consistency estimates indicated good reliability for the GHRI (see Table C.1). Coefficients ranged from 0.88 to 0.90 for adults ($N = 3521$) across HIE study sites and across groups differing in educational attainment. Test-retest reliability estimates (two- to six-week intervals) were somewhat lower than internal-consistency estimates.

We have estimated the stability of GHRI scores for HIE adults for time intervals of one, two, and three years between administrations.

These results vary somewhat by HIE site and years of the experiment, but median stability coefficients for intervals of one, two, and three years are, respectively, 0.66, 0.59, and 0.56. Hence, a substantial gain in precision can be realized by using the GHRI in a before-after study design as opposed to an after-only design.

Validity and Generalizability

The validity of the GHRI and its subscales has been substantiated in numerous investigations over the past 10 years. Norms are available for aged and nonaged general adult populations (Davies and Ware, 1981; Ware et al., 1978). Cross-sectional and longitudinal analyses have been conducted for general and targeted populations to address questions about: (1) the kinds of individual differences in health status measured; (2) similarities and differences in results compared with other widely used health status measures; (3) sensitivity to the impact of different diseases; (4) sensitivity to individual differences in disease severity; (5) sensitivity to changes in health status over time, owing to treatment, aging, and other factors; (6) predictive validity in relation to use of medical services, future morbidity, and survival; and (7) who, in sociodemographic terms (e.g., age, sex), scores high and low on the GHRI and its subscales.

Initial validity studies examined the factor structure of the HPQ and correlations between HPQ subscales and other survey measures of health status and health-related behaviors in four general population studies. The factor structure of the HPQ was very similar across populations, indicating that the HPQ had construct validity; it included six correlated health perceptions factors (the basis for the six HPQ health subscales). The HPQ subscales correlated substantially with survey measures of physical and mental health and with self-reported use of health services (Ware et al., 1978; Davies and Ware, 1981). A pattern of age and sex differences has been consistently replicated: Women and older persons score lower on the GHRI.

Subsequent validity studies provided the psychometric basis for the GHRI summary score. They examined correlations between the GHRI (and HPQ subscales) and 35 measures of health status and health-related validity variables using data from all HIE sites (Davies and Ware, 1981). These studies expanded the list of validity variables measured by self-report, and they also added measures of chronic disease based on clinical examinations and subsequent use of health services reported on HIE insurance claim forms. Results indicated that the GHRI is a sensitive indicator of individual differences in disease status, limitations in physical and role functioning owing to

poor health, acute physical and psychosomatic symptoms, and mental health (symptoms of psychological distress and well-being). (Correlations between the GHRI and measures of social contacts and resources were significant but very weak.) Other HIE analyses have shown that the GHRI is substantially related to subsequent use of general medical and mental health services in the ambulatory system and that it has considerable incremental validity in these predictions, i.e., it adds information not contained in the other health status measures listed above (Manning et al., 1982; Ware et al., 1984a).

Validity studies now in progress will increase understanding of the clinical and social relevance of individual differences in scores for the GHRI and its subscales and the meaning of changes in scores over time. Preliminary findings suggest that the GHRI discriminates between those with and without a chronic disease (e.g., hypertension, arthritis, chronic obstructive airway disease, diabetes). Moreover, it is sensitive to individual differences in disease severity and to changes over time in both physical and mental health.

Appendix D

DESCRIPTION OF PHYSIOLOGIC AND HEALTH HABITS MEASURES

INTRODUCTION

This appendix provides information for all remaining measures reported on in the main text. Further information about these measures and about the clinical definitions of specific conditions can be found in the following reports: blood pressure (Brook et al., 1980); cholesterol (Brook et al., 1981); vision (Rubenstein et al., 1982); smoking (Stewart et al., 1979); and weight (Stewart 1982; Stewart et al., 1980). As explained in Appendix B, the self-administered Medical History Questionnaire (MHQ) was given to all adults at the start and at the completion of their HIE enrollment period. The multiphasic screening examination was administered to a randomly selected 60 percent of the sample at enrollment (see Smith et al., 1978) and to all participants at exit.

BLOOD PRESSURE

Measurement

Screening examination. At enrollment in Dayton, sitting and standing systolic and diastolic blood pressures in the participant's left arm were measured by a technician with an Infrasonde 3000 electronic sphygmomanometer. At enrollment in all other sites, sitting blood pressures were taken with a Physiometrics Automatic Blood Pressure Recorder. The hard-copy record produced by this instrument was read by a technician and verified by a registered nurse. Adult and large adult cuffs were available in all sites.

If either the systolic or diastolic sitting blood pressure was elevated (≥ 140 mm Hg or ≥ 90 mm Hg, respectively), sitting measurements were repeated after at least 10 minutes of rest. If any examinee was taking antihypertensive medications (as determined by a response to questions asked at the screening examination), a standing measurement was also obtained.

At exit in all sites, blood pressure was measured with a Model 2230 electronic Bloodpressure Station. Sitting measures were obtained from all enrollees; standing and/or repeat sitting measures after a 10-minute rest were obtained from any examinee who met the same criteria as above.

Medical history questionnaire. At enrollment in all sites except Dayton and at exit in all sites, the MHQ contained a Blood Pressure battery of questions about blood pressure checks, physician diagnosis, prescription and use of hypertension medications, use of salt, diet, recency of physician visit about high blood pressure, and impact of high blood pressure on the person's life. Only persons who responded "yes" to the question "Has a doctor ever said that you had high blood pressure?" were instructed to answer the remaining questions in the battery; all other persons skipped out. The Dayton enrollment battery was an earlier, less detailed version of this battery.

Reliability and Validity of Measures

Reliability of blood pressure measurements. A randomly selected subsample numbering 191 individuals received a second set of blood pressure readings at the close of the enrollment screening examination. Compared with the original set of readings, both systolic and diastolic readings showed some between-measurement variation; second measures tended to be lower, as would be expected from clinical experience. The mean decrease between the first and second sets of systolic measurements was 2.3 mm Hg, with a standard deviation of 7.5 mm Hg. The mean decrease for diastolic measurements was 1.8 mm Hg, with a standard deviation of 6.1 mm Hg.

Validity of MHQ battery. Using the screening examination measure as "truth," we assessed the validity of the MHQ by calculating its sensitivity and specificity for correctly classifying enrollees as to hypertension status. The sensitivity (for accurately identifying persons as hypertensive) was 0.71, with a specificity (for correctly identifying persons as normotensive) of 0.93.

CHOLESTEROL

Measurement

Screening examination. In all sites, cholesterol was part of the SMA 12/60 panel. Measurement techniques differed among the sites and between enrollment and exit because local laboratories were used to process screening examination tests (see Appendix B). They did not

differ by plan, so this problem should not bias the analysis of plan effects. Site- and time-specific details are as follows:

At enrollment in Dayton, Seattle, and the South Carolina sites, the Lieberman-Burchard test without extraction was used. In Dayton and Seattle, the Huang modification with Technicon reagent was used; in South Carolina, the Hycel reagent. In both Massachusetts sites, the enzymatic oxidase technique was used. For all sites, any enrollee with a value higher than normal for the laboratory performing the test received a repeat determination, and the second value was used in all analyses. The enzymatic oxidase method gives systematically lower readings than does the Lieberman-Burchard method.

At exit in Dayton (three-year enrollees), Seattle and Massachusetts (five-year enrollees), and South Carolina (three- and five-year enrollees), the enzymatic oxidase method was used. At exit for Dayton five-year enrollees and for three-year enrollees in Seattle and Massachusetts, the Lieberman-Burchard technique without extraction was used (performed on the SMAC machine with Technicon reagent). Repeat determinations were performed for any individual whose cholesterol value exceeded 260 mg/dl (milligrams per deciliter), and the second (i.e., repeat) value was used in all analyses.

Medical history questionnaire. The Cholesterol battery in the MHQ included questions about previous cholesterol tests, physician diagnosis of and care for high cholesterol, use of medications or special diet, and the impact of hypercholesterolemia on the individual's life. Only persons who responded "yes" to "Has a doctor ever said you have high blood cholesterol?" were instructed to answer the remainder of the items in this battery.

Reliability and Validity of Measures

Reliability of cholesterol test. A random sample totaling 318 persons at the enrollment screening examination had extra blood drawn at venipuncture. The blood was split into two samples, independently labeled, and sent to the laboratory in different batches so that they would be processed at different times. Comparisons were made of the two cholesterol determinations as an indicator of the reliability of this measure. The mean difference between the first and second determinations was -1.83 mg/dl. The standard deviation, 13.79 mg/dl, was close to the upper range (about 11 mg/dl) of a sample of "lipid laboratories" studied by the Centers for Disease Control at about the same period.

Validity of MHQ battery. Data on 126 individuals were available to study the validity of the MHQ in terms of its sensitivity and specificity for correctly classifying people as to their cholesterol status at

enrollment. MHQ sensitivity was estimated to be 0.32, with a corresponding specificity of 0.99.

VISION

Measurement

Screening examination. At enrollment and exit in all sites, measurements of visual acuity were obtained for each eye separately. For far vision, illiterate examinees were tested with the Illiterate E Chart and all other examinees with the Snellen Eye Chart. For near vision, the Rosenbaum Near Vision Card was used for all individuals. If the examinee wore glasses or contact lenses, both the "corrected" vision (i.e., with the available vision aid) and uncorrected (natural) vision were tested. Persons with contact lenses were given 10 minutes for adjustment of eyes before testing of natural vision was done. The medical assistant recorded the type of corrective aid brought by the examinee to the test.

Far-vision testing began with the examinee asked to read the line equivalent to 20/40. If more than one letter was missed, the examinee was asked to read the next line up (of larger print) until only one letter on a line was missed. If one or no letter was missed on the 20/40 line, the examinee was asked to proceed down the chart (smaller print) until more than one letter was missed or the examinee reached the line equivalent to 20/15. A test of pinhole acuity was given to examinees with far vision worse than 20/20 who did not wear corrective aids and to examinees with corrected far vision worse than 20/20. Testing was done at 15 feet at enrollment (with results converted to the 20/20 Snellen scale) and at 20 feet at exit.

Near-vision testing proceeded in a manner similar to far-vision testing. Both corrected and uncorrected vision were tested in each eye separately. Level of acuity was determined from the smallest line the examinee could read with no more than one error in character identification. Testing was done at 14 inches from the eye at enrollment and at 16 inches at exit, because of a change in the card used.

Medical history questionnaire. The main Eyesight battery in the MHQ consisted of a series of questions about eye tests, use of corrective lenses, ability to do simple functional tasks (read a newspaper, recognize a friend across the street), and impact of vision impairment on daily life. The battery had no introductory "skip" question, so all enrollees were expected to answer all questions that pertained to them.

Reliability and Validity of Measures

Reliability of vision testing. At enrollment, 112 of 3313 persons tested with complete far-vision data were retested for far vision and 108 of 3314 persons with complete near-vision data were retested for near vision. Reliability was assessed as the degree of agreement between the first and second sets of results for vision in the *better* eye, with or without glasses. Acuity values were converted to corresponding line numbers from the eye chart ($20/15 = 1$, $20/20 = 2$, etc.) for the analysis.

For far vision, the difference between the first and second measurements was never greater than two lines; the average difference was 0.16 lines with a standard deviation of 0.46. For near vision, the largest difference was three lines, the average difference, 0.17 lines, and standard deviation, 0.50.

Approximately 4 percent of enrollees apparently had worse vision with lenses than without them. For about half of this group, this problem could be attributed to inappropriate testing with lenses (e.g., using "reading glasses" for far-vision testing). In all analyses involving "corrected" (i.e., functional) vision, acuity-without-lenses was substituted as a measure of functional vision whenever that value was better than the acuity-with-lenses value.

Validity of MHQ questions. The validity of the task-related items in the Eyesight battery was evaluated in terms of sensitivity and specificity for correctly labeling persons as to whether they had near- or far-vision impairments, using the screening examination results as "truth." The items had poor sensitivity (with high specificity) except at quite marked levels of impairment and were judged to be invalid measures of vision deficit.

SMOKING

Measurement

Medical history questionnaire. The Smoking battery appeared on the MHQ at enrollment at all sites except Dayton and all sites at exit (the Dayton enrollment version was shorter and less detailed). As with the general health measures described in Appendix C, the battery was administered more often than at enrollment and exit, namely, at the end of the study's second year in each site, at the end of the study's third year in Dayton (for five-year enrollees only), and at the end of the fourth year in all other sites (for five-year enrollees only). The

battery included ten items with categorical responses; some items were asked only if the enrollee had given a specific response (typically, "yes") to the immediately preceding item.

