Health Insurance and the Demand for Medical Care

Evidence from a Randomized Experiment

Willard G. Manning, Joseph P. Newhouse,
Naihua Duan, Emmett Keeler,
Bernadette Benjamin, Arleen Leibowitz,
M. Susan Marquis, Jack Zwanziger

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 016B-80.

Library of Congress Cataloging in Publication Data

Health insurance and the demand for medical care.

Based on data from the RAND Health Insurance Study.
"Prepared under a grant from the U.S. Department of Health and Human Services."
"R-3476-HHS."
"February 1988."
Bibliography: p.
1. Insurance, Health—United States. 2. Medical care, Cost of—United States. I. Manning, Willard G. II. RAND Corporation. III. RAND Health Insurance Study. IV. United States. Dept. of Health and Human Services.
[DNLM: 1. Expenditures, Health—United States. 2. Health Services—research—United States. 3. Insurance, Health—economics—United States. W 275 AA1 H42]
HD7102.U4H413 1987 368.3'82'00973 87-17587
ISBN 0-8330-0864-1

Published by The RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138

R-3476-HHS

Health Insurance and the Demand for Medical Care

Evidence from a Randomized Experiment

Willard G. Manning, Joseph P. Newhouse,
Naihua Duan, Emmett Keeler,
Bernadette Benjamin, Arleen Leibowitz,
M. Susan Marquis, Jack Zwanziger

February 1988

Supported by the
U.S. Department of Health and Human Services

40 Years
1948-1988

RAND

PREFACE

This report examines the effects of varying levels of cost sharing on the demand for medical care and other health services. It presents the final results of the RAND Health Insurance Experiment with respect to annual utilization of medical services in the fee-for-service system. The experiment was a large scale social experiment designed to investigate the effects of alternative health insurance plans on the utilization of health services, health status, the quality of care, and patient satisfaction. The report updates the interim results presented in:

- Newhouse, J. P., et al., "Some Interim Results from a Controlled Trial of Cost Sharing in Health Insurance," *New England Journal of Medicine,* Vol. 305, December 17, 1981, pp. 1501–1507. (See also The RAND Corporation, R-2847-HHS, January 1982.)

Other reports have dealt with cost sharing and the demand for medical care and the effects on health status. They include:

- Brook, R. H., et al., "Does Free Care Improve Adults' Health? Results from a Randomized Controlled Trial," *New England Journal of Medicine,* Vol. 309, December 8, 1983, pp. 1426–1434. Also, R. H. Brook et al., *The Effect of Coinsurance on the Health of Adults,* The RAND Corporation, R-3055-HHS, December 1984.
- Keeler, E. B., et al., "How Free Care Reduced Hypertension of Participants in the RAND Health Insurance Experiment," *Journal of the American Medical Association,* Vol. 154, October 11, 1985, pp. 1926–1931.
- Keeler, E. B., et al., *The Demand for Episodes of Medical Treatment in the Health Insurance Experiment,* The RAND Corporation, R-3454-HHS, March 1988.
- Keeler, E. B., et al., "The Demand for Supplementary Health Insurance, or Do Deductibles Matter?" *Journal of Political Economy,* Vol. 85, August 1977, pp. 789–802.
- Keeler, E. B., et al., *The Demand for Episodes of Medical Treatment: Interim Results from the Health Insurance Experiment,* The RAND Corporation, R-2829-HHS, December 1982.

- Leibowitz, A., et al., "Effect of Cost Sharing on the Use of Medical Services by Children: Interim Results from a Randomized Controlled Trial," *Pediatrics*, Vol. 75, May 1985, pp. 942–951.
- Lohr, K. N., et al., "Use of Medical Care in the RAND Health Insurance Experiment: Diagnosis- and Service-Specific Analyses," *Medical Care* (Supplement), September 1986.
- Newhouse, J. P., et al., "The Findings of the RAND Health Insurance Experiment: A Response to Welch et al.," *Medical Care*, Vol. 25, February 1987, pp. 157–179.
- Newhouse, J. P., et al., "The Effect of Deductibles on the Demand for Medical Care Services," *Journal of the American Statistical Association*, Vol. 75, September 1980, pp. 525–523.
- O'Grady, K. F., et al., "The Impact of Cost Sharing on Emergency Department Use," *New England Journal of Medicine*, Vol. 313, August 22, 1985, pp. 484–490.
- Valdez, R. B., *The Effects of Cost Sharing on the Health of Children*, The RAND Corporation, R-3270-HHS, March 1986.
- Valdez, R. B., et al., "The Consequences of Cost Sharing for Children's Health," *Pediatrics*, Vol. 75, May 1985, pp. 957–961.

There have also been a series of reports on comparisons of the use of health services and health status between fee-for-service and health maintenance organizations. Those reports include:

- Davies, A. R., J. E. Ware, Jr., R. H. Brook, J. R. Peterson, and J. P. Newhouse, "Consumer Acceptance of Prepaid and Fee-for-Service Medical Care: Results from a Randomized Trial," *Health Services Research*, Vol. 21, August 1986, pp. 429–452.
- Manning, W. G., et al., "A Controlled Trial of the Effect of a Prepaid Group Practice on Use of Services," *New England Journal of Medicine*, Vol. 310, June 7, 1984, pp. 1505–1510. (See also The RAND Corporation, R-3029-HHS, September 1985.)
- Sloss, E. M., E. B. Keeler, R. H. Brook, B. H. Operskalski, G. A. Goldberg, and J. P. Newhouse, "Effect of a Health Maintenance Organization on Physiologic Health: Results from a Randomized Trial," *Annals of Internal Medicine*, Vol. 106, January 1987, pp. 130–138.
- Ware, J. E., et al., "Comparison of Health Outcomes at a Health Maintenance Organization with Those of Fee-for-Service Care," *The Lancet*, Vol. 1, No. 848, May 3, 1986, pp. 1017–1022.

The reports on oral health status and the demand for dental and mental health services include:

- Bailit, H., et al., "Does More Generous Dental Insurance Coverage Improve Oral Health?" *Journal of the American Dental Association*, Vol. 110, May 1985, pp. 701–707.
- Manning, W. G., et al., "The Demand for Dental Care: Evidence from a Randomized Trial in Health Insurance," *The Journal of the American Dental Association,* Vol. 110, June 1985, pp. 895–902. (See also The RAND Corporation, R-3225-HHS, August 1986.)
- Manning, W. G., et al., "Cost Sharing and the Demand for Ambulatory Mental Health Services," *American Psychologist,* Vol. 39, October 1984, pp. 1090–1100.
- Manning, W. G., et al., "How Cost Sharing Affects the Use of Ambulatory Mental Health Services," *Journal of the American Medical Association*, Vol. 256, October 1986, pp. 1930–1934.
- Wells, K. B., et al., *Cost Sharing and the Demand for Ambulatory Mental Health Services*, The RAND Corporation, R-2960-HHS, September 1982.

Most of these studies have used statistical models developed to address the difficult estimation problems faced with the highly skewed data on the use of health services, and economic models developed to estimate the effect of price in the absence of upper limits on the out-of-pocket costs paid by the family. These include:

- Duan, N., "Smearing Estimate: A Nonparametric Retransformation Method," *Journal of the American Statistical Association*, Vol. 78, September 1983, pp. 605–610.
- Duan, N., et al., "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Economic and Business Statistics,* Vol. 1, April 1983, pp. 115–126. Also The RAND Corporation, R-2754-HHS, January 1982.
- Duan, N., et al., "Choosing Between the Sample-Selection Model and the Multi-Part Model," *Journal of Business and Economic Statistics,* Vol. 2, July 1984, pp. 283–289.
- Keeler, E. B., et al., *The Demand for Episodes of Medical Treatment in the Health Insurance Experiment,* The RAND Corporation, R-3454-HHS, March 1988.
- Keeler, E. B., et al., "Deductibles and the Demand for Medical Care Services: The Theory of a Consumer Facing a Variable Price Schedule Under Uncertainty," *Econometrica,* Vol. 45, April 1977, pp. 641–656.