At exit, a smoking supplement was included in the MHQ in all sites. It asked about type of cigarette (filter or nonfilter), size, and brand usually smoked.

From these data, three primary dependent variables were constructed:

Cigarette smoking status: nonsmoker, ex-cigarette smoker, and current cigarette smoker.

Current amount smoked: average number of cigarettes smoked per day at one of four levels (10, 20, 40, or 50 cigarettes).

Overall smoking status: an estimate of the future risk of mortality from all causes (based on pre-1960 cigarettes). The ratio values are given in Table 2 of the main text.

A number of other measures were also available from this battery, including type of tobacco smoked; length of time smoked and number of pack-years smoked for current cigarette smokers; for ex-cigarette smokers, amount smoked when smoking, length of time smoked, number of pack-years, and length of time stopped; and whether a doctor had advised the person to stop or cut down smoking.

Reliability and Validity

The reliability and validity of the HIE smoking measures were not directly assessed. Various lines of evidence, such as findings from the literature, overall high quality of the HIE data from the MHQ, absence of "socially desirable response set" bias, and similarity of the HIE distribution to that from a national probability sample, supported the conclusion that the HIE smoking measures had sufficient reliability and validity for testing group differences. The one-year stability coefficient for the smoking index was estimated to be 0.90.

WEIGHT

Measurement

Medical history questionnaire. The MHQ in all sites at both enrollment and exit asked one item on height ("How tall are you without shoes on?") and weight ("How much do you weigh without heavy clothes?"). Enrollees were to respond by writing in the appropriate

feet and inches or pounds (respectively). Height was converted to meters by multiplying height in inches by 0.0254; weight was converted to kilograms by multiplying weight in pounds by 0.4536. From these data, numerous measures of weight adjusted for height and relative recommended weight were constructed. For the analyses in this report, a measure that relates weight (W) to a power of height (H) was used: Men: $W/(H^2)$; Women: $W/(H^{1.5})$.

The MHQ also contained a separate Weight battery, with the majority of questions directed only at persons who answered "yes" to the question "Do you think you weigh too much now, or has a doctor or nurse told you that you now weigh too much?" Remaining questions dealt with efforts to lose weight and impact of being overweight on the person's life. This battery was administered at enrollment and exit in all sites (with some minor differences in the Dayton enrollment version).

Screening examination. Weight in pounds and height in inches were obtained twice on all individuals who received the enrollment and exit screening examinations: an initial measure done early in the screening examination and an *immediate* repeat measurement. Weight in pounds and height in inches were computed as a rounded average of these two values; if the two values differed by more than 3 percent of the average, the person was given a missing value (which occurred in less than 1 percent of the cases). These averaged values were then converted to kilograms and meters (respectively) as with the MHQ data.

All measures constructed from MHQ data were also constructed from screening examination data. For analyses in this report, the MHQ data were used so that the analytic sample would not need to be restricted to the random sample who took the enrollment screening examination. In cases where one self-report value did not look reasonable (e.g., if person reported himself or herself as more than 50 percent overweight or underweight), we used information from either the screening examination or the other self-report.

Reliability and validity of MHQ data. We assessed the reliability and validity of the self-reported weight and height data in numerous analyses (see Stewart, 1982). For example, we tested the correlation of the self-reported value with the "true" value (from the screening examination) and estimated the mean of the differences between the self-reported and true values. Overall, self-reported weight and height were very accurate measures of actual weight and height. From the enrollment data, we estimated average under-reporting of weight to be 2.4 pounds and average over-reporting of height to be 0.6 inches. The correlation of self-reported weight with measured weight was 0.99; the correlation for height was 0.97.

RISK OF DYING

This index is derived from systolic blood pressure, cholesterol, and smoking measures. Values are calculated according to a specific formula based on findings from several major epidemiologic studies of cardiovascular disease (see McGee and Gordon, 1976). The formula is as follows:

$$100 \exp(\text{Index}) / (1 + \exp(\text{Index})), \text{ where Index is equal to:} \\ 1.28 (\text{smoking scale}) + 0.0023 (\text{cholesterol}) + 0.023 \\ (\text{systolic blood pressure}) - 4.92.$$

We carried out no reliability or validity analyses on this measure.

Appendix E

DETAILED RESULTS OF REGRESSION EQUATIONS FOR GENERAL HEALTH MEASURES AND SMOKING

INTRODUCTION

This appendix presents the regression equations for the five general health measures and the smoking index. The five general concepts and their related technical variable names are as follows: physical functioning, Personal Functioning Index; role functioning, Role Limitations Index; mental health, Mental Health Inventory; social contacts, Social Contacts Scale; and health perceptions, General Health Ratings Index. The opening section defines all the "main-effects" (i.e., single) explanatory variables and all interactions; some of these also appear in the equations that will be described in Appendix F. The second section and its tables give the variable coefficients, associated t-test values, and other quantitative information about the equations.

DEFINITION OF VARIABLES USED IN REGRESSION EQUATIONS

This section gives the names of the variables used in the regression analyses together with a brief definition. The shortened version of a variable name is given in parentheses in some instances to simplify the designations of the interactions. The interactions are defined after the single variables; the * indicates multiplication.

Because some types of variables are "dummy" or "indicator" variables, there is by definition a set of "omitted" variables. These are the following: for sex, female; for insurance plan, free care; for site, Seattle; and for term of participation, five years. Thus, comparisons implicitly refer to persons with these characteristics (women, on the free plan, in Seattle, and enrolled for five years). However, comparisons cannot be interpreted without reference to the relevant interactions. Unless otherwise specified, dummy variables are scored 1 if "yes" and 0 otherwise.

<i>Variable Name</i>	<i>Variable Definition</i>
Catastrophic (Catast.)	A dummy variable indicating whether the participant was assigned to one of the three income-related catastrophic plans.
Intermediate (Interm.)	A dummy variable indicating whether the participant was assigned to one of the nine intermediate coinsurance plans.
Individual Deductible (ID)	A dummy variable indicating whether the participant was assigned to the individual deductible plan.
Physical Functioning Role Functioning	The Personal Functioning Index measured at enrollment. A dichotomous index related to ability to perform role activities and expressed as a percentage; it is calculated as $100 - 100(\text{Role Limitations})$, where Role Limitations is scored 1 if the participant reported any role limitations and 0 otherwise. Role Functioning is thus reported as positively scored to facilitate its interpretation in relation to the other general health measures.
Mental Health	The Mental Health Inventory measured at enrollment.
Social Contacts	The Social Contacts Scale measured at enrollment. (This variable appears only in the regression for exit Social Contacts.)
Health Perceptions Smoking	The General Health Ratings Index measured at enrollment. The value of the Smoking Index at enrollment. See Table 2 in the text for actual scoring. (This variable appears only in the regression for Smoking at exit.)
Age	The age of the participant at the date of enrollment.
Male	A dummy variable indicating whether the participant was male (scored 1 if male, 0 if female).
Income	Participant's family income as measured at baseline (three to nine months before enrollment). This value is computed by (1) standardizing the family's reported income for the two years before baseline to 1974 dollars using cost-of-living adjustments, (2) correcting for intersite differences in the cost of living, (3) adding \$1000, (4) dividing by a family-size adjustment factor (this factor is 1 for a family with two adults and two children), and (5) taking the natural logarithm of this value. To report income in this report in 1982 dollars, we multiplied 1973 dollars by 2.17, which represents the ratio of the Consumer Price Index for an urban family for those two years (1982, 289.1 CPI; 1973, 133.1 CPI). In all interactions, income is measured in its log form and centered at its mean of 9.33 (roughly, \$11,270).
Three-Year Term (Term3)	A dummy variable indicating whether the participant had a three-year term of enrollment (scored 1 if three years, 0 if five).
Took Physical	A dummy variable indicating whether the participant took the physical screening examination at enrollment.
Dayton	A dummy variable for participants in the Dayton, Ohio, site.

Fitchburg	A dummy variable for participants in the Fitchburg, Massachusetts, site.
Franklin	A dummy variable for participants in the Franklin County, Massachusetts, site.
Charleston	A dummy variable for participants in the Charleston, South Carolina, site.
Georgetown	A dummy variable for participants in the Georgetown County, South Carolina, site.
"Health"	A generic term referring to the enrollment health measure that corresponds to the dependent variable for any given regression equation. The "health" variables are centered at their means: 89.04 for Personal Functioning; 0.0720 for Role Limitations; 74.37 for Mental Health; 72.42 for Social Contacts; 70.15 for Health Ratings; and 1.298 for Smoking.
Income*Catast.	The interaction between centered income and Catastrophic.
Income*Interm.	The interaction between centered income and Intermediate.
Income*ID	The interaction between centered income and Individual Deductible.
Health*Catast.	The interaction between the centered initial health variable and Catastrophic.
Health*Interm.	The interaction between the centered initial health variable and Intermediate.
Health*ID	The interaction between the centered initial health variable and Individual Deductible.
Health*Income	The interaction between the centered initial health variable and centered income.
Health*Term3	The interaction between the centered initial health variable and the three-year term. This variable models an intertemporal correlation that weakens over time.
Health*Income *Catast.	The three-way interaction between the centered initial health variable, centered income, and Catastrophic.
Health*Income *Interm.	The three-way interaction between the centered initial health variable, centered income, and Intermediate.
Health*Income *ID	The three-way interaction between the centered initial health variable, centered income, and Individual Deductible.

RESULTS OF REGRESSION ANALYSES FOR FINAL RESULTS

We used a standard linear regression model for all dependent variables except Role Functioning. For the four general measures, a higher score means better health. For Role, we used a standard logistic ("logit") regression model (namely, maximum likelihood logit). The

actual regression runs used the variable Role Limitations, for which a higher score means worse health. A transformation was later made to reverse the scoring, so that results for this measure could be interpreted in the same way as the other general measures. For Smoking, a higher score is consistent with higher consumption of cigarettes.

In the remainder of this appendix, we present the detailed results of the regression analyses. Each table includes the variable name, the coefficient and t-test value of that coefficient for each variable in each of the six equations, and certain other information about the equation (number of observations, R-squared value, and residual standard error (standard error of the error term)).

Note that for health variables such as Personal Functioning or Social Contacts, the relevant value for the *explanatory* variable is the one at enrollment.

CORRECTIONS FOR INTRAFAMILY CORRELATION

The standard errors and the t-tests were computed using Huber's (1967) formula for the variance of a robust regression. To apply Huber's formula, we considered the family as the unit of observation and linear regression on individuals as an M-estimator. (An M-estimator is a type of robust estimator.) Linear regression is not the maximum likelihood estimator because individuals within a family have correlated responses. To overcome this, we calculate $R_2 R_1 R_2$, which is an asymptotically consistent estimate of the covariance matrix of the regression parameters, regardless of the form of intrafamily correlation or heteroskedasticity (nonconstant variance of the error term).

R_1 and R_2 are defined as follows:

$$R_1 = \sum_{\substack{\text{across} \\ \text{families}}} (\sum_{\substack{\text{within} \\ \text{family}}} X_i r_i)' (\sum_{\substack{\text{within} \\ \text{family}}} X_i r_i)$$

$$R_2 = (X'X)^{-1} \sigma^2.$$

X_i stands for the matrix of observed data and r_i for the vector of residuals for family member i .

Table E.1
REGRESSION COEFFICIENTS, t-TEST VALUES, AND OTHER EQUATION
RESULTS: PERSONAL FUNCTIONING, ROLE LIMITATIONS

Explanatory Variable and Other Measures	Dependent Variable Equation			
	Personal Functioning		Role Limitations ^a	
	Coeff.	t-Test	Coeff.	t-Test
Intercept	74.51	24.48	-0.708	-1.48
Catastrophic	0.744	0.68	-0.0305	-0.15
Intermediate	-0.267	-0.27	0.0742	0.40
Individual Deductible	-0.359	-0.35	0.1416	0.76
Personal Functioning	0.474	9.56	-0.0164	-5.35
Role Functioning	Not Used		1.429	3.81
Mental Health	0.0162	0.47	-0.0003	-0.06
Social Contacts	Not Used		Not Used	
Health Perceptions	0.306	8.88	-0.0321	-5.96
Smoking	Not Used		Not Used	
Age	-0.445	-12.66	0.0415	7.65
Male	3.673	4.94	-0.0771	-0.54
Income	-0.975	-0.86	-0.0570	-0.25
Three-Year Term	1.126	1.33	0.0805	0.50
Took Physical	0.576	0.74	-0.0815	-0.60
Dayton	-1.587	-1.35	0.284	1.36
Fitchburg	-2.185	-1.54	0.247	1.05
Franklin County	1.595	1.34	-0.234	-0.96
Charleston	0.930	0.71	0.044	0.19
Georgetown County	2.116	1.68	0.059	0.28
Income*Catast.	3.753	2.10	-0.388	-1.24
Income*Interm.	3.430	2.07	-0.0165	-0.06
Income*ID	-0.0139	-0.01	0.0427	0.14
Health*Catast.	-0.0631	-1.04	-0.569	-1.14
Health*Interm.	-0.0419	-0.77	-0.130	-0.32
Health*ID	0.0606	1.18	-0.989	-2.23
Health*Income	0.1334	2.16	-0.219	-0.41
Health*Term3	-0.0732	-1.66	-0.274	-0.80
Health*Income*Catast.	-0.1281	-1.52	1.367	1.64
Health*Income*Interm.	-0.0138	-1.61	0.665	0.96
Health*Income*ID	-0.0196	-0.25	0.032	0.05
Sample size	3862	—	3861	—
R-squared or Chi-square	0.3189	—	539.8 (28 degrees of freedom)	—
Residual standard error	22.889	—	NA ^b	NA ^b

^aLogit regressions (maximum likelihood logit) were run on the variable Role Limitations, for which 0 means no limitations and 1 means any limitations. Thus, signs on the coefficients are interpreted opposite the way they are interpreted for the other four general health measures (i.e., for Role Limitations, a negative sign means better health).