- Keeler, E. B., et al., *The Demand for Episodes of Medical Treatment: Interim Results from the Health Insurance Experiment,* The RAND Corporation, R-2829-HHS, December 1982.
- Manning, W., et al., "A Two-Part Model of the Demand for Medical Care: Preliminary Results from the Health Insurance Experiment," in J. van der Gaag and M. Perlman (eds.), *Health, Economics, and Health Economics,* North Holland, Amsterdam, 1981.
- Manning, W. G., *Estimating Health Demand Functions with Health Insurance Data,* The RAND Corporation, N-2729-HHS, March 1988.
- Manning, W. G., et al., "Monte Carlo Evidence on the Choice Between Sample Selection and Two-Part Models," *Journal of Econometrics,* Vol. 35, May 1987.
- Newhouse, J. P., et al., "On Having Your Cake and Eating It Too: Econometric Problems in Estimating the Demand for Health Services," *Journal of Econometrics,* Vol. 13, August 1980, pp. 365–390.
- Phelps, C. E., and J. P. Newhouse, *Coinsurance and The Demand for Medical Services,* The RAND Corporation, R-964-1-OEO/NC, October 1974.

The present report should be of interest to persons studying the demand for health services and the role of health insurance, as well as those interested in applied econometrics and statistics. An abridged version of this report (less appendixes and some technical material) was published in *The American Economic Review,* June 1987.

SUMMARY

Over the past four decades, medical care costs have grown about 4 percent per year in real terms, and the share of GNP devoted to medical care has increased from 4 to 11 percent between 1950 and 1984 (Levit et al., 1985). The most prominent explanation of this rapid increase has emphasized the spread of health insurance, which has generated demand for both a higher quality and an increased quantity of medical services.

No one has shown, however, that the spread of health insurance can quantitatively account for most of the sustained rise in health expenditure. If it cannot, the widespread presumption that distorted prices (because of insurance) are inducing excess resources in medical care is not necessarily correct. Central to appraising the quantitative role of insurance, of course, is the magnitude of the demand response to changes in insurance. The literature exhibits substantial disagreement, by a factor of 10 or more, about the price elasticity, or coinsurance elasticity, of demand.

Such disagreement is not surprising in light of the problems of using nonexperimental data to estimate elasticities, including the problem of adverse selection—the incentive for those in poor health to acquire more generous health insurance policies.

In light of the uncertainty about how demand responds to insurance-induced changes in price, and the importance for both public and private decisions of quantifying that response, the federal government initiated the RAND Health Insurance Experiment (HIE) in 1974, one aim of which was to narrow uncertainty about this issue. In this study, we report the results of that experiment. Our findings have implications for the role of insurance in explaining the postwar increase in medical expenditure, as well as for the magnitude of the welfare loss from health insurance.

DATA AND SAMPLE

Design of the RAND Health Insurance Experiment

Between November 1974 and February 1977, the HIE enrolled families in six sites: Dayton, Ohio; Seattle, Washington; Fitchburg, Massachusetts; Franklin County, Massachusetts; Charleston, South Carolina; and Georgetown County, South Carolina.

Of the 7,791 persons who participated in the experiment, 5,809 were assigned to one of 14 different fee-for-service insurance plans; another 1,982 were enrolled in prepaid group practice plans. The fee-for-service insurance plans, the main focus of this report, had different levels of cost sharing, which varied over two dimensions: the coinsurance rate (percentage paid out of pocket) and an upper limit on annual out-of-pocket expenses. The coinsurance rates were 0, 25, 50, or 95 percent. Each plan had an upper limit (the Maximum Dollar Expenditure or MDE) on annual out-of-pocket expenses of 5, 10, or 15 percent of family income, up to a maximum of $1,000. Beyond the MDE, the insurance plan reimbursed all covered expenses in full.

Families were assigned to these insurance plans using the Finite Selection Model. This model was used to achieve as much balance across plans as possible while retaining randomization; that is, it minimizes the correlation between the experimental treatments and health, demographic, and economic covariates. The enrolled sample is for the most part a random sample of each site's nonaged population, but some groups were not eligible: (1) those 62 years of age and older at the time of enrollment; (2) those with incomes in excess of $25,000 in 1973 dollars (or $58,000 in 1984 dollars); this excluded 3 percent of the families contacted; (3) those eligible for the Medicare disability program; (4) those in jails or institutionalized for indefinite periods; (5) those in the military or their dependents; and (6) veterans with service-connected disabilities.

Medical Use

We focus primarily on the use of medical services other than outpatient psychotherapy and dental services. We do, however, summarize results for dental services. All use measures were derived from claims data.

Plan of Analysis

We contrast the use of services for alternative insurance plans using analysis of variance (ANOVA) and multiple regression methods that also control for site, health status, sociodemographic, and economic variables. The reliability of these and other methods are examined in a split sample analysis.

We also examine differences in the response to health insurance across subgroups of interest—children versus adults, the sick versus the well, and the poor versus the rich.

Empirical Results

Main Effects. The data from the HIE clearly show that the use of medical services responds to changes in the amount paid out of pocket. The per capita expenses on the free plan (no out-of-pocket costs) are 45 percent higher than those on the plan with a 95 percent coinsurance rate, subject to an upper limit on out-of-pocket expenses of at most $1,000 dollars per year. Spending rates on plans with an intermediate level of cost sharing lie between these two extremes.

The largest decreases in the use of outpatient services occurs between the free and 25 percent plans, with smaller but statistically significant differences between the 25 percent and other family coinsurance plans ($p < 0.01$).

Cost sharing affects the number of medical contacts, but not the intensity of each of those contacts. In other words, the differences in expenditures across plans reflect real variation in the number of contacts rather than an increase in the intensity or charge per service. For example, outpatient expenses on the free plan are 67 percent higher than those on the 95 percent plan, whereas outpatient visit rates to physicians and other health providers are 66 percent higher than those on the 95 percent plan. A similar pattern holds more weakly for inpatient care: Inpatient expenses are 30 percent higher on the free plan than on the 95 percent plan, whereas admission rates are 29 percent higher.

There are no significant differences among the family coinsurance (25, 50, and 95 percent) plans in the use of inpatient services. There is also no significant effect of cost sharing on the use of inpatient services by children.

The individual deductible plan exhibits a somewhat different pattern from the other cost sharing plans. This plan has free inpatient care, but a 95 percent coinsurance rate (up to a $150 per person or $450 per family annual maximum) for outpatient services. Total expenditures on this plan are significantly less than on the free plan ($p < 0.02$). This overall response is the sum of a one-third reduction in outpatient expenses ($p < 0.0001$), and an insignificant, less than one-tenth, reduction in inpatient expenses. Thus, this plan looks like a combination of the 50 or 95 percent plans for outpatient care and the free or 25 percent plan for inpatient care. The admission rate for the individual deductible plan lies roughly midway between the free plan and family coinsurance plan rates, suggesting a nontrivial cross price elasticity between inpatient and outpatient services.

Use by Subgroups. We do not observe a statistically significant difference in the response to insurance plan across income groups.

We observed about the same outpatient response to insurance plan for children (ages less than 18) as for adults, but children are less plan-responsive for inpatient care. As we observed with a subset of these data (Newhouse et al., 1981, 1982; Leibowitz et al., 1985), we cannot reject the hypothesis that admission rates for children show no response to insurance coverage. By contrast, adults have significantly lower use of inpatient services on the family pay plans than they do on the free plan.

Although health status was a strong predictor of expenditure levels, we observed no differential response to health insurance coverage between the healthy and the sickly.

The six sites in the HIE were selected to reflect a spectrum of city sizes, waiting times to appointment, and physician to population ratios. One concern was that the response to insurance coverage could vary according to the complexity of the medical market or to the excess demand in the medical delivery system, but this was not observed.

We enrolled families for three or five years to see if the response to insurance changed over time and if the duration of enrollment mattered. The free plan might generate transitorily high demand; the 95 percent plan might generate postponement of demand at the end of the experiment. Neither effect was found for medical care, but there was a major transitory shift in demand for dental care at the beginning of the study.

COMPARISONS WITH THE LITERATURE

Our results indicate that demand elasticities for medical care are nonzero and indeed that the response to cost sharing is nontrivial. To make comparisons with other studies, we have to estimate the price response in the absence of an upper limit on out-of-pocket expenditures. The response to insurance plan is a mix of the response to a positive price below the limit and free care (no out-of-pocket expense) above the limit. Thus, the plan response is less than the pure price response in the absence of such a limit.

We used three alternative methods to estimate the pure price response. These three methods suggest that price elasticities for medical care are in the −0.1 to −0.2 range—values that are consistent with those in the lower range of the nonexperimental literature and which vary from −0.1 to −2.1.

Our estimates of demand response imply that, given technology, the spread of health insurance can account for only a modest portion of the postwar rise in medical expenditure.

THE WELFARE LOSS FROM MORAL HAZARD

With health insurance, individuals purchase more medical care than they would have if they paid full price. As a result, the individual values the additional service at less than its market price, which reflects the cost of the additional service to society. Under a number of strong assumptions (including that gross medical care prices are competitive and there are no externalities), our estimates imply a non-trivial welfare loss from first dollar health insurance coverage. An approximation of the loss from moving from a universal 95 percent plan (with a $1,000 upper limit on out-of-pocket expenses) to the free care plan is $37 to $60 billion, as against an expenditure around $200 billion on these services in 1984 by the under 65 population.

From the $37–$60 billion figure must be deducted some amount for the reduced risk in the free plan relative to the 95 percent plan. Usual values for risk aversion, however, would suggest that the deduction is small in the presence of a $1,000 cap. Thus, we expect that the welfare loss from more generous, first dollar coverage health insurance is substantial.

Central to issues of welfare loss is the extent to which insurance induced technological change that consumers would not have been willing to pay for. This question cannot be addressed by the experiment and is very difficult to address with nonexperimental data. However, because countries with widely varying institutional arrangements exhibit similar rates of change, consumers may be willing to pay for much of the new technology even if faced with its full cost.

WAS IT WORTH IT?

Between 1982 and 1984, the time during which results from the experiment were first published, there was a substantial increase in cost sharing among major employers and a decline in hospital days among the under 65. If only a small portion of this change was attributable to the experimental results, the experiment paid for itself rather quickly.

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 Andre, Rae Archibald, Marie Brown, Maureen Carney, Lorraine Clasquin, and Kenneth Krug (administration); Darlene Blake, Carol Edwards, Janet Hanley, Joan Keesey, Bryant Mori, Suzanne Polich, Martin Seda, David Stewart, and Beatrice Yormark (data processing); and Geraldine Jaimovitch, Marilyn Martino, and Joice Polin (secretarial assistance).

The survey data were collected by the National Opinion Research Center and by Mathematica, Inc.; the utilization data were collected by Glen Slaughter and Associates and Group Health Cooperative of Puget Sound. 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 Haveman and three anonymous reviewers for the *American Economic Review* provided thoughtful and constructive critiques of this report, for which we thank them.

CONTENTS

Appendix

TABLES

I. INTRODUCTION

Over the past four decades medical care costs have grown about 4 percent per year in real terms, and the share of GNP devoted to medical care has increased from 4.4 percent in 1950 to 10.6 percent in 1984 (Levit et al., 1985). The most prominent explanation of this rapid increase has emphasized the spread of health insurance, which has generated demand for both a higher quality and an increased quantity of medical services (Feldstein, 1971, 1977). In turn, the spread of health insurance has been linked to the exemption of employer-paid health insurance premiums from the individual income tax (Feldstein and Allison, 1974; Feldstein and Friedman, 1977; Pauly, 1986). Thus, the increase in expenditure is often portrayed as a type of market failure induced by public policy, although such an argument is not universally accepted (Barer et al., 1979; Evans, 1984; Goddeeris and Weisbrod, 1985).

No one has shown, however, that the spread of health insurance can quantitatively account for most of the sustained rise in health care expenditure (Pauly, 1986). If it cannot, the widespread presumption that distorted prices (because of insurance) are inducing excess resources in medical care is not necessarily correct. Central to appraising the quantitative role of insurance, of course, is the magnitude of the demand response to changes in insurance. The literature exhibits substantial disagreement, by a factor of 10 or more, about the price elasticity, or coinsurance elasticity, of demand (Rosett and Huang, 1973; Davis and Russell, 1972; Phelps and Newhouse, 1974a; Goldman and Grossman, 1978; Colle and Grossman, 1978; Newhouse and Phelps, 1974; Newhouse and Phelps, 1976).[1]

Such disagreement is not surprising in light of the problems of using nonexperimental data to estimate elasticities (Newhouse et al., 1980a). In cross-sectional data insurance is endogenous; those who expect to demand more services have a clear incentive to obtain more complete insurance, either by selecting a more generous option at the place of employment, by working for an employer with a generous insurance plan, or by purchasing privately more generous coverage.

Ignoring this selection issue (i.e., treating insurance as exogenous) has generally produced results showing that demand for medical care responds to insurance-induced variation in price. Treating insurance

[1]The elasticity estimates at the mean vary from around −0.1 to −2.1.

as endogenous, however, has generally led to coefficients with confidence intervals that are insignificantly different from zero at conventional levels (Newhouse and Phelps, 1976).[2]

That upward bias may be present is suggested by results from several natural experiments that compared demands of the same individuals before and after their group insurance changed (Scitovsky and Snyder, 1972; Scitovsky and McCall, 1977; Phelps and Newhouse, 1972; Beck, 1974). In these cases the change in insurance is presumptively exogenous, and the elasticity estimates cluster near the low end of those cited above. But natural experiments have no control group, so that any other factor that changed over time is perfectly confounded with the insurance change. Moreover, the samples available in such studies are not necessarily representative of the general population, and the changes in insurance that could be studied were limited to those that occurred in the natural experiment. Hence, these results too have been suspect.[3]

In light of the uncertainty about how demand responds to insurance-induced changes in price, and the importance for both public and private decisions of quantifying that response, the federal government initiated the RAND Health Insurance Experiment (HIE) in 1974, one aim of which was to narrow uncertainty about this issue (Newhouse, 1974). In this article we report the results of that experiment. Our findings have implications for the role of insurance in explaining the postwar increase in medical expenditure, as well as for the magnitude of the welfare loss from health insurance.

The HIE had several objectives other than improved estimates of how demand responds to insurance. Four such objectives merit mention here:

1. Many poor individuals are insured through public programs; whether the demand response differs for the poor is therefore an issue in decisions on the scope of these programs.