^bNot applicable.

Table E.2
REGRESSION COEFFICIENTS, t-TEST VALUES, AND OTHER EQUATION
RESULTS: MENTAL HEALTH, SOCIAL CONTACTS

Explanatory Variable and Other Measures	Dependent Variable Equation			
	Mental Health		Social Contacts	
	Coeff.	t-Test	Coeff.	t-Test
Intercept	65.037	36.63	74.12	20.06
Catastrophic	0.111	0.18	-0.128	-0.10
Intermediate	0.076	0.13	0.747	0.60
Individual Deductible	0.309	0.51	0.404	0.32
Personal Functioning	0.0287	2.80	0.0090	0.42
Role Functioning	Not Used		Not Used	
Mental Health	0.3607	8.88	0.0474	1.22
Social Contacts	Not Used		0.3546	7.91
Health Perceptions	0.0917	4.76	0.0289	0.77
Smoking	Not Used		Not Used	
Age	0.0200	1.20	-0.3191	-8.23
Male	0.975	2.67	-1.956	-2.62
Income	0.589	0.94	-0.753	-0.50
Three-Year Term	0.008	0.02	1.77	1.70
Took Physical	-0.100	-0.23	-1.283	-1.35
Dayton	0.072	0.11	-0.72	-0.47
Fitchburg	0.410	0.56	3.30	2.29
Franklin County	0.276	0.38	-0.263	-0.19
Charleston	1.350	1.67	0.425	0.28
Georgetown County	0.528	0.77	-1.293	-0.83
Income*Catast.	-0.724	-0.74	2.31	1.02
Income*Interm.	-0.200	-0.22	1.06	0.49
Income*ID	-0.460	-0.47	1.88	0.86
Health*Catast.	0.0994	1.94	-0.0292	0.58
Health*Interm.	0.0719	1.57	0.0255	0.55
Health*ID	0.0238	0.77	0.0694	1.43
Health*Income	0.1062	2.12	0.1026	1.84
Health*Term3	0.0645	1.74	0.0674	1.64
Health*Income*Catast.	-0.0442	-0.56	-0.0068	-0.08
Health*Income*Interm.	-0.0306	-0.42	0.0362	0.45
Health*Income*ID	-0.0060	-0.07	-0.0592	-0.70
Sample size	3862	—	3827	—
R-squared	0.2792	—	0.2066	—
Residual standard error	11.74	—	25.296	—

Table E.3
REGRESSION COEFFICIENTS, t-TEST VALUES, AND OTHER EQUATION
RESULTS: HEALTH PERCEPTIONS, SMOKING

Explanatory Variable and Other Measures	Dependent Variable Equation			
	Health Perceptions		Smoking	
	Coeff.	t-Test	Coeff.	t-Test
Intercept	67.25	35.92	1.373	41.54
Catastrophic	0.670	1.06	-0.0089	-0.71
Intermediate	0.635	1.14	0.0020	0.16
Individual Deductible	0.495	0.84	-0.0005	-0.04
Personal Functioning	0.0497	4.55	0.0002	0.80
Role Functioning	Not Used		Not Used	
Mental Health	0.0388	2.06	-0.0005	-1.30
Social Contacts	Not Used		Not Used	
Health Perceptions	0.5679	15.87	-0.0007	-2.16
Smoking	Not Used		0.7009	21.54
Age	-0.1618	-9.73	-0.0014	-4.16
Male	-0.286	-0.74	0.0100	1.26
Income	1.637	2.52	0.0090	0.72
Three-Year Term	-0.235	-0.50	0.0330	3.46
Took Physical	0.243	0.55	-0.0237	-2.54
Dayton	-3.344	-4.92	0.0476	3.50
Fitchburg	-1.824	-2.49	0.0343	2.23
Franklin County	-0.053	-0.08	0.0250	1.51
Charleston	-1.461	-1.98	0.0230	1.55
Georgetown County	-3.010	-4.30	0.0073	0.54
Income*Catast.	-1.038	-1.03	-0.0044	-0.22
Income*Interm.	-0.061	-0.62	0.0005	0.03
Income*ID	-0.688	-0.65	-0.0349	-1.88
Health*Catast.	0.0413	0.92	0.0215	0.56
Health*Interm.	-0.0109	-0.27	0.0356	1.06
Health*ID	0.0068	0.17	0.0440	1.20
Health*Income	0.1463	3.26	0.0238	0.63
Health*Term3	0.0029	0.08	0.0778	2.63
Health*Income*Catast.	-0.0854	-1.13	-0.0132	-0.20
Health*Income*Interm.	-0.1022	-1.29	0.0231	0.41
Health*Income*ID	-0.0090	-0.13	-0.0052	-0.08
Sample size	3843	—	3758	—
R-squared	0.4113	—	0.6093	—
Residual standard error	11.964	—	0.24429	—

Appendix F

DETAILED RESULTS OF REGRESSION EQUATIONS FOR CHOLESTEROL, WEIGHT, VISION, BLOOD PRESSURE, AND RISK OF DYING

INTRODUCTION

This appendix contains the regression equations for the remaining five measures reported in the main text. For definitions for most of the explanatory variables and their abbreviations, see Appendix E. As in Appendix E, the * indicates multiplication. The variables used only in these equations are defined below.

"Predicted Health": For persons who did *not* take the screening examination: the linear combination of enrollment values for site, age, sex, and response to relevant MHQ questions that best predicted exit values of the variable in question.

For persons who did take the examination: the linear combination of the enrollment values for those same variables and for the dependent health measure in question.

Elevated: Elevated risk. The (arbitrarily defined) least healthy 25 percent of the distribution of predicted exit values.

Low: The other 75 percent of the distribution of predicted exit values.

These three variables are described more fully in the next section.¹

DEFINING "PREDICTED HEALTH" AND "ELEVATED-RISK GROUP"

Our sample was representative of the nondisabled and noninstitutionalized general population under 65 years of age. This relatively young and healthy population does not have much chronic disease. To

¹Appendix G presents further analyses on the elevated-risk group designed to clarify whether using different group sizes would have influenced our results.

carry out our analyses, therefore, we had to use information as efficiently as possible.

Because chronic conditions are fairly stable, it was helpful to use as covariates the enrollment physiologic measurements on the 60 percent of the sample for whom we had such data. We could not afford, however, to ignore the other 40 percent of our sample. Fortunately, we had other data such as age, sex, and answers to MHQ questions about the condition that gave some indication of what their enrollment status might have been. Indeed, even physical measurements such as blood pressure on the day of the screening examination are not perfect measures of the participants' "true" status because of day-to-day fluctuations around the true value.

To highlight the effects of insurance plan, we used whatever enrollment data were available to "predict" what exit values would be on average, as a first step in identifying the elevated-risk group. This section presents the equations used to predict exit values for persons with and without initial measurements. We did such predictions for all physiologic measures except weight. Weight is extremely stable and accurately measured, so the initial value was used to define the elevated-risk group. Since we used the MHQ weight measure, which is highly reliable, we did not need to worry about those who missed the initial physical.

Tables F.1 through F.4 present the coefficients and t-test values for the variables used to predict exit values for individuals with and without enrollment measures from the screening examination. Also given are the number of persons analyzed, the equation's R-squared value, and the residual standard error (S.E.).

As with other regression analyses, some explanatory variables are dummy variables. The "omitted" categories include: for site, Dayton; for sex, female; for enrollment term, five years; for child (<17) versus adult, adult; and for various MHQ items that are identified by a "?," a "no" response (or a response indicating normal health). The regression model for vision (see Table F.1) includes a variable called "natural vision," which refers to vision at exit *without* glasses measured in terms of the denominator of the 20/40 ratio (i.e., someone with 20/40 vision is assigned the value 40). All equations except that for vision include interaction terms combining site with three-year term of enrollment (shown, for instance, as Dayton3).

Table F.1
REGRESSION EQUATIONS TO PREDICT EXIT VALUES FOR
DETERMINING ELEVATED RISK:
FUNCTIONAL FAR VISION

Variable	With Initial Values		Without Initial Values	
	Coeff.	t-Test	Coeff.	t-Test
Intercept	1.29	2.4	2.52	4.9
Male	0.003	0.04	-0.01	-0.2
Age	0.009	2.5	0.01	4.0
Child	0.14	0.9	0.22	1.9
Term3	-0.11	-1.5	-0.11	-1.7
Wear glasses?	0.27	5.0	0.54	2.0
Read newsprint?	0.05	0.6	0.13	1.8
Spot a friend?	-0.19	-2.0	-0.14	-1.8
Restrict activity?	-0.14	-1.9	-0.32	-5.0
Self-rated vision	0.26	2.7	0.19	2.2
Natural vision	0.003	1.8	0.003	2.6
Seattle	0.23	1.9	0.28	3.1
Fitchburg	0.56	4.7	0.66	6.7
Franklin County	0.20	1.7	0.24	2.5
Charleston	0.75	5.6	0.77	7.5
Georgetown County	0.33	2.5	0.49	4.7
Initial vision	0.36	8.8	—	—
Sample size	1143	—	1841	—
R-squared	0.20	—	0.11	—
Residual S.E.	1.19	—	1.22	—

Dividing the Sample into Those at Elevated Risk and Those Not

Because medical care can be expected to benefit most persons who have or will develop health problems, we used the predictions to divide our sample into groups likely to be normal or abnormal at exit. Choosing the size of the elevated-risk group was an arbitrary decision. Defining the group too tightly risked excluding those whose condition got unexpectedly worse during the study or those who had a good day

on the day of the enrollment physical. Defining the group too loosely risked diluting the real effects by including too many normal people whom medical care would not be expected to benefit.

For our analyses, we judged that studying the least healthy 25 percent would be a reasonable definition of elevated risk; we picked 25

Table F.2
REGRESSION EQUATIONS TO PREDICT EXIT VALUES
FOR ELEVATED RISK: CHOLESTEROL

Variable	With Initial Values		Without Initial Values	
	Coeff.	t-Test	Coeff.	t-Test
Intercept	64.53	6.3	202.29	18.6
Male	-2.11	-0.2	-61.72	-6.6
Age	0.53	1.4	0.64	1.6
Age-squared	0.00	0.0	0.01	2.3
Age * male	0.72	1.3	4.26	7.6
Age-squared * male	-0.01	-2.1	-0.06	-8.1
Told high chol?	-12.02	-3.2	-22.15	-5.1
Seattle	17.04	4.9	14.11	4.2
Fitchburg	8.63	2.5	4.00	1.0
Franklin County	10.37	3.2	9.44	2.5
Charleston	14.43	4.1	18.36	4.5
Georgetown County	17.40	5.3	18.37	4.9
Dayton3	34.10	12.2	32.41	11.5
Seattle3	-14.11	-4.2	-12.49	-3.9
Fitchburg3	-9.04	-2.7	-11.55	-2.8
Franklin3	-10.63	-3.5	-14.98	-4.1
Charleston3	-3.09	-0.9	-12.73	-3.1
Georgetown3	-3.94	-1.2	-12.25	-3.4
Standardized weight	-0.06	-0.1	2.62	3.8
Initial cholesterol	0.65	39.6	—	—
Sample size	1984	—	3381	—
R-squared	0.61	—	0.29	—
Residual S.E.	2.6	—	3.7	—

percent as the criterion before seeing any of the results except blood pressure. This avoided misstatements of the significance of our results, which would follow if we had chosen cutoffs that produced the biggest plan differences and treated the statistical tests so obtained as if they came from the only regression model tried. Nonetheless, our uniform and arbitrary choice of "elevated risk" may have missed some important results by focusing on the wrong size group. The point is taken up again in Appendix G.