2. Insurance need not be uniform across various medical services. In fact, second-best pricing implies that coverage should be more generous for less price elastic (or less insurance elastic) services (Ramsey, 1927; Zeckhauser, 1970; Baumol and Bradford, 1970). We therefore wished to learn if insurance elasticities differed for various types of

[2]Although many believe this failure to reject the null hypothesis when insurance is treated as endogenous occurs because the insurance variable is only weakly identified, the magnitude of any upward bias in elasticity estimates from treating insurance as exogenous remains unknown. Hausman (or Wu) type tests have not been used to test for endogeneity, but if they failed to reject the null hypothesis of exogeneity, it could be for lack of power because of a lack of a useful set of instruments.

[3]For reviews of the nonexperimental demand literature and a discussion of its methodological problems, see Newhouse (1978, 1981).

medical services. In particular, are demand elasticities greater for outpatient physician services, psychotherapy, and preventive services, which would accord with the observed lesser coverage of these services?[4]

3. The public financing of medical care has been justified by its status as a merit good (Musgrave, 1959) and in particular the claim that the consumption of medical services leads to improved health, which can generate externalities (Lindsay, 1969; Culyer, 1971, 1976, 1978; Pauly, 1971; Evans, 1984). Thus, we sought to quantify how the change in the consumption of medical services at the margin might affect health. The answer to this question would inform the political debate about the benefits of public financing of medical care services for the indigent and would also inform the insurance decisions of private agents such as employers and unions.

4. For the past decade public policy has promoted Health Maintenance Organizations (HMOs) on the grounds that such organizations were more efficient in the delivery of services. Almost all evidence of lower cost, however, came from uncontrolled settings, leaving unresolved the question of whether selection of healthier members or more efficient treatment was responsible for lower costs in HMOs (Luft, 1981). In the latter case effects on health outcomes were usually assumed to be zero. Therefore, we sought to decompose the observed lower use of services at one HMO into the pure effect of the HMO, on the one hand, and treating a possibly less sickly group of enrollees, on the other. Moreover, we sought to determine whether any reduced use of services affected health status and satisfaction.

The focus of this report is on the first two questions, although we summarize briefly our findings on the latter two.

---

[4]Other explanations, not mutually exclusive, for the lower coverage of these services include greater loading charges and asymmetric information between insurer and insured.

II. DATA AND SAMPLE

DESIGN OF THE RAND HEALTH INSURANCE EXPERIMENT[1]

Between November 1974 and February 1977, the HIE enrolled families in six sites:[2] Dayton, Ohio; Seattle, Washington; Fitchburg, Massachussetts; Franklin County, Massachusetts; Charleston, South Carolina; and Georgetown County, South Carolina. The sites were selected to represent the four census regions; to represent the range of city sizes (a proxy for the complexity of the medical delivery system); to cover a range of waiting times to appointment and physician per capita ratios (to test for the sensitivity of demand to nonprice rationing); and to include both urban and rural sites in the North and the South. Table 2.1 displays the characteristics for these six sites.

Families participating in the experiment were assigned to one of 14 different fee-for-service insurance plans or to a prepaid group practice; additionally, some members already enrolled in the prepaid group practice were enrolled as a separate group. The fee-for-service insurance plans—the main focus of this report—had different levels of cost sharing, which varied over two dimensions: the coinsurance rate (percentage paid out of pocket) and an upper limit on annual out-of-pocket expenses. The coinsurance rates were 0, 25, 50, or 95 percent. Each plan had an upper limit (the Maximum Dollar Expenditure or MDE) on annual out-of-pocket expenses of 5, 10, or 15 percent of family income, up to a maximum of $1000.[3] Beyond the MDE, the insurance plan reimbursed all covered expenses in full.

Covered expenses included virtually all medical services.[4] One plan had different coinsurance rates for inpatient and ambulatory medical services (25 percent) than for dental and ambulatory mental health

[1]Newhouse (1974) and Brook et al. (1979) provide fuller descriptions of the design. Newhouse et al. (1979) discuss the measurement issues for the second generation of social experiments, to which the HIE belongs. Ware et al. (1980a, 1980b) discuss many aspects of data collection and measurement for health status.

[2]Single persons constituted a one-person family.

[3]The maximum was $750 in some site-years for the 25 percent coinsurance plans.

[4]See Clasquin (1973) for a discussion of the rationale for the HIE structure of benefits. Nonpreventive orthodontia and cosmetic surgery (related to preexisting conditions) were not covered. Also excluded were outpatient psychotherapy services in excess of 52 visits per year per person. In the case of each exclusion, it is questionable whether anything could have been learned about steady-state demand during the three- to five-year lifetime of the experiment.

Table 2.1

CHARACTERISTICS OF SITES

Site	Census Region	Population of Urbanized Area or County (1970)	Primary Care Physicians per 100,000 Population (1972)[a]	Days Spent Waiting for an Appointment with a Primary Care Physician, New Patient[b] (1973, 1974)	Median Family Income, $ (1969)	Percent over Age 24 with Less Than 5 Years Education (1970)	Percent Black (1970)	Number of Enrollees
Seattle, WA	West	1,200,000	59	4.1	11,800	1.8	3	1,220[c]
Dayton, OH	North Central	690,000	41	7.5	11,400	3.3	13	1,140
Charleston, SC	South	230,000	33	15.9	8,300	6.2	25	780
Fitchburg-Leominster, MA	Northeast	78,000	30	25.0	10,000	4.3	1	724
Franklin County, MA	Northeast	59,000	46	9.2	9,900	2.8	1	891
Georgetown County, SC	South	34,000	44	0	6,400	20.6	48	1,061
United States	—	—	46	7.1	9,600	5.5	11	—

[a]Includes general practitioners, family practitioners, internists, and pediatricians.

[b]Physicians who do not use appointment systems and take patients on a first-come, first-served basis are valued as having zero wait time. All physicians sampled in Georgetown County at the time of the survey accepted patients on this basis. For other sites, the values are negligibly changed if only physicians using appointment systems are included.

[c]An additional 1892 participants were enrolled in the Group Health Cooperative of Puget Sound.

services (50 percent). And on one plan, the families faced a 95 percent coinsurance rate for outpatient services, subject to a $150 annual limit on out-of-pocket expenses per person ($450 per family). In this plan, all inpatient services were free; in effect, this plan had approximately an outpatient individual deductible.[5]

A simple example illustrates how an HIE plan worked. Consider a family facing a 25 percent coinsurance rate for all health services and an upper limit on out-of-pocket expenses of $1,000. For the first $4,000 of expenditures on any health service (dental, medical, or mental health), the family paid 25 percent of the bill and the insurance company paid 75 percent. Beyond that point, the family paid nothing more out of pocket that year, because it had already paid $1,000 (= 0.25 × $4,000). At the beginning of the new accounting year, the family again paid 25 percent of the bill, until it reached its annual upper limit on out-of-pocket expenses.

As this example illustrates, the family's response to an experimental insurance plan is an amalgam of responses to the coinsurance rate below the upper limit on out-of-pocket expenses and free care beyond the limit. In this report, we examine the overall effect of the insurance plan. Keeler et al. (1988) provide a detailed decomposition of the family's response into the separate responses to the coinsurance rate and the upper limit; in Sec. V we do provide some information on the response to a constant price schedule. See Keeler et al. (1977a) for a theoretical discussion of the response to a multipart tariff, such as the experimental insurance plans, in the face of uncertainty.

Families were assigned to these insurance plans using the Finite Selection Model (Morris, 1979). This model was used to achieve as much balance across plans as possible while retaining randomization; that is, it minimizes the correlation between the experimental treatments and health, demographic, and economic covariates.

Families were enrolled as a unit with only eligible members participating. No choice of plan was offered; the family could either accept the experimental plan or choose not to participate. To prevent refusals, families were given a lump-sum payment greater than the worst-case outcome in their experimental plans relative to their previous plan; thus, families were always better off financially for accepting the enrollment offer. Moreover, because of a bonus for completion,

[5]The coinsurance rate for the 95 percent and individual deductible plans was actually 100 percent in the first year of Dayton, the first site. The rate was changed to 95 percent for all other site-years of the experiment to increase the incentive to file claims, although there was no statistical evidence at that time of underfiling. Subsequent analysis has shown that the mean outpatient physician expenditure on the 95 percent coinsurance plans relative to the free-care plan is understated by about 5 to 10 percent because of a lower propensity to file claims (Rogers and Newhouse, 1985).

they were always better off completing the study. Hence, there is a theoretical presumption of no bias from refusal or attrition. In fact, we have detected negligible effects from refusal and attrition (Brook et al., 1983; O'Grady et al., 1985; Newhouse et al., 1987).[6]

The lump-sum payment was an unanticipated change in income and should negligibly affect the response to cost sharing. We show below that there was no measurable effect of these payments.

The family's nonexperimental coverage was maintained for the family by the HIE during the experimental period, with the benefits of the policy assigned to the HIE. If the family had no coverage, the HIE purchased a policy on its behalf. Thus, no family could become uninsurable as a result of participation in the study.

To study methods effects, the HIE employed four randomized subexperiments. First, to increase precision in measuring changes in health status, 60 percent of the households were given a physical examination at entry into the study; to test for a possible stimulus to utilization, the remaining households received no examination. Second, to measure sick and work-loss days, and telephone consultations with physicians, some households filled out a diary on contacts with the health care system and on time lost as a result of illness. To test for a possible stimulus to utilization, some households filled out no forms, some filled them out weekly, and some biweekly. Third, to test for transitory aspects of the study (Metcalf, 1973; Arrow, 1975), 70 percent of the households were enrolled for three years; the remainder for five years. Fourth, to make families financially no worse off for participating in the study, participants were paid a lump-sum payment. To test for a possible stimulus to utilization, 40 percent of the families were given an unanticipated increase in their lump-sum payment during the second to last year of the study.

There are two potential threats to the balance of health and other characteristics across the insurance plans: nonrandom refusal of the offer to participate and nonrandom attrition from the study. Refusals of the plan offer varied from 6 percent on the free plan to 23 percent on the 95 percent coinsurance plans in the non-Dayton sites (see Brook et al., 1983).[7] Analysis of these refusals to participate indicate

[6]The details of the lump-sum payment rules can be found in Clasquin and Brown (1977).

[7]Data from Dayton are incomplete and hence have not been analyzed, but the refusal of the enrollment offer across all plans in Dayton was only 7 percent. Additionally, we have compared the group that enrolled on all plans with the group that completed baseline interviews but did not enroll. The only significant difference was that children are overrepresented by a modest amount in the group that enrolled (Morris, 1985). No significant preexperimental differences were found for self-reported utilization and health status (Morris, 1985). Age is explicitly controlled for in our analysis.

that the only significant difference between those who accepted and those who rejected the offer was that the latter had lower education and income. Income is controlled for in our analysis and education had no detectable (partial) effect on use. There is no evidence that those who rejected the offer to participate were sicker, or that there was an interaction among plan, sickness, and refusal of the offer.

Individuals on the cost sharing plans were more likely to leave the study early than were individuals on the free plan. These early departures were also sicker on average than those who stayed. Thus people on the cost sharing plans at the end of the study are healthier on average than those on the free plan. This could lead to an overestimate of the response to insurance plan. To correct for such bias, we include health status measures as covariates.

THE SAMPLE

The enrolled sample is for the most part a random sample of each site's nonaged population, but some groups were not eligible.[8] Table 2.2 gives the sample by plan and site; it excludes the 1,982 persons in the HMO experiment. Note that plans are not perfectly balanced by site; in particular, no one was enrolled on the 50 percent plan in

Table 2.2

ENROLLMENT SAMPLE, BY SITE

Plan	Dayton	Seattle	Fitch-burg	Franklin County	Charles-ton	Georgetown County	Total
Free	301	431	241	297	264	359	1893
25 percent[a]	260	253	125	152	146	201	1137
50 percent	191	0	56	58	26	52	383
95 percent	280	253	113	162	146	166	1120
Individual deductible	105	285	188	220	196	282	1276
Total	1137	1222	723	889	778	1060	5809

[a]Includes those with 50 percent coinsurance for dental and mental health and 25 percent coinsurance for all other services.

[8]The ineligible groups include: (1) those 62 years of age and older at the time of enrollment; (2) those with incomes in excess of $25,000 in 1973 dollars (or $58,000 in 1984 dollars)—this excluded 3 percent of the families contacted; (3) those eligible for the Medicare disability program; (4) those in jails or institutionalized for indefinite periods; (5) those in the military or their dependents; and (6) veterans with service-connected disabilities.

Seattle, and about half of those on the 50 percent plan are in Dayton, whereas only 20 percent of all participants are in Dayton. Table 2.3 gives the sample size used in our estimates below.[9]

DEPENDENT VARIABLES

In this report, we focus primarily on the use of medical services other than outpatient psychotherapy and dental services.[10] We do, however, summarize results for those services below. The medical services we do consider include all inpatient services and all purchases of drugs and supplies. We examine five different measures of use: the probability of any medical use during the year, the probability of any inpatient use during the year, the number of outpatient visits to any

Table 2.3

ESTIMATION SAMPLE

Plan	Person-Years
Free	6,822
25 percent	4,065
50 percent	1,401
95 percent	3,727
Individual deductible	4,175
Total	20,190

[9]About 3 percent of the actual participant-years are truncated because the participant withdrew partway through an accounting year. We do not use such participants in the estimation sample because the four-part model (see below) requires equal time periods for each observation. If a person is only observed for one quarter and the expenditure distribution is lognormal, the annual distribution is not simply the quarterly distribution scaled up by a factor of four; i.e., the lognormal does not convolute. The sample used in this analysis more specifically includes enrollees during each full year that they participated and the last accounting year in the study for those who died. We excluded data on partial years of participation by newborns. (Their expenses in the hospital at the time of birth, however, are attributed to the mother.) We tested the legitimacy of excluding those with partial years by comparing expenditure rates of part-year persons, adjusted for time at risk, with what they would have spent if they behaved like full-year people. Specifically, we regressed actual expenditure minus (time at risk times the four-part prediction) on plan dummy variables. We could not reject the null hypothesis of no difference by plan (χ^2 (4) = 2.67, $p > 0.50$). The estimated effect of including part-year participants is to negligibly increase the estimated response to plan.

[10]See Manning et al. (1984b) and Wells et al. (1982) for additional results on the use of mental health care, and Manning et al. (1985a) for additional results on dental use. Mental health care use is on the order of 4 percent of the expenditures discussed here.

health provider,[11] the number of admissions to a hospital or nursing home, and total expenditures on medical care.[12] All of the measures of use are derived from claims data.[13]

Appendix A contains summary statistics for medical expenditures.

INDEPENDENT VARIABLES

Although we present sample means by plan, we also present results controlling for site, health status, sociodemographic, and economic variables.

Insurance Plan Variables

Rather than impose a functional form, we have conservatively used dummy variables for insurance plans. We have grouped the insurance plans into five groups: (1) the free plan (no out-of-pocket cost to the family); (2) 25 percent coinsurance rate plans for medical services; (3) 50 percent coinsurance rate plans for medical services; (4) 95 percent coinsurance rate plans for medical services; and (5) the plan with a 95 percent coinsurance rate for outpatient services (subject to a limit of $150 per person or $450 per family per year) and free inpatient care. The middle three groups we call the family pay plans.