Combining Results from Persons Who Did and Did Not Receive the Enrollment Screening Examination

The predicted values were also used to combine the findings from the screened and unscreened group. Using the Dagenais (1971) procedure, we computed the appropriate weight for the unscreened group, whose initial value contains somewhat less information. In other words, the 40 percent without initial values are included but are appropriately down-weighted in the final regressions.

The Dagenais procedure to determine the weights and the subsequent multilinear regressions are iterative; thus, combining all the information related to site, age, sex, and initial MHQ and screening values greatly reduces the computing expense. Because these factors are orthogonal (i.e., unrelated) to insurance plan, our estimates of the effects of plan on health status are unaffected by this simplification.

RESULTS OF REGRESSION ANALYSES FOR FINAL RESULTS

Table F.5 gives the coefficients, t-test values, and other information for the equations for cholesterol and weight that produced the results shown in Tables 5 and 7 of the main text. Table F.6 gives the equivalent information for functional far vision, diastolic blood pressure, and the risk of dying index, in this case for the "average" individual (i.e., for the entire sample).

Table F.7 contains similar information for the three equations that produced the results reported in Table 8 of the main text. These equations pertain just to the elevated-risk subgroups as defined above.

Table F.3
REGRESSION EQUATIONS TO PREDICT EXIT VALUES FOR
ELEVATED RISK: DIASTOLIC BLOOD PRESSURE

Variable ^a	With Initial Values		Without Initial Values	
	Coeff.	t-Test	Coeff.	t-Test
Intercept	32.34	6.6	74.35	19.6
Male	2.00	0.6	0.31	0.1
Age	0.44	3.4	0.62	5.9
Age-squared	-0.00	-1.8	-0.00	-2.7
Age * male	0.15	0.8	0.35	2.3
Age-squared * male	-0.00	-1.3	-0.01	-2.7
Seattle	-5.21	-1.4	0.35	0.4
Fitchburg	-2.02	-0.6	1.89	1.7
Franklin County	-4.55	-1.3	0.44	0.4
Charleston	-4.39	-1.2	2.05	1.8
Georgetown County	-2.17	-0.6	4.19	4.1
Dayton3	2.59	2.7	2.44	3.2
Seattle3	0.84	0.7	0.81	0.9
Fitchburg3	-0.69	-0.6	-0.17	-0.2
Franklin3	0.16	0.2	-0.08	-0.1
Charleston3	0.83	0.7	0.31	0.3
Georgetown3	-0.30	-0.3	-1.18	-1.2
Ever told HBP?	2.38	2.0	4.38	4.4
More than once?	4.69	5.1	6.87	7.8
On drugs for HBP?	4.44	4.4	5.21	4.7
BP now high?	-0.64	-0.5	-4.10	-3.3
Advised "no salt"?	—	—	-1.64	-1.7
Worry about BP?	—	—	-1.30	-2.0
Dayton diastolic	-0.03	-0.7	—	—
Initial diastolic	0.32	11.5	—	—
Initial systolic	0.11	6.5	—	—
Sample size	2042	—	3472	—
R-squared	0.39	—	0.28	—
Residual S.E.	9.2	—	10.2	—

^aBP = blood pressure; HBP = high blood pressure.

Table F.4

REGRESSION EQUATIONS TO PREDICT EXIT VALUES
FOR ELEVATED RISK: RISK OF DYING INDEX

Variable ^a	With Initial Values		Without Initial Values	
	Coeff.	t-Test	Coeff.	t-Test
Intercept	1.691	8.2	3.597	23.8
Male	0.229	1.4	0.049	0.4
Age	-0.013	-1.9	-0.012	-2.3
Age-squared	0.000	3.4	0.000	5.6
Age * male	0.001	0.1	0.018	2.4
Age-squared * male	-0.000	-0.8	-0.000	-3.3
Seattle	-0.002	-0.0	0.024	0.5
Fitchburg	0.097	1.6	0.154	2.8
Franklin County	0.031	0.6	0.149	3.0
Charleston	0.042	0.7	0.091	1.7
Georgetown County	0.085	1.5	0.137	2.7
Dayton3	0.050	1.1	-0.016	-0.4
Seattle3	-0.098	-1.7	-0.036	-0.8
Fitchburg3	-0.066	-1.1	-0.071	-1.3
Franklin3	-0.090	-1.7	-0.123	-2.5
Charleston3	0.116	1.9	0.098	1.8
Georgetown3	-0.003	-0.0	-0.031	-0.6
Told HBP?	0.119	1.9	0.193	3.9
More than once?	0.105	2.3	0.304	7.7
On drugs for HBP?	0.179	3.5	0.284	6.2
BP now high?	0.001	0.0	-0.153	-2.6
Told high chol?	-0.066	-1.0	-0.077	-1.3
Smoking index	0.358	6.5	1.041	46.6
TOOKPHYS	—	—	-0.067	-3.7
Initial diastolic	0.002	1.9	—	—
Initial risk index	0.504	13.9	—	—
Sample size	1943	—	3217	—
R-squared	0.58	—	0.51	—
Residual S.E.	0.45	—	0.49	—

^aBP = blood pressure; HBP = high blood pressure; TOOK-
PHYS = took enrollment screening examination.

Table F.5
REGRESSION COEFFICIENTS, t-TEST VALUES, AND OTHER
EQUATION RESULTS: CHOLESTEROL AND WEIGHT

Explanatory Variables and Other Measures	Dependent Variable Equation			
	Cholesterol		Weight ^a	
	Coeff.	t-Test	Coeff.	t-Test
Intercept	6.25	0.9	6.78	3.9
Male	Not Used		0.69	1.9
Predicted Health	0.98	34.7	0.93	41.6
Low	-2.46	-0.9	-0.69	-1.1
Catastrophic	-0.001	-0.0	0.16	0.5
Intermediate	-2.75	-1.8	-0.088	-0.3
Individual Deductible	1.31	1.0	0.31	1.0
Elevated*Catast.	-4.41	-1.2	-0.74	-0.8
Elevated*Interm.	1.81	0.5	-0.46	-0.5
Elevated*ID	- 85	-0.2	-0.075	-0.1
Sample size	3381	—	2804	—
R-squared	0.50	—	0.83	—
Residual S.E.	28	—	6.1	—

^aFor weight, "predicted health" is the initial weight reported on the MHQ.

Table F.6

REGRESSION COEFFICIENTS, t-TEST VALUES, AND OTHER EQUATION
RESULTS FOR THE ENTIRE SAMPLE: FUNCTIONAL FAR VISION,
DIASTOLIC BLOOD PRESSURE, AND RISK OF DYING

Explanatory Variables and Other Measures	Dependent Variable Equation					
	Functional Far Vision ^a		Diastolic Blood Pressure		Risk of Dying	
	Coeff.	t-Test	Coeff.	t-Test	Coeff.	t-Test
Intercept	-0.19	-1.0	4.87	1.5	-0.36	-2.3
Predicted Health	1.01	15.0	0.95	26.7	1.05	37.7
Low	Not Used		-1.64	-2.0	0.12	2.6
Catastrophic	0.25	3.2	1.28	2.4	-0.006	-0.2
Intermediate	0.15	2.0	0.61	1.3	-0.033	-1.5
Individual Deductible	0.18	2.3	0.60	1.2	0.004	0.2
Elevated*Catast.	Not Used		-1.07	-0.9	0.13	2.0
Elevated*Interm.	Not Used		-0.11	-0.1	0.10	1.7
Elevated*ID	Not Used		0.92	0.8	0.12	2.0
Sample size	1841	—	3495	—	3317	—
R-squared	0.18	—	0.35	—	0.56	—
Residual S.E.	1.2	—	9.3	—	0.45	—

^aSample for far vision includes only persons whose natural far vision was worse than 20/20 in the better eye.

Table F.7

REGRESSION COEFFICIENTS, T-TEST VALUES, AND OTHER EQUATION RESULTS WITH INCOME AND ELEVATED-RISK INTERACTIONS: FUNCTIONAL FAR VISION, DIASTOLIC BLOOD PRESSURE, AND RISK OF DYING

Explanatory Variables and Other Measures	Dependent Variable Equation					
	Functional Far Vision*		Diastolic Blood Pressure		Risk of Dying	
	Coeff.	t-Test	Coeff.	t-Test	Coeff.	t-Test
Intercept	-0.21	-0.6	4.33	1.4	-0.36	-2.3
Predicted Health	1.01	8.7	0.96	26.7	1.05	37.6
Low	0.022	0.2	-1.70	-2.1	0.11	2.4
Income	-0.16	-2.4	-0.57	-1.1	-0.015	-0.6
Catastrophic	0.27	3.5	1.26	2.24	-0.007	-0.2
Intermediate	0.09	1.4	0.65	1.31	-0.033	-1.4
Individual Deductible	0.29	3.5	0.71	1.33	0.005	0.2
Elevated*Income	0.053	0.4	1.20	1.0	0.063	1.2
Elevated*Catast.	-0.066	-0.4	-1.07	-0.9	0.12	1.8
Elevated*Interm.	0.11	0.8	-0.03	-0.0	0.091	1.5
Elevated*ID	-0.21	-1.3	0.81	0.7	0.11	1.9
Income*Catast.	0.17	1.3	-0.49	-0.6	-0.010	-0.3
Income*Interm.	0.24	2.2	-0.01	-0.0	-0.003	-0.1
Income*ID	-0.38	-2.4	0.58	0.7	0.002	0.1
Elevated*Income						
*Catast.	-0.38	-1.4	-0.12	-0.1	-0.035	-0.4
Elevated*Income						
*Interm.	-0.53	-2.2	-2.26	-1.4	-0.076	-0.9
Elevated*Income						
*ID	-0.41	-1.7	-2.50	-1.5	-0.053	-0.6
Sample Size	1841	—	3495	—	3317	—
R-squared	0.18	—	0.36	—	0.56	—
Residual S.E.	1.2	—	9.3	—	0.45	—

*Sample for vision includes only persons whose natural far vision was worse than 20/20 in the better eye.

Appendix G

SENSITIVITY OF RESULTS TO THE SIZE OF THE ELEVATED-RISK GROUP

INTRODUCTION

Our main experimental analyses focused on persons who were abnormal at enrollment or had a good chance at becoming abnormal during the study. The size of these groups may vary from condition to condition. As noted elsewhere, we reported on an “a priori,” or arbitrarily defined, group at elevated risk specified as the least healthy 25 percent, but this is not necessarily the “optimal” group for such investigations. This appendix describes additional work we did to elucidate how the size of the elevated-risk group might have influenced our results.

LEGALIZED SEARCH FOR THE BEST ELEVATED-RISK GROUP

The “Maximum *t*” Procedure

We analyzed the problem of picking the subgroups at elevated risk with a method that has two advantages relative to the approach described above: It does not rely on an arbitrary cutoff (e.g., the least healthy 25 percent), and it allows the size of the elevated-risk group to vary from disease to disease. The procedure¹ was as follows: We explicitly searched for the least healthy subgroup that yielded the *largest* value for the *t*-test of the differences between the free and cost-sharing plans—hence the term “maximum *t*” procedure. We then tested significance against the distribution that would result from this search if insurance plan had had no effect, instead of against the usual *t* distribution. In effect, we “dredge” the data to find the biggest effects, but the higher standard of statistical significance applied to these effects legitimizes the procedure.

¹We are indebted to Professor Carl Morris of the University of Texas, Austin, for suggesting this method.

Specifically, we computed the significance of the difference between the free and cost-sharing plans for groups defined by the least healthy 40, 35, 30, 25, 20, 15, 10, and 5 percentiles of predicted exit values. We then selected whichever group has the maximum (highest) t value as our preferred elevated-risk group. Under the null hypothesis that insurance plan has no effect on the health condition, this maximum t has a distribution that we have empirically computed. Thus, we can see how likely the observed results would be if insurance plans really had had no effect on health.

Empirical Presentation of the Maximum t

Table G.1 shows the empirical distribution of the maximum t . For contrast, the table also presents the distribution of the usual t ; these are the appropriate values if the size of the elevated-risk group had been selected before the results were examined. Because our sample size is large, this usual distribution is simply standard normal.

The top row of Table G.1 shows the values for maximum t (a kind of one-tailed) test at the specified percentile from 50 to 99.5. The middle row gives the absolute values of the maximum t , which correspond essentially to a two-tailed test because effects in either direction might have the larger absolute value. Finally, the bottom row presents the familiar values of an ordinary t -test.

The empirical distribution was derived as follows. Under the null hypothesis, the difference between the free and cost-sharing plans for

Table G.1

VALUES OF MAXIMUM t , MAXIMUM ABSOLUTE t , AND USUAL t
AT SELECTED POINTS IN THE DISTRIBUTION

Type of t -Test	Percentiles of the Distribution				
	50	95	97.5	99	99.5
Maximum t^a	0.74	2.20	2.48	2.84	3.11
Maximum $ t $	1.42	2.45	2.78	3.10	3.31
Usual t	0	1.65	1.96	2.33	2.58

^aSee text. This is the empirical distribution of M_8 .

each 5-percent increment is independent, identically distributed, and approximately normal. Thus, let Z_1, Z_2, \dots, Z_k be independent $N(0,1)$. Let S_i be the sum of the first Z_i , and let $T_i = S_i / (i)^{0.5}$. Finally, let M_n be the maximum of the first n T_i . Then, M_g has the desired distribution. This is easy to simulate, and Table G.1 comes from a set of 2000 randomly generated M_g s.

PREDICTED EXIT VALUES FOR THE OPTIMAL ELEVATED-RISK GROUP

Table G.2 shows the percentile cutoff that would provide the optimal sample size for blood pressure, vision, cholesterol, risk of dying, and weight. Also given are the differences between the free and cost-sharing plans when the elevated-risk group is so defined, and the t-test and p values corresponding to that difference. This is a one-tailed

Table G.2

OPTIMAL SAMPLE PERCENTILES AND OTHER EVALUATIVE MEASURES, BY HEALTH CONDITION

Health Conditions	Percentile Defining the Optimal Size of Elevated-Risk Group	Predicted Exit Values and Significance Levels for Optimal Elevated-Risk Group		
		Free Minus Cost-Sharing	Maximum t-Test	P Value
Diastolic Blood Pressure	40	-1.2 mm Hg	-2.08	0.07
Functional Far Vision	100 ^a	-0.2 lines	-3.29	<0.005
Cholesterol	5	-5.1 mg/dl	-0.68	>0.5
Risk of Dying	5	-0.32 logits	-2.97	0.01
Weight	5	-1.8 kg	-0.87	0.5

^aThis 100-percent figure is with reference to *only* those adults whose vision was worse than 20/20—i.e., the 53 percent of the entire sample who were eligible to be in the analysis.

analysis with the free plan hypothesized to be better than cost-sharing plans.

These results can be compared with those in Table 7 of the main text, where the elevated-risk group was arbitrarily defined a priori. The optimal elevated-risk group for blood pressure appeared to be rather larger (top 40 percent of the distribution, rather than the top 25 percent), but this may occur because a single blood pressure measurement cannot precisely define those who would benefit from treatment. Indeed, though this "optimal" group showed stronger effects than the "a priori" group, the strongest effects (statistically speaking) were observed for the "average" enrollee (see Table 5 of main text).

As it happens, the optimal group for vision was the same as that selected for the main analyses: 100 percent of all adults with natural vision worse than 20/20, who accounted for 53 percent of all adults in the study, fell into the analytic sample. (For vision, the percentiles were tried in 10 steps of 10 percent each.) The maximum t 's for cholesterol and weight occurred for the smallest elevated-risk groups analyzed (defined at the 5th percentile). For both cholesterol and weight analyzed at this cutoff, the free plan was better, but not significantly so, than the cost-sharing plans.

Finally, the risk of dying index also had its biggest t values when the elevated-risk group was defined by the 5th percentile cutoff. This result was quite significant in favor of the free plan, more so than in the main analyses. For short time periods—say, in the month following the end of the study—the difference of -0.32 logits shown in Table G.2 implies that mortality on the free plan would be approximately 73 percent of that on the cost-sharing plans ($\exp(-0.32) = 73$ percent) for this small group of extremely high-risk individuals. These results suggest that the medical treatment of health habits related to this risk index (cholesterol, smoking) or of blood pressure that is obtained under free-care circumstances is most valuable for people with extreme values.

Appendix H

RESULTS OF SPECIAL ANALYSES ON THE EFFECT OF SITE

INTRODUCTION

To determine if site would make a difference in the reported health status results or their interpretation, we undertook a special set of analyses for three different variables—one of the general health measures and two of the physiologic measures. The basic approach was to compute predicted exit values for each site using the same values for the explanatory (independent) variables as in the all-sites analyses.

HEALTH PERCEPTIONS

For the general health measures, we used health perceptions (i.e., the General Health Ratings Index) as the illustrative dependent variable. Findings were as expected: The sites did not differ in any statistically significant way (Table H.1). The F-test for all site differences was 1.5 (with 110 and 3705 degrees of freedom), and the F-test for plan by site effects alone was 0.6 (with 15 and 3811 d.f.). Neither of these values was statistically significant at $p < 0.05$.

We observed an insignificant difference between the plan effects in the two South Carolina sites and the remaining four sites; the free plan had higher (better) exit scores than the cost-sharing plans in those two sites, whereas the reverse was true for all other sites. We hypothesized that this might be explained by the large proportions of persons of low income or poor health in those sites. This proved not to be the case, however, because the predicted exit values for the low-income enrollees in ill health in the South Carolina sites were about the same on the free and the cost-sharing plans.

Table H.1
PREDICTED EXIT VALUES FOR HEALTH PERCEPTIONS,*
BY SITE AND PLAN

Site	Free Plan	Cost-Sharing Plans	Free Minus Cost-Sharing Plans	t-Test Value
Dayton	65.2	66.2	-1.0	-0.8
Seattle	68.5	70.2	-1.8	-1.9
Fitchburg	67.3	68.0	-0.6	-0.5
Franklin County	68.6	69.3	-0.7	-0.7
Charleston	68.9	67.6	1.3	1.0
Georgetown County	66.9	66.4	0.5	3.5
All sites	67.4	68.0	-0.6	-1.3

*The General Health Ratings Index is positively scored from 0 to 100.

DIASTOLIC BLOOD PRESSURE AND FUNCTIONAL FAR VISION

For both physiologic measures, we observed large differences among the sites in measured initial values and in differences between enrollment and exit scores. Screening examination procedures did not change *within* a site or "time of screening" (i.e., enrollment, three-year exit, or five-year exit), but personnel, equipment, and test procedures for measuring physiologic health did vary among sites and times. The various methods used to assess each physiologic variable were described in Appendixes B and D and by Smith et al. (1978).

Laboratory tests were performed in each site (so that local physicians would be familiar with the labs); hence, they could not be completely standardized across the experiment. Some differences between entry and exit can be attributed to aging of the participants, but most differences undoubtedly arose from measurement techniques.

Because plan assignments were not correlated with enrollment period or site, measurement differences by site will not bias the simple (i.e., uncontrolled) plan-difference results. To eliminate all possible effects of such measurement differences, however, we controlled for site and period of enrollment (as shown, for instance, in Appendix F). This

adjustment was then carried over to the final equations in Appendix F through the "predicted health" variable. Generally, we used 11 dummy variables for all but one of the site-enrollment term possibilities in our regressions (six sites times two enrollment terms minus an omitted category).

Because measurement techniques were not standardized, we were precluded from making any meaningful interpretation of differences between sites at enrollment or of trends from the beginning to the end of the experiment. Such differences would be better determined from large standardized national surveys, such as the periodic Health and Nutrition Examination Surveys carried out by the National Center for Health Statistics.

Tables H.2 and H.3 show the exit values for blood pressure and vision for each of the six sites. The differences in plan effects among the sites were small, as expected.

For a formal conservative test, we ran a regression of final blood pressure and corrected vision with the five site-free-plan interactions. In each case, if the sites are not different, then a statistic derived from the R-squared values of this and the original regression has an F distribution. With the large number of degrees of freedom (5, 3000), this statistic is essentially a chi-square with 5 degrees of freedom. The

Table H.2
PREDICTED EXIT VALUES FOR DIASTOLIC BLOOD PRESSURE,
BY SITE AND PLAN

Site	No.	Cost-Sharing Plans				Free Plan	Free Minus Cost-Sharing
		Catastrophic	Inter-med.	Ind. Deduct.	Total		
Dayton	709	78.7	78.4	78.3	78.5	78.7	0.2(-1.7,2.1) ^a
Seattle	742	77.9	78.9	78.5	78.4	78.1	-0.3(-1.8,1.2)
Fitchburg	451	79.7	78.5	78.5	78.9	77.1	-1.8(-3.7,0.1)
Franklin	550	80.7	78.2	78.9	79.3	77.8	-1.5(-3.1,0.1)
Charleston	442	79.3	79.0	78.9	79.1	77.9	-1.2(-3.0,0.7)
Georgetown	601	77.9	78.1	78.9	78.3	77.9	-0.4(-2.4,1.6)

^a95-percent confidence intervals in parentheses.

chi-square value for blood pressure was 3.2 ($p = 0.7$), and the value for vision was 10.6 ($p = 0.06$). The latter was close to significant, owing mainly to one person in Charleston whose eyesight went from 20/200 to 20/25.

Table H.3
PREDICTED EXIT VALUES FOR CORRECTED VISION,
BY SITE AND PLAN

Site	No. ^a	Cost-Sharing Plans					Free Minus Cost-Sharing ^b
		Cata- strophic	Inter- med.	Ind. Deduct.	Total	Free Plan	
Dayton	411	2.88	2.83	2.74	2.81	2.65	-0.16(-0.36,0.04)
Seattle	400	3.14	2.71	3.13	2.99	2.72	-0.27(-0.47,-0.07)
Fitchburg	260	3.10	3.00	3.02	3.04	2.47	-0.57(-0.84,-0.30)
Franklin	302	2.90	3.29	2.91	3.03	2.81	-0.23(-0.46,0.005)
Charleston	229	2.85	2.70	2.79	2.97	2.87	-0.10(-0.57,-0.37)
Georgetown	239	3.33	2.89	2.86	3.03	3.01	-0.01(-0.39,0.37)

^aSample in table includes just those with less than perfect natural vision in the better eye at exit. The numbers in text included the 47 percent of the population with normal natural vision (a score of 2.00).

^b95-percent confidence intervals in parentheses.

Appendix I

SAMPLE LOSS FOR PHYSIOLOGIC MEASURES, CHOLESTEROL, AND WEIGHT

Details about the reasons for sample loss are presented in this appendix. Special attention is given to how we arrived at the final samples for functional far vision, diastolic blood pressure, cholesterol, weight, and risk of dying.¹ We gave special attention to these measures because the self-reported general health measures had considerably fewer missing values owing to the high response rates for the attrition MHQ.

As noted in the main text, initially we enrolled 3958 adults in the various fee-for-service plans. Of these, 14 had neither enrollment nor exit data on these five measures, and 393 had enrollment but no exit data. Table I.1 distributes these persons across plans. By and large, persons who left the experiment voluntarily (i.e., attrited) are concentrated in the cost-sharing plans. Individuals who left the study prematurely for other reasons, including health-related reasons, are spread fairly evenly across all plans.

In Table I.2, the reasons for sample loss are given for the five health measures noted above, *over and above* the total sample loss of 407 (14 + 393 = 407) shown in Table I.1. Table I.2 losses are shown sequentially. For example, persons with no exit data from the MHQ who were also under 18 years of age are counted in the row labeled "no exit data," but not in the row labeled "under 18 at enrollment."

In Table I.2, "no exit data" means that the person had no data at exit on the particular health measure under consideration, although he or she may well have had exit data on any or all of the other health measures. "No exit data" for the risk of dying index could have arisen from missing data on any of its components.

Not having enrollment data was an important reason for sample loss only for weight. Other restrictions to the weight definition were also important, however. For example, we excluded teenagers 14-17

¹"Final sample" here included persons who may not have received the enrollment screening examination. Thus, the figures shown for "final sample" may differ from sample size numbers given in the main text or in other appendixes because those reflect "analytic samples"; the latter may exclude persons with "planned" or "unplanned" missing data (depending on the analysis under consideration).