We did not differentiate the plans by the size of the upper limit on out-of-pocket expenses. Our data show that the lower the limit, the greater the likelihood that a family will exceed the limit and receive free care for part of the year (see Appendix B). Nonetheless, when we included variables for the upper limit, the variables were insignificant ($p > 0.50$) and over two-thirds of the coefficients were of the wrong sign. Using results on episodes of treatment and the timing of treatment (Keeler et al., 1982), we estimate that the differences between the upper limit groups for a given level of coinsurance is less than 4 percent of the plan mean. At the level of annual expenditure per person, we do not have the precision to pick up such a small effect.

In the analysis of inpatient use, we grouped the insurance plans into six groups. For adults (aged 18 or over), we had three plans—free, family pay, and individual deductible. The family pay plan consists of all plans with a medical coinsurance rate of 25, 50, or 95 percent. For

[11]A count of visits with positive charges.

[12]Expenditures include out-of-pocket payments and payments by the insurance carrier.

[13]As noted above, there is a small amount of differential underfiling. In light of the difficulty of ascribing this to particular individuals and its modest levels, we have not corrected for it in this analysis.

children, we had the same three plan groupings. Analysis of a subsample of HIE data (Newhouse et al., 1981, 1982) indicated that there was no appreciable or significant difference in inpatient use among the family pay plans, but that there was a significant and appreciable difference in the response of children and adults to cost sharing. This lack of detectable difference in inpatient use is to be expected, because the majority of those with any inpatient use on the family pay plans exceeded the MDE amount, irrespective of the plan; hence to most of those on the family pay plans, the cost of a hospitalization was $1,000 (less for the poor).

Measures of Health Status

We used four measures of health status to increase the precision of our estimates of the consumption of medical services: (1) general health perceptions;[14] (2) physical or role limitations;[15] (3) chronic disease status;[16] and (4) mental health status.[17] Each of these

[14]The General Health Index (GHINDX) is a continuous score (0–100) based on 22 questionnaire items for individuals aged 14 and over, and 7 items for children (aged under 14) measuring perceptions of health at the present, in the past, and in the future; the items also measure resistance to illness and health worry. GHINDX refers to health in general and does not specify a particular component of health.

The construct is a subjective assessment of personal health status. The reliability and validity of GHINDX have been extensively studied and documented (Ware, 1976; Davies and Ware, 1981; and Eisen et al., 1980). For example, the impact of chronic diseases, everything equal, is equivalent to 5.6 GHINDX points for hypertension, and 10 GHINDX points for chronic obstructive pulmonary disease or diabetes (Brook et al., 1983). The death rate in the study was 25/1,000 for those with GHINDX under 63, 6/1,000 for those with GHINDX from 63 to 76 and 1/1,000 for those with GHINDX from 76 to 100.

[15]The physical or role limitations measure is scored dichotomously (PHYSLM: 1 = limited, 0 otherwise) to indicate the presence of one or more limitations resulting from poor health. It is based on 12 questionnaire items for adults and five items for children measuring four categories of limitations: self-care (eating, bathing, dressing); mobility (confined, or able to use public or private transportation); physical activity (walking, bending, lifting, stooping, climbing stairs, running); and usual role activities (work, home, school). The reliability and validity of these measures have been studied and documented by Stewart et al. (1977, 1978, 1981a, 1981b) and Eisen et al. (1980).

[16]The disease measure is a simple count of the number of diseases or health problems (out of a possible 26) for individuals aged 14 or older (Manning et al., 1982). The disease list includes kidney disease and urinary tract infections, eye problems, bronchitis, hay fever, gum problems, joint problems, diabetes, acne, anemia, heart problems, stomach problems, varicose veins, hemorrhoids, hearing problems, high blood pressure, hyperthyroidism, and ten other diseases or problems.

[17]The Mental Health Inventory (MHI) for adults is a continuous score (0–100) based on 38 questionnaire items measuring both psychological distress and psychological well-being, as reflected in anxiety, depression, behavioral and emotional control, general positive affect and interpersonal ties. The reliability and validity of this measure has been studied and documented by Veit and Ware (1983); Ware et al. (1979, 1980b); and Williams et al. (1981). A similar construct has been developed for children aged 5 to 13, based on 12 questionnaire items (Eisen et al., 1980).

measures is based on the self-administered Medical History Questionnaire for individuals 14 years or older. Measures for children are based on questionnaires filled out by parents. All of the health status data used in this report were collected at the beginning of the study; a summary description of each is presented below.

Other Covariates

In addition to variables for insurance plan and health status, we also included covariates for age, sex, race, family income, family size, and site. With the exception of family size and income, the data were collected before or at enrollment in the study. The value for family size varies by year. Family income data are from 1975 in Dayton, 1978 for the three-year group in South Carolina, and 1976 for all other participants.[18] Health status measures are described more fully in Brook et al. (1983, 1984), Valdez et al. (1985), and Valdez (1986).

Although we have not tested for all possible interactions among covariates, we did examine some that are important for policy purposes (e.g., income and plan). As a result, we have included interactions between being a child and plan in the inpatient and outpatient use equations (see below), between plan and income in the probabilities of any use of medical care and of any inpatient use (see below), and between sex and age in all equations. The remaining interactions were neither significant nor appreciable and have been omitted.

UNIT OF ANALYSIS

The unit of analysis is a person-year. We use the year as the time frame for ease of interpretation and because the upper limit on out-of-pocket expenses is an annual limit. We use the person as the unit of observation because most major determinants of the use of services are individual (e.g., age, sex, and health status) rather than family (e.g., insurance coverage and family income).

---

[18]1975 was the first year of participation for the Dayton participants, the South Carolina three-year group began participation beginning in late 1978 (about a quarter participated for two months and another quarter for one month), and the remainder of the sample enrolled in 1976 or early 1977; most of the enrollment was in the latter half of 1976. We used these data because we believed the income measure was more reliable than the data on preexperimental income. The data we used were collected on forms keyed to income tax returns, whereas data on preexperimental income were responses to a personal interview.

III. STATISTICAL METHODS

In addition to estimates based on analysis of variance (ANOVA), we present more robust estimates based on a four-equation model developed by Duan et al. (1982, 1983, 1984). This model gains over ANOVA (and analysis of covariance, ANOCOVA) by exploiting three characteristics of the distribution of medical expenses. First, a large proportion of the participants use no medical services during the year. Second, the distribution of expenses among users is highly skewed. Third, the distribution of medical expenses is different for individuals with only outpatient use than for individuals with inpatient use.

Because of these three characteristics, ANOVA (and ANOCOVA) yields imprecise though consistent estimates of the effects of health insurance, health status, and socioeconomic status on the use of medical services, even for a sample size on the order of 21,000 (not all independent) observations. As Duan et al. (1982, 1983) and we, in Appendix C, show, a four-equation model that exploits the characteristics of the medical expense distribution yields consistent estimates with lower mean square error than ANOVA.

THE FOUR-EQUATION MODEL

We partition the participants into three groups: nonusers, users of only outpatient services, and users of any inpatient services. We examine the expenses of the last two groups of users separately.

The first equation of the model is a probit equation for the probability that a person will receive any medical service during the year—from either inpatient or outpatient sources. Thus, this equation separates users from nonusers and addresses the first characteristic described above—a large proportion of the population does not use medical services during the year. The second equation is a probit equation for the conditional probability that a user will have at least one inpatient stay, given that he has some medical use. This equation separates the two user groups and thus addresses the third characteristic noted above—different distributions of medical expenses for inpatient and outpatient users.