Table I.1
REASONS FOR SAMPLE LOSS FOR PHYSIOLOGIC MEASURES,
CHOLESTEROL, AND WEIGHT, BY PLAN

Reason for Sample Loss of Eligible Adults	Total Number	Number by Plan			
		Free Plan	Inter- med.	Cata- strophic	Ind. Deduct.
No. without enrollment and exit data	14	2	5	2	5
No. with entry data only	393	70	94	119	110
Left voluntarily	174	4	39	82	49
Terminated for nonhealth reasons	113	32	26	22	33
Terminated for health reasons	37	13	12	2	10
Died	37	10	11	6	10
Completed study, but had no exit data	32 ^a	11	6	7	8
Total with some exit data	3551	1222	925	638	766
No. eligible adults in sample	3958	1294	1024	759	881

^aThree people on the individual deductible plan had exit data on weight only and are included in this row; thus, they are *not* included in the "total with some exit data" row.

because we believed the growth during those years would make any weight change less meaningful for them than for older participants.

Of the 347 persons with no enrollment measure, 117 were persons in the three-year South Carolina samples who had not initially been included in these analyses. The weight analyses were rerun to include them, bringing the sample to 2921. Changes in results, relative to those shown in Tables 5 and 7, were negligible.

Table I.2
 REASONS FOR SAMPLE LOSS AND FINAL SAMPLE SIZE,
 BY INDIVIDUAL HEALTH MEASURE AND PLAN

Reason for Sample Loss of Eligible Adults	Number by Plan				
	Total Number	Free Plan	Inter- med.	Cata- strophic	Ind. Deduct.
Total with some exit data	3551	1222	925	638	766
Cholesterol					
No exit data	170	67	40	24	39
Final sample	3381	1155	885	614	727
Weight					
No exit data	5	2	0	0	3
No enrollment measure	347	127	79	57	84
Under 18 at enrollment or pregnant at exit	398	150	105	80	63
Final sample ^a	2804	943	741	501	619
Vision					
No exit data	74	40	13	13	8
Final sample	3477	1182	912	625	758
Blood Pressure					
No exit data	56	31	8	6	11
Final sample	3495	1191	917	632	755
Risk of Dying					
No exit data	234	85	55	39	55
Final sample	3317	1137	870	599	711

^aThe final sample for weight is shown as 2804 rather than 2801 (3551 less the sum of 5, 347, and 398) because three people had exit data only for weight. They had *not* been included in the row entitled "total with some exit data" on Table I.1.

Appendix J

RESULTS OF SPECIAL ANALYSES ON EFFECTS OF MISSING DATA: PHYSIOLOGIC MEASURES AND CHOLESTEROL

INTRODUCTION

In defining the samples for the analyses (reported in the main text) of the physiologic measures and cholesterol, we had made two decisions relating to persons with missing data. The first was to *include* people who did not take the enrollment screening examination (a randomly selected 40 percent of enrollees). The second was to *exclude* people who did not have data at exit. This appendix presents results of some tests made to determine how sensitive our main results were to these decisions.¹

The basic approach for this analysis is to calculate results (i.e., calculate predicted exit values) for persons with complete enrollment and exit data (as done in Table J.1), to compute results with various persons included or omitted from the analysis, and then to compare these sets of results with those given in the main body of the report. The methods are the same as those used to produce the information in Table 5.

INCLUDING PEOPLE WHO LACKED ENROLLMENT SCREENING DATA

The data in Table J.1 indicate that *including* individuals who lacked data from the screening examination at enrollment did not influence our results materially, except to make the estimates more precise because of the larger sample size. Comparing the differences shown in the far right column of Table J.1 with those in the equivalent column in Table 5 shows almost no discrepancy for any of these health measures in the differences between free and cost-sharing plans. For example, Table 5, which included people with no enrollment examination, gives a free -cost-sharing difference for the risk of dying of -0.02; in Table J.1, the difference is -0.01. We concluded from this analysis

¹Weight was not analyzed here because it was based on MHQ, not screening examination, data and thus already included persons who did not receive the examination.

Table J.1
PREDICTED EXIT VALUES FOR ONLY THOSE PERSONS WITH
COMPLETE ENROLLMENT AND EXIT DATA

Health Status Measures	No.	Cost-Sharing Plans				Free Plan	Free Minus Cost-Sharing Plans
		Catastrophic	Inter-med.	Ind. Deduct.	Total		
Cholesterol (mg/dl)	1984	202	201	203	202	202	0.0 (-3,3)*
Far vision (Snellen lines)	2158	2.58	2.54	2.54	2.55	2.42	-0.12 (-0.02,0.23)
Blood pressure (mm Hg)	2066	78.4	78.3	78.6	78.4	77.9	-0.5 (-1.4,0.4)
Risk of dying (%)	1943	0.99	0.98	1.04	1.00	1.00	-0.01 (-0.05,0.04)

*95-percent confidence intervals in parentheses.

that including the people without initial examination data was reasonable.

EXCLUDING PEOPLE WHO LACKED EXIT DATA

Theoretical Background

To test the effects of noncompletion for our final results, we imputed exit values to noncompleters according to the following formula:

$$Y_1 \mid \text{quit} = \hat{Y}_1 + \Delta Y(1 - \rho).$$

\hat{Y}_1 is the predicted value of Y_1 ; ρ is the coefficient of Y_0 in a regression of exit values on initial values and other covariates for the sample of persons with both measurements; and ΔY is the residual average value in a regression of initial values on covariates in the subset of noncompleters in the group of interest. (Quit is used interchangeably with noncompletion.)

This correction is prompted by the possibility that those who did not complete the experiment may have been systematically different from those who did complete even from the outset of the study. The *degree* of difference is estimated by comparing the initial (enrollment) values for noncompleters on each plan with equivalent values for completers. (We have eliminated noncompleters who did not take the entry physical examination, because we have too little information to make a useful exit prediction.) The theoretical justification for this approach assumes that noncompletion is partially a function of the physiologic health measure being studied, and it makes a correction for the additional piece of information that the person has in fact left the study.

This correction is not as conservative as assuming that all noncompleters have extreme values on the measure, but it is aesthetically more appealing and intuitively more correct. It avoids huge corrections when Y_0 is not correlated with Y_1 but is associated with leaving the study early. Moreover, it may hold under assumptions other than the ones given below to derive it.

The formula rests on three assumptions. They basically say that completion of the study depends on the true value of the underlying Y , which is stable.

1. Assume $Y_{it} = X_{it}B + u_i + e_{it}$ where u and e are normal and uncorrelated with X and with each other. Define $T_i = X_iB + u_i$ to be the true value. Let $s^2 = s_u^2 + s_e^2$ where s_u^2 is the variance of unobserved stable effects, and s_e^2 is the variance of "measurement" error. Let $\rho = s_u^2 / s^2$.

2. The measurement error in Y , defined by e , is not error in the collection instrument but rather is mainly variation over time around the true average value T . We assume that error is large enough to worry about but that measurement error in X is negligible.

3. $\text{Log}(\text{Prob}(\text{quit})) = aT + X\gamma = au + X\theta, \theta = (aB + \gamma)$. The coefficient a can vary between groups defined by plan and reason for quitting the study early. In the derivation, we assume that we are focusing on one particular group, so a is unique. In practice, we did not distinguish among noncompleting groups.

Under these assumptions, u given Y_0 and XB is normally distributed with mean $\rho(Y_0 - XB)$ and variance $(\rho - \rho^2)s^2$. Since XB does not provide information about u , the distribution of u given XB is normal with mean 0 and variance $s^2(\rho)$. Now since the $\text{prob}(\text{quit}) = +e^{au}g(X)$, we can use log-normal theory to show that $(u | Y_0, XB, \text{quit})$ is normally distributed with mean $\rho(Y_0 - XB) + a(\rho - \rho^2)s^2$. (See Aitchison and Brown (1957), theorem 2.6.) Thus, $E(u | Y_0, XB, \text{quit}) = \rho(Y_0 - XB) + a(\rho - \rho^2)s^2$.

Let $\hat{Y}_1 = XB + \hat{u} = XB + \rho(Y_0 - XB)$. If we know the subject quit, then

$$E(Y_1 | Y_0, XB, \text{quit}) = \hat{Y}_1 + a(\rho - \rho^2)s^2.$$

On the other hand, $E(\Delta Y_0 | \text{quit}) = E(u | XB, \text{quit}) = a(\rho)s^2$, where $\Delta Y_0 =$ the mean of $Y_0 - XB$. By substitution, we obtain the formula given above. The average residual ΔY_0 will be the coefficient of a dummy indicating noncompleters on a given plan in a regression of initial values on covariates in the full initial sample.

Empirical Findings

Table J.2 gives the differences in *enrollment* values between non-completers and the sample with complete entry and exit data for the three physiologic measures and cholesterol. In all cases, we controlled for age, sex, site, and income and computed differences as noncompleters minus completing sample. Only 200 noncompleters had enrollment values, and none of the differences was significant. On the free plan, noncompleters had lower cholesterol, better vision, and lower blood pressure than completers, but not a lower risk of dying. By contrast, on the catastrophic plan, noncompleters had better vision and blood pressure and a lower risk of dying than completers. Put another way, except on the intermediate plan, noncompleters had lower values

Table J.2

DIFFERENCES IN ENROLLMENT VALUES BETWEEN NONCOMPLETERS AND COMPLETING SAMPLE,* BY PLAN

Health Status Measures	Catastrophic	Intermediate	Ind. Deduct.	Free Plan
Cholesterol (mg/dl)	4.8	4.1	-0.9	-6.7
Far vision (Snellen lines)	-0.03	0.11	0.12	-0.08
Blood pressure (mm Hg)	-2.3	1.4	-1.3	-1.4
Risk of dying (%)	-0.05	0.03	0.11	0.03

*Completing sample is the one in Table J.1 for whom entry and exit data were complete. A positive value means that noncompleters had a higher value than completers.

of blood pressure at enrollment than completers; except on the catastrophic plan, noncompleters had a higher risk of dying.

The question of interest, of course, is what effect *including* these noncompleters might have had on our final exit results. To explore this point, we computed the predicted exit values for these four health status measures with the noncompleters, for whom an exit value had been imputed according to the formula given above. We then did essentially the same "free minus cost-sharing" calculation as reported in the main text, followed by a calculation of the "discrepancy" (i.e., difference between the differences) between the main-text results and these results.

The discrepancy is computed according to the following general formula:

$$\text{Discrepancy} = (\text{Free} - \text{Cost-sharing})_{CO} - (\text{Free} - \text{Cost-sharing})_{NC}$$

where *CO* means "with completers only" and *NC* means "with non-completers."

"With completers only" is equivalent to the sample used for Table 5. For these four health measures, in cases where the free plan had better outcomes than the cost-sharing plan, the free-minus-cost-sharing value will be negative in the above equation. Hence, in that case, a negative value for the *discrepancy* has the effect of reducing the advantage of the free over the cost-sharing plan.

Table J.3 gives the results of that analysis, shown as the actual arithmetic value of the discrepancy just described for each health measure. Generally, the effects were minor.

For example, the -0.14 value in the first column of Table J.3 means that including noncompleters would have narrowed the difference between the free and the catastrophic plan in diastolic blood pressure by 0.14 mm Hg; the difference in the risk of dying would have been narrowed by 0.4 percent. Including noncompleters would have widened the difference in vision between the free and individual deductible and free and intermediate plans by very small amounts. Thus, we concluded that the decision to omit from the analysis those persons who had not finished the experiment did not influence our results in any material way.

The main reason for not having included them to begin with was that we wished to keep our analyses as "clean" as possible. The various corrections needed to include noncompleters would have made our main analyses very complicated. Given that the exclusion did not affect our principal results, we believe that the decision was reasonable.

Table J.3

DISCREPANCIES IN THE DIFFERENCES BETWEEN FREE AND COST-SHARING PLANS IN PREDICTED EXIT VALUES WHEN NONCOMPLETERS ARE INCLUDED IN ANALYSIS AND WHEN THEY ARE NOT, BY HEALTH STATUS MEASURE

Health Status Measure	Discrepancy in Predicted Exit Values ^a			
	Free Minus Catastrophic	Free Minus Inter-med.	Free Minus Ind. Deduct.	Free Minus Total Cost-Sharing
Cholesterol (mg/dl)	0.3	0.2	0.1	0.2
Far vision (Snellen lines)	0.0	0.005	0.008	0.004
Blood pressure (mm Hg)	-0.14	0.07	-0.04	-0.003
Risk of dying (%)	-0.004	0.0	0.005	0.0

^aDiscrepancy between free-cost-sharing differences calculated as plan differences with completers only minus plan differences with noncompleters; see formula in text. A negative entry in this table means that including noncompleters would have increased the corresponding "free minus cost-sharing" value reported in Table 5.

Appendix K

RESULTS OF SPECIAL ANALYSES ON EFFECTS OF MISSING DATA: HEALTH PERCEPTIONS

INTRODUCTION

We undertook an exhaustive study of the effects of missing values for the general health measures, again using health perceptions (i.e., the General Health Ratings Index, or GHRI), as the dependent variable. Both enrollment and exit data were considered. Generally, our findings reflect what would be found for other self-assessment health variables.

SOURCES OF ENROLLMENT DATA

Altogether, 2290 of the total 3958 adults studied had "true" values on the health perceptions measure from their own enrollment MHQ. Another 816 in the South Carolina sites who were enrolled for three years (two years into the experimental period in that site) had "true" values from an MHQ administered when they were actually enrolled.

In Dayton, we did not obtain GHRI scores at enrollment, but we subsequently constructed an excellent predicted score from a number of other health measures (such as the Personal Functioning Index, Role Limitations, and the Mental Health Inventory). For the few enrollment values still missing, we either imputed a value based on age and the score on an earlier version of the GHRI or assigned the person the mean enrollment value for the entire adult sample. We judged the small methods bias induced by these imputation procedures to be less important than the bias that would have been produced by omitting or downweighting the observations.

As shown in Table K.1, the plans did not differ in the degree to which these sources of enrollment GHRI scores were used. Because of the manner in which all MHQ data were gathered and the instruments edited in the field, the health perceptions findings should parallel those for the other self-assessed general health measures. Thus, we are confident that our level of missing enrollment data and the outcomes of our procedures for overcoming what problems did arise did not differ meaningfully by plan.

Table K.1

DISTRIBUTION OF SOURCES OF ENROLLMENT VALUES FOR THE
GENERAL HEALTH RATINGS INDEX, BY SOURCE AND PLAN

Source of Enrollment Value	Free Plan		Cost-Sharing Plans		Total	
	No.	%	No.	%	No.	%
Actual enrollment MHQ	779	60	1511	57	2290	58
MHQ for South Carolina three-year sample	280	22	536	20	816	21
Predicted from other enrollment health variables ^a	204	16	551	21	755	19
Imputed from age and pre-enrollment value	26	2	62	2	88	2
Assigned mean for entire sample	5	(b)	4	(b)	9	(b)
Total	1294	100	2664	100	3958	100

^aAll for Dayton.^bLess than 0.5 percent.

SOURCES OF EXIT DATA

In general, for exit we accepted only "real" data for this (and other) dependent variables. If the exit MHQ was unavailable, we used a value from the attrition MHQ (see Appendix B). If that was unavailable, we used a value from the annual Health Questionnaire (completed during the experiment) that was closest to exit. In all other cases, the individual was dropped from the analysis.

Table K.2 shows the distribution by plan of these three sources of exit data. The overwhelming majority of people had exit values from the full exit MHQ. We excluded 115 individuals from the analysis because of missing exit data: Four had completed the study, 64 had withdrawn prematurely from the study, 30 had been terminated for reasons not related to health status, seven had been terminated for health reasons, and ten had died.

Table K.2

DISTRIBUTION OF SOURCES OF EXIT SCORES FOR THE GENERAL
HEALTH RATINGS INDEX, BY SOURCE AND PLAN

Source of Exit Score	Free Plan		Cost-Sharing Plans		Total	
	No.	% ^a	No.	% ^a	No.	%
Actual exit MHQ	1216	95	2326	91	3542	92
Attrition MHQ	31	2	147	6	178	5
Annual health questionnaire	27	2	96	4	123	3
All sources of exit data	1274	99	2569	101	3843	100
No exit data	20	—	95	—	115	—
Total	1294	—	2664	—	3958	—

^aPercentages do not sum to 100 because of rounding.

EFFECTS OF REMOVING PEOPLE WITH MISSING DATA

The effect of removing each source of missing data is documented in Table K.3, which shows the predicted exit values of the GHRI for the sample of 3843 reported in the main text (the row labeled "Original sample") and for successively smaller subsets of persons defined by who was removed. The row labeled "Attrition MHQ" means that just those individuals who had exit values only from that source were excluded from the analysis; the row labeled "Health questionnaire" excludes those omitted in the previous row *as well as* those who had exit values only from an annual health questionnaire. The fourth row labeled "Enrollment imputations" excludes all persons omitted from the previous rows as well as those who had had fairly crude enrollment values (imputed from age and pre-enrollment data or from the mean value for the entire sample). Finally, the last row labeled "Other predictions or substitutions" excludes *anyone* with a health perceptions score at either exit or enrollment that did not come from the actual exit or enrollment MHQ. The sample sizes shown in the last two rows of Table K.3 will not correspond exactly to the figures that would be calculated if one simply subtracted the relevant numbers shown in

Table K.1, because some individuals had imputed data for both enrollment and exit and had already been subtracted at an earlier step.

By and large, using the different samples (i.e., moving down the rows of Table K.3) does not qualitatively change our conclusions about the effects of cost-sharing (versus free care) on the health perceptions measure. That is, cost-sharing values are always higher than free-plan values, but the difference is never statistically significant.

Theoretically, the differences among the rows are large enough to matter in some situations, and other analysts, making other judgments or assumptions, might come to different conclusions than we have. To avoid unwittingly "manufacturing" a result of our own choosing, we selected our models based on preliminary information (outcomes in earlier years) and thoroughly discussed the merits of all our modeling decisions without reference to the direction of their effect. That is, we specified our final models without knowing or taking into account whether a given model would yield results that favored cost-sharing over free care or the reverse. We are satisfied that our decisions about how to handle persons with missing data have not distorted our findings in any material way.

Table K.3

PREDICTED EXIT VALUES FOR GENERAL HEALTH RATINGS INDEX
WHEN PERSONS WITH MISSING VALUES ARE
ELIMINATED FROM THE ANALYSIS

Successive Exclusions from Sample ^a	Sample Size	Free Plan	Cost-Sharing Plans	Free Minus Cost-Sharing Plans	t-Test Value
Original sample ^b	3843	67.4	68.0	-0.6	-1.3
Attrition MHQ	3665	67.5	68.0	-0.5	-1.1
Health questionnaire	3542	67.6	68.0	-0.4	-0.9
Enrollment imputations	3471	67.6	68.0	-0.4	-0.9
Other predictions or substitutions	2777	68.3	68.7	-0.4	-0.7

^aRows are successively smaller subsets and exclude persons whose health perceptions data came from the indicated sources.

^bOriginal means the sample used in the main text (see Table 5).

Appendix L

RESULTS FROM CLASSIFYING PEOPLE AS SIMPLY SICK OR WELL

INTRODUCTION

The measures for which we reported significant effects of health insurance plan—functional far vision and diastolic blood pressure—are continuous measures. The medical care system, however, sometimes views people as being in one of two classes—sick or well. We were interested, therefore, in determining if our results would hold even when we used dichotomous definitions for each of these health variables. We present here results for enrollees classified as sick using standard definitions of abnormal: diastolic blood pressure 90 mm Hg or higher and corrected vision 20/40 or worse. (The latter is a criterion often used to determine whether people must wear glasses to drive.)

UNADJUSTED VALUES

Findings are based on both raw and adjusted exit values, and they are essentially equivalent. Table L.1 shows the percentage distribution of exit blood pressure values by plan, and Table L.2 the distribution for far functional vision. (For Table L.2, line 2 is 20/20, line 3 is 20/25, line 4 is 20/30, line 5 is 20/40, etc.) Since the plans were initially balanced (as shown in Table 3 of the main text), these results should be a fairly accurate representation of the effects of plan.

As expected (see the first three rows of Table L.1), the free plan had somewhat fewer people in the higher ranges of blood pressure than did the cost-sharing plans. For instance, 14.2 percent of those on the free plan were at or above 90 mm Hg, as contrasted with 16.8 percent on the other plans. Similarly, the percentages of people with normal vision (see the first three rows of Table L.2) were higher on the free than the cost-sharing plans.

Table L.1

PERCENTAGE DISTRIBUTION OF ANALYTIC SAMPLE,
BY LEVEL OF DIASTOLIC BLOOD PRESSURE
AT EXIT AND BY PLAN

Diastolic Blood Pressure Group at Exit (mm Hg)	Cost-Sharing Plans				
	Cata- strophic	Inter- med.	Ind. Deduct.	Total	Free Plan
≤79	50	53	51	51	54
80-89	31	32	32	30	32
≥90	19	15	17	16.8	14.2
≥90, adjusted*	—	—	—	16.9	14.1

*Adjusted by multiple regression methods to eliminate any imbalance in initial values.

Table L.2

PERCENTAGE DISTRIBUTION OF ANALYTIC SAMPLE,
BY LEVEL OF FUNCTIONAL FAR VISION
AT EXIT AND BY PLAN

Functional Far Vision Group at Exit	Cost-Sharing Plans				
	Cata- strophic	Inter- med.	Ind. Deduct.	Total	Free Plan
20/20 or better	47	53	48	50	59
20/25 - 20/30	42	36	39	38	31
20/40 or worse	11.0	11.2	13.2	11.8	10.0
20/40 or worse, adjusted*	—	—	—	12.0	9.6

*Adjusted by multiple regression methods to eliminate any imbalance in initial values.

ADJUSTED VALUES

A more precise assessment of the effects of plan on the probability of being "sick" was done for both conditions. For diastolic blood pressure, a logistic regression was performed on the probability of being 90 mm Hg and above. We applied the coefficients from that regression to the entire sample and computed the expected number of people who would have exceeded that blood pressure level if everyone had had free care and if everyone had had cost-sharing insurance. For blood pressure, the regression coefficient on the free plan was -0.27 ($t = -2.36$, which was a slight overestimate because family correlation was not corrected for).

The results (last row of Table L.1, labeled ≥ 90 , adjusted) showed that if everyone had had free care, 14.1 percent would have been expected to be at or above 90 mm Hg, whereas if everyone had had cost-sharing plans, 16.9 percent would have been expected to be in that elevated group. These differences were very close to the raw values.

For vision, the logistic regression coefficient for the free plan was -0.28 ($t = -1.6$). As shown in the last row of Table L.2, if everyone had had free care, 9.6 percent of those with impaired natural vision would have had corrected far vision of 20/40 or worse. If everyone had had to share in the costs of care, 12.0 percent would have had 20/40 vision or worse. Again, these values were close to the raw values.

Of all enrollees, 47 percent had had perfect natural vision. Thus, we estimate that, with free care, 5.1 percent of the remainder of the population (9.6 percent times 53 percent) would have had serious impaired vision. With nonfree care, analogous arithmetic suggests that 6.4 percent would have had a serious vision problem.

Appendix M

PHYSICAL HEALTH MEASURES FOR PREGNANT WOMEN

Pregnant women might have scored differently on the Personal and Role Functioning measures solely because of their pregnancy. If rates of pregnancy were different on the free and cost-sharing plans, this could bias our final results.

To evaluate whether this was a problem, we ran our regressions with and without women who were pregnant at the time they left the experiment and completed the exit MHQ. The results are given in Table M.1.

Excluding pregnant women tended to raise the average values for both of these self-reported measures, but only for the cost-sharing plans. This did *not* account for the observed differences between the free and cost-sharing plans, so we concluded that pregnancy had not caused any bias in our results. Our reported analyses, therefore, included these women.

Table M.1

RESULTS ON PHYSICAL HEALTH MEASURES WITH AND WITHOUT PREGNANT WOMEN

Sample in Analysis	Sample Size	Free Plan	Cost-Sharing Plans	Free Minus Cost-Sharing Plans	t-Test Value
<i>Personal Functioning</i>					
Pregnant women					
Included	3862	85.3	85.3	-0.0	-0.0
Excluded	3797	85.3	85.4	-0.1	-0.1
<i>Role Functioning</i>					
Pregnant women					
Included	3861	95.4	95.1	0.3	0.4
Excluded	3796	95.4	95.4	0.0	0.1

REFERENCES

- Aitchison, J., and J.A.C. Brown, *The Lognormal Distribution, with Special Reference to its Uses in Economics*, Cambridge University Press, Cambridge, Massachusetts, 1957.
- Beck, A. T., *Depression—Causes and Treatment*, University of Pennsylvania Press, Philadelphia, 1967.
- Benham, L., and A. Benham, "The Impact of Incremental Medical Services on Health Status, 1963-1970," in R. Andersen, J. Kravitz, and O. W. Anderson (eds.), *Equity in Health Services: Empirical Analyses in Social Policy*, Ballinger Publishing Company, Cambridge, Massachusetts, 1975.
- Berwick, D., S. Cretin, and E. Keeler, *Children, Cholesterol, and Heart Disease: An Analysis of Alternatives*, Oxford University Press, New York, 1980, Table 7.1.
- Brook, R. H., D. M. Berman, K. N. Lohr, et al., *Conceptualization and Measurement of Physiologic Health for Adults: Vol. 3, Hypertension*, The Rand Corporation, R-2262/3-HHS, August 1980.
- Brook, R. H., K. N. Lohr, E. B. Keeler, et al., *Conceptualization and Measurement of Physiologic Health for Adults: Vol. 11, Hypercholesterolemia*, The Rand Corporation, R-2262/11-HHS, September 1981.
- Brook, R. H., J. E. Ware, Jr., W. H. Rogers, et al., "Does Free Care Improve Adults' Health? Results from a Randomized Controlled Trial," *The New England Journal of Medicine*, Vol. 309, 1983, pp. 1426-1434.
- Brown, M., *Lessons Learned from the Administration of the Rand Health Insurance Experiment*, The Rand Corporation, R-3095-HHS, May 1984.
- Carlson, R. J., *The End of Medicine*, John Wiley & Sons, New York, 1975.
- Clasquin, L., and M. E. Brown, *Rules of Operation for the Rand Health Insurance Study*, The Rand Corporation, R-1602-HEW, May 1977.
- Cochrane, A. L., A. S. St. Leger, and F. Moore, "Health Service 'Input' and Mortality 'Output' in Developed Countries," *Journal of Epidemiology and Community Health*, Vol. 32, 1978, pp. 200-205.
- Comrey, A. L., *Comrey Personality Scales*, Educational and Industrial Testing Services, San Diego, California, 1970.