The third equation is a linear regression for the logarithm of total annual medical expenses of the outpatient-only users. The fourth equation is a linear regression for the logarithm of total annual medical expenses for the users of any inpatient service. This last equation

includes both outpatient and inpatient expenses for users of any inpatient services.[1]

More formally, the specifications of the four equations are as follows: The first equation is a probit equation for the dichotomous event of zero versus positive medical expense:

$$I_{1i} = x_i\beta_1 + \epsilon_{1i},$$

$$(\epsilon_{1i} \mid x_i) \sim N(0, 1),$$

where medical expense is positive if $I_{1i} > 0$, and 0 otherwise.

The second equation is a probit equation for having zero versus positive inpatient expense, given that the person is a positive user of medical services:

$$I_{2i} = x_i\beta_2 + \epsilon_{2i},$$

$$(\epsilon_{2i} \mid I_{1i} > 0, x_i) \sim N(0, 1),$$

where inpatient expenses are positive if $I_{1i} > 0$ and $I_{2i} > 0$, and zero if $I_{1i} > 0$ and $I_{2i} \leq 0$; the equation is defined only for $I_{1i} > 0$. There are only outpatient expenses if $I_{1i} > 0$ and $I_{2i} \leq 0$, and 0 otherwise.

The third equation is a linear model on the log scale for positive medical expenses if only outpatient services are used:

$$ln(\text{MEDICAL}\$_i \mid I_{1i} > 0 \text{ and } I_{2i} \leq 0) = x_i\beta_3 + \epsilon_{3i},$$

where

$$E(\epsilon_{3i} \mid x_i, I_{1i} > 0, I_{2i} \leq 0) = 0.$$

The fourth equation is also a linear model on the log scale for positive medical expenses if any inpatient services are used:[2]

$$ln(\text{MEDICAL}\$_i \mid I_{1i} > 0 \text{ and } I_{2i} > 0) = x_i\beta_4 + \epsilon_{4i},$$

---

[1]Grouping expenses by person rather than the more natural all-inpatient and all-outpatient expenditure eliminates the need to account for across equation correlation in estimating standard errors of total expenditure.

[2]Note that this equation includes *both* inpatient and outpatient medical expenses for any-inpatient users.

where

$$E(\epsilon_{4i} \mid x_i, I_{1i} > 0, I_{2i} > 0) = 0.$$

For the last two equations, the errors are assumed to be identically distributed but not necessarily normally distributed.

The likelihood function for this model is multiplicatively separable because of the way the conditional densities are calculated. (The separability does *not* depend on any assumption of independence among errors in the four equations. In fact, the errors may be correlated; see Duan et al., 1984.) Separability implies that estimating the four equations by maximum likelihood *separately* provides the global full-information maximum-likelihood estimates (Duan et al., 1983, 1984). We therefore estimate the four equations separately.

The logarithmic transformation of annual expenses practically eliminates the undesirable skewness in the distribution of expenses among users—the second characteristic noted above. In particular, the logarithmic transformation yields nearly symmetric and roughly normal error distributions, for which the least squares estimate is efficient. Further details are available in Duan et al. (1982, 1983) and Appendix C.

Although our use of the four-equation model is motivated by our desire to have the stochastic term approximate the normal assumption as closely as possible (to obtain robust estimates), the error distributions for the two levels of expense equations still deviate from the normal assumption. As a result, if we were to use the normal theory retransformation from the logarithmic scale to the raw dollar scale,

$$(exp(\sigma^2/2)),$$

the predictions would be inconsistent. Instead we use a nonparametric estimate of the retransformation factors, the smearing estimate, developed by Duan (1983), which is the sample average of the exponentiated least squares residuals:

$$\hat{\varphi}_j = \Sigma_i exp(\hat{\epsilon}_{ij})/n_j \tag{3.1}$$

where n_j = sample size for equation j,
$\epsilon_{ij} = \ln(y_{ij}) - x_i\hat{\beta}_j$
$\hat{\beta}_j$ = OLS estimate of β_j.

The smearing estimate is weakly consistent (asymptotically unbiased) for the retransformation factor if the error distribution does not depend on the characteristic x_i.[3]

A consistent estimate of the expected medical expense for medical services based on the four-equation model is given by

$$\text{E(Medical Expenditure}_i) = \hat{p}_i[(1 - \hat{\pi}_i)\exp(x\hat{\beta}_3)\hat{\varphi}_3 + \hat{\pi}_i \exp(x_i\hat{\beta}_4)\hat{\varphi}_4] \quad (3.2)$$

where $\hat{p}_i = \Phi(x_i\hat{\beta}_1)$ = estimated probability of any medical use,

$\hat{\pi}_i = \Phi(x_i\hat{\beta}_2)$ = estimated conditional probability for a medical user to have any inpatient use,

$\exp(x_i\hat{\beta}_3)\hat{\varphi}_3$ = estimate of the conditional expense for medical services if outpatient only,

$\exp(x_i\hat{\beta}_4)\hat{\varphi}_4$ = estimate of the conditional expense for medical services if any inpatient,

$\hat{\varphi}_3, \hat{\varphi}_4$ = estimated retransformation ("smearing") factor of the error terms for level of outpatient only and any inpatient expenditure equations.

Our estimates of predicted expenditure presented below are based on Eq. (3.2). We used Eq. (3.2) to predict medical expenditure for each person we enrolled, alternatively placing that person on each plan (by successively turning on plan dummy variables). We then averaged within plans over each predicted value to obtain a mean value for each plan. Standard errors of the predicted values are obtained by the delta method (Duan et al., 1982, pp. 40, 48). The regression equations underlying our predicted values are presented in Appendix A.

---

[3]Moreover, when the normal assumption does hold, the smearing factor has high efficiency (90 percent or more) relative to the normal retransformation for a wide range of parameter values, including those in this analysis (see Duan, 1983, Section 5, and Mehran, 1973). In the results presented below, the smearing factors for the log level of expense for outpatient-only users are estimated separately by plan and year to allow for heteroscedasticity. For the log level of expenses for users of any inpatient services, the smearing factor is a constant. See Duan et al. (1982, 1983) and Appendix C for a comparison of normal theory and nonparametric retransformations.

CORRELATION IN THE ERROR TERMS

Although we have over 20,000 observations, we do not have the same number of *independent* observations, because of substantial positive correlations in the error terms among family members and over time among observations on the same person. These correlations exist in all four equations. Failure to account for them in the analysis would yield inefficient estimates of the coefficients and statistically inconsistent estimates of the standard errors. In the results presented below we have corrected the inference statistics (t, F, and χ^2) for this positive correlation using a nonparametric approach.[4]

SELECTION MODELS

The econometric literature provides an additional class of models for continuous but limited dependent variables such as medical expenditure. These models include the Tobit model (Tobin, 1958), the Adjusted Tobit model (van de Ven and van Praag, 1981a, 1981b), and sample selection models (Maddala, 1983). Like our four-equation model, these are multi-equation models, with an equation (often a probit) for whether there is a positive amount, and another equation for the level of the positive amount. These models differ from ours in that they explicitly model the correlation between the probability of any use and the level of use. Although they may appear to be more general, in fact for this problem they are not (Duan et al., 1984). In particular, the four-equation model just described is not nested within the sample selection model.

Appendix C provides a fuller discussion of these models and, using a split-sample validation, shows that the four-equation model has less bias than the sample selection model and is statistically indistinguishable on the basis of mean square error. In a separate Monte Carlo study, Manning et al. (1987b) show that models such as the four-equation model are more robust and are no worse than selection models when the data are truly generated by a selection model.