- Costello, C. G., and A. L. Comrey, "Scales for Measuring Depression and Anxiety," *The Journal of Psychology*, Vol. 66, 1967, pp. 303-313.
- Dagenais, M. G., "Further Suggestions Concerning the Utilization of Incomplete Observations in Regression Analysis," *Journal of The American Statistical Association*, Vol. 66, 1971, pp. 93-98.
- Davies, A. R., and J. E. Ware, Jr., *Measuring Health Perceptions in the Health Insurance Experiment*, The Rand Corporation, R-2711-HHS, October 1981.
- Davies, A. R., C. D. Sherbourne, J. R. Peterson, and J. E. Ware, Jr. (eds.), *Scoring Manual: Adult Health Status and Patient Satisfaction Measure Used in Rand's Health Insurance Experiment*, The Rand Corporation, N-2190-HHS, forthcoming.
- Dohrenwend, B. S., B. P. Dohrenwend, and D. Cook, "Ability and Disability in Role Functioning in Psychiatric Patient and Non-Patient Groups," in J. D. Wing and H. Hafner (eds.), *Roots of Evaluation*, Oxford University Press, London, 1973, pp. 337-360.
- Dohrenwend, B. P., P. E. Shrout, G. Egri, and F. S. Mendelsohn, "Nonspecific Psychological Distress and Other Dimensions of Psychopathology," *Archives of General Psychiatry*, Vol. 37, 1980, pp. 1229-1236.
- Donald, C. A., and J. E. Ware, Jr., *The Quantification of Social Contacts and Resources*, The Rand Corporation, R-2937-HHS, October 1982.
- Donald, C. A., and J. E. Ware, Jr., "The Measurement of Social Support," in J. F. Greenley (ed.), *Research in Community and Mental Health*, Jai Press, Greenwich, Connecticut, 1984 (in press).
- Donald, C. A., J. E. Ware, Jr., R. H. Brook, et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. IV, Social Health*, The Rand Corporation, R-1987/4-HEW, August 1978.
- Hadley, J., *More Medical Care, Better Health?* The Urban Institute, Washington, D.C., 1982.
- Huber, P. J., *The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions*. *Fifth Berkeley Symposium*, 1965, University of California Press, Berkeley, California, 1967, pp. 221-233.
- Hulka, B. S., and J. C. Cassel, "The AAFP-UNC Study of the Organization, Utilization, and Assessment of Primary Medical Care," *American Journal of Public Health*, Vol. 63, 1973, pp. 494-501.
- Illich, I., *Medical Nemesis: The Expropriation of Health*, Pantheon Books, New York, 1977.

- Keeler, E. B., and J. E. Rolph, "How Cost Sharing Reduced Medical Spending of Participants in the Health Insurance Experiment," *Journal of the American Medical Association*, Vol. 249, 1983, pp. 2220-2222.
- Manning, W. G., Jr., J. P. Newhouse, and J. E. Ware, Jr., "The Status of Health in Demand Estimation: Beyond Excellent, Good, Fair, and Poor," in V. R. Fuchs (ed.) *Economic Aspects of Health*, The University of Chicago Press, Chicago, 1982, pp. 143-184.
- McGee, D., and T. Gordon, "Section 31. The Results of the Framingham Study Applied to Four Other US-Based Epidemiologic Studies of Cardiovascular Disease," in W. B. Kannel and T. Gordon (eds.), *The Framingham Study: An Investigation of Cardiovascular Disease* (DHEW Publication No. (NIH) 76-1083), National Heart and Lung Institute, National Institutes of Health, Bethesda, Maryland, April 1976.
- Miller, A. E., and M. G. Miller, *Options for Health and Health Care. The Coming of Post-Clinical Medicine*, John Wiley & Sons, New York, 1981.
- Miller, M. K., and C. S. Stokes, "Health Status, Health Resources, and Consolidated Structural Parameters: Implications for Public Health Care Policy," *Journal of Health and Social Behavior*, Vol. 19, 1978, pp. 263-379.
- Morris, C. N., "A Finite Selection Model for Experimental Design of the Health Insurance Study," *Journal of Econometrics*, Vol. 11, 1979, pp. 43-61.
- Myers, J. K., J. J. Lindenthal, M. P. Pepper, et al., "Life Events and Mental Status: A Longitudinal Study," *Journal of Health and Social Behavior*, Vol. 13, 1972, pp. 398-406.
- National Center for Health Statistics, *Limitations of Activity and Mobility Due to Chronic Conditions: United States, 1972*, (DHEW Publication No. (HRA) 75-1523), National Center for Health Statistics, Rockville, Maryland, 1974.
- Newhouse, J. P., "A Design for a Health Insurance Experiment," *Inquiry*, Vol. 11, 1974, pp. 5-27.
- Newhouse, J. P., and L. Friedlander, "The Relationship Between Medical Resources and Measures of Health: Some Additional Evidence," *Journal of Human Resources*, Vol. 15, 1980, pp. 200-218.
- Newhouse, J. P., W. G. Manning, C. N. Morris, et al., "Some Interim Results from a Controlled Trial of Cost Sharing in Health Insurance," *New England Journal of Medicine*, Vol. 305, 1981, pp. 1501-1507. (Also published as R-2847-HHS, The Rand Corporation, January 1982).

- Patrick, D. L., J. W. Bush, and M. M. Chen, "Toward an Operational Definition of Health," *Journal of Health and Social Behavior*, Vol. 14, 1973, pp. 6-23.
- Rice, D. P., and D. Wilson, "The American Medical Economy: Problems and Perspectives," *Journal of Health Politics, Policy, Law*, Vol. 1, 1976, pp. 151-172.
- Rogers, D. E., and R. J. Blendon, "Special Communication. The Changing American Health Scene. Sometimes Things Get Better," *Journal of the American Medical Association*, Vol. 237, 1977, pp. 1710-1714.
- Rogers, W. H., and P. Camp, *Refusal and Attrition in the Health Insurance Experiment*, The Rand Corporation, N-2195-HHS, forthcoming.
- Rubenstein, R. S., K. N. Lohr, R. H. Brook, et al., *Conceptualization and Measurement of Physiologic Health for Adults: Vol. 12. Vision Impairments*, The Rand Corporation, R-2262/12-HHS, July 1982.
- Smith, L. H., G. A. Goldberg, R. H. Brook, et al., *The Health Insurance Study Screening Examination Procedures Manual*, The Rand Corporation, R-2101-HEW, September 1978.
- Starr, P., "The Politics of Therapeutic Nihilism," *Working Papers*, Vol. IV, No. 2, 1976, pp. 48-55.
- Stewart, A. L., "The Reliability and Validity of Self-Reported Weight and Height," *Journal of Chronic Diseases*, Vol. 35, 1982, pp. 295-309.
- Stewart, A. L., R. H. Brook, and R. L. Kane, *Conceptualization and Measurement of Health Habits for Adults in the Health Insurance Study: Smoking*, The Rand Corporation, R-2374/1-HEW, June 1979.
- Stewart, A. L., R. H. Brook, and R. L. Kane, *Conceptualization and Measurement of Health Habits for Adults in the Health Insurance Study: Overweight*, The Rand Corporation, R-2374/2-HEW, July 1980.
- Stewart, A., J. E. Ware, Jr., and R. H. Brook, "The Meaning of Health: Understanding Functional Limitations," *Medical Care*, Vol. 15, 1977, pp. 939-952.
- Stewart, A. L., J. E. Ware, Jr., and R. H. Brook, "Advances in the Measurement of Functional Status: Construction of Aggregate Indexes," *Medical Care*, Vol. 19, 1981, pp. 473-488.
- Stewart, A. L., J. E. Ware, Jr., and R. H. Brook, *Construction and Scoring of Aggregate Functional Status Indexes*, The Rand Corporation, R-2551-1-HHS, August 1982a.

- Stewart, A. L., J. E. Ware, Jr., and R. H. Brook, *Construction and Scoring of Aggregate Functional Status Indexes: Appendixes*, The Rand Corporation, N-1706-1-HHS, August 1982b.
- Stewart, A. L., J. E. Ware, Jr., R. H. Brook, et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol II, Physical Health in Terms of Functioning*, The Rand Corporation, R-1987/2-HEW, July 1978.
- U.S. Public Health Service, "Life Tables," *Vital Statistics of the United States, Vol. II, Sec. 5-2* (DHHS Publication No. (PHS) 81-1104), U.S. Government Printing Office, Hyattsville, Maryland, 1980.
- Veit, C., and J. E. Ware, Jr., "The Structure of Psychological Distress and Well-Being in General Populations," *Journal of Consulting and Clinical Psychology*, Vol. 51, 1983, pp. 730-742.
- Ware, J. E., Jr., "Scales for Measuring General Health Perceptions," *Health Services Research*, Vol. 11, 1976, pp. 396-415.
- Ware, J. E., Jr., "The General Health Rating Index," in N. K. Wenger, M. E. Mattson, C. D. Furberg, and J. Elinson (eds.), *Assessment of Quality of Life in Clinical Trials of Cardiovascular Disease*, Le Jacq Publishing Co., New York, 1984 (in press).
- Ware, J. E., Jr., and A. H. Karmos, *Development and Validation of Scales to Measure Perceived Health and Patient Role Propensity: Vol. II of a Final Report*, NTIS Publication No. PB 288-331, National Technical Information Service, Springfield, Virginia, 1976.
- Ware, J. E., Jr., A. Davies-Avery, and C. A. Donald, *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. V, General Health Perceptions*, The Rand Corporation, R-1987/5-HEW, September 1978.
- Ware, J. E., Jr., A. Davies-Avery, and R. H. Brook, *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. VI, Analysis of Relationships Among Health Status Measures*, The Rand Corporation, R-1987/6-HEW, November 1980.
- Ware, J. E., Jr., R. H. Brook, A. R. Davies, et al., *Choosing Measures of Health Status for Individuals in General Populations*, The Rand Corporation, N-1642-HHS, January 1981.
- Ware, J. E., Jr., S. A. Johnston, A. Davies-Avery, et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. III, Mental Health*, The Rand Corporation, R-1987/3-HEW, December 1979.
- Ware, J. E., W. G. Manning, N. Duan, et al., "Health Status and the Use of Outpatient Mental Health Services," *American Psychologist*, Vol. 39, No. 10, 1984a (in press).

- Ware, J. E., Jr., C. T. Veit, and C. A. Donald, *Refinements in the Measurement of Mental Health for Adults in the Health Insurance Experiment*, The Rand Corporation, R-2737-HHS, 1984b (forthcoming).
- Williams, A. W., C. A. Donald, and J. E. Ware, Jr., *The Effects of Life Events and Social Supports on the Psychological Well-Being of Women and Men*, The Rand Corporation, 1984 (forthcoming).
- Williams, A. W., J. E. Ware, Jr., and C. A. Donald, "A Model of Mental Health, Life Events, and Social Supports Applicable to General Populations," *Journal of Health and Social Behavior*, Vol. 22, No. 4, 1981, pp. 324-336.
- World Health Organization, "Constitution of the World Health Organization," in *Basic Documents*, Geneva, 1948.
- Zung, W. K., "A Self-Rating Depression Scale," *Archives of General Psychiatry*, Vol. 12, 1965, pp. 63-70.

RAND/R-3055-HHS