---

[4]The correction is similar to that for the random effects least-squares model, or equivalently the intracluster correlation model (Searle, 1971). The model is described in Brook et al. (1984), based on prior work by Huber (1967) on the variance of a robust regression.

IV. EMPIRICAL RESULTS AND MAIN EFFECTS OF INSURANCE PLAN

ANOVA ESTIMATES

The data from the HIE clearly show that the use of medical services responds to changes in the amount paid out of pocket. Table 4.1 provides the sample means and standard errors for several measures of use of services—the probability of being treated, visit and admission rates, and total expenses. The per capita expenses on the free plan (no out-of-pocket costs) are 45 percent higher than those on the plan with a 95 percent coinsurance rate, subject to an upper limit on out-of-pocket expenses. Spending rates on plans with an intermediate level of cost sharing lie between these two extremes. The right-most column shows that adjusting for the site imbalance in plan assignments (see Table 2.2) makes little difference.

Cost sharing affects the number of medical contacts but not the intensity of each of those contacts. In other words, the differences in expenditures across plans reflect real variation in the number of contacts rather than an increase in the intensity or charge per service.[1] For example, outpatient expenses on the free plan are 67 percent higher than those on the 95 percent plan, whereas outpatient visit rates to physicians and other health providers are 66 percent higher than those on the 95 percent plan. A similar pattern holds more weakly for inpatient care; for example, inpatient expenses are 30 percent higher on the free plan than on the 95 percent plan, whereas admission rates are 29 percent higher.

The largest decreases in the use of outpatient services occur between the free and 25 percent plans, with smaller but statistically significant differences between the 25 percent and other family pay plans ($\chi^2(2) = 9.48, p < 0.01$).

There are no significant differences among the family coinsurance (25, 50, and 95 percent) plans in the use of inpatient services. For the probability of any inpatient use, total admission rates, and inpatient expenses, the contrasts between the 25, 50, and 95 percent plans have p values greater than 0.50. As noted above, this lack of a significant difference is probably due to the effect of the upper limit on out-of-

[1]Keeler et al. (1982) found that cost sharing affected the number of episodes of treatment, rather than the size of the episode. They used data from the first three years of the Dayton site. Lohr et al. (1986) found a similar result for diagnosis-specific episodes.

Table 4.1

SAMPLE MEANS FOR ANNUAL USE OF MEDICAL SERVICES PER CAPITA

(Standard error in parentheses)

Plan	Face-to-Face Visits	Outpatient Expenses (1984 $)	Admissions	Inpatient Dollars (1984 $)	Prob. Any Medical (%)	Prob. Any Inpatient (%)	Total Expenses (1984 $)	Adjusted Total Expenses (1984 $)[a]
Free	4.55	340	0.128	409	86.8	10.3	749	750
	(0.168)	(10.9)	(0.0070)	(32.0)	(0.817)	(0.45)	(38.7)	(39)
Family pay								
25 percent	3.33	260	0.105	373	78.8	8.4	634	617
	(0.190)	(14.70)	(0.0090)	(43.1)	(1.38)	(0.61)	(52.8)	(49)
50 percent	3.03	224	0.092	450	77.2	7.2	674	573
	(0.221)	(16.8)	(0.0116)	(139)	(2.26)	(0.77)	(143.5)	(100)
95 percent	2.73	203	0.099	315	67.7	7.9	518	504
	(0.177)	(12.0)	(0.0078)	(36.7)	(1.76)	(0.55)	(44.8)	(47)
Individual deductible	3.02	235	0.115	373	72.3	9.6	608	630
	(0.171)	(11.9)	(0.0076)	(41.5)	(1.54)	(0.55)	(46.0)	(56)
Chi-squared (4)	68.8	85.3	11.7	4.1	144.7	19.5	15.9	17.0
P value for chi-squared (4)	<0.0001	<0.0001	0.02	n.s.	<0.0001	0.0006	0.003	0.002

NOTES: All standard errors are corrected for intertemporal and intrafamily correlation. Dollars are expressed in June 1984 dollars. Visits are face-to-face contacts with MD, DO, or other health providers; excludes visits for only radiology, anesthesiology, or pathology services. Visits and expenses exclude dental care and outpatient psychotherapy. n.s. = not significant.

[a]The figures in this column are adjusted for the imbalance of plans across sites as follows: The site-specific responses on each plan (simple means by site) are weighted by the fraction of the sample in each site and summed across sites. In the case of the 50 percent plan, which has no observations in Seattle, the weights are renormalized excluding Seattle.

pocket expenses. Seventy percent of people with inpatient care exceeded their upper limit. Hence, the out-of-pocket cost of a hospitalization was at most $1,000 (in current dollars) and did not vary much among the pay plans (other than the individual deductible).[2]

The individual deductible plan exhibits a somewhat different pattern from the other cost sharing plans. This plan has free inpatient care, but a 95 percent coinsurance rate (up to a $150 per person or $450 per family annual maximum) for outpatient services. Total expenditures on this plan are significantly less than the free plan ($t = -2.34, p < 0.02$). This overall response is the sum of a one-third reduction in outpatient expenses ($t = -6.67$) , and a less than one-tenth reduction in inpatient expenses ($t = -0.68$). Thus, this plan looks like a combination of the 50 or 95 percent plans for outpatient care and the free or 25 percent plan for inpatient care. The admission rate for the individual deductible plan lies roughly midway between the free plan and family coinsurance plan rates, suggesting a nontrivial crossprice elasticity between inpatient and outpatient services.

FOUR-EQUATION ESTIMATES

Because sample means are quite sensitive to the presence of catastrophic cases, we used the four-equation model to provide more robust estimates of the plan responses.[3] The use of covariates in these equations further enhances precision and removes the relatively minor imbalances across plan, including the site imbalance. Table 4.2 presents estimates from this model of plan response for the probability of any use of medical services, the unconditional probability of any inpatient use, and total medical expenses. Figure 4.1 displays the expenditure results.

Mean predicted expenditure in the free care plan is 46 percent higher than in the 95 percent plan ($p < 0.001$), almost exactly the difference found in the sample means.[4] Like the sample means, these

[2]This is a good example of the difference between the response to a marginal price or coinsurance and the response to plan.

[3]For example, the ANOVA estimates of the response to cost sharing for total expenses (not adjusted for site) show a statistically insignificant reversal between the 50 percent and 25 percent plans. Although such a reversal is compatible with theory (because of the MDE) the reversal is almost certainly due to chance. One participant on this plan had a very expensive hospitalization (total medical expenses of $148,000 in one year); that single observation, which was the largest observation in the sample, adds $106 dollars to the 50 percent plan mean (16 percent of that plan's mean).

[4]It may seem that this is a trivial result that follows from the orthogonality of plan and covariates. Such is not the case because of the nonlinear transformations in the four-part model. Using the logarithm of expenditure plus $5, for example, as a depen-

Table 4.2

VARIOUS MEASURES OF PREDICTED MEAN ANNUAL USE OF MEDICAL SERVICES, BY PLAN

(Standard error in parentheses)

Plan	Likelihood of Any Use (%)	Likelihood of One or More Admissions (%)	Medical Expenses (1984 $)
Free	86.7	10.37	777
	(0.67)	(0.420)	(32.8)
Family pay			
25 percent	78.8	8.83	630
	(0.99)	(0.379)	(29.0)
50 percent	74.3	8.31	583
	(1.86)	(0.400)	(32.6)
95 percent	68.0	7.75	534
	(1.48)	(0.354)	(27.4)
Individual deductible	72.6	9.52	623
	(1.14)	(0.529)	(34.6)

NOTES: Medical services exclude dental and outpatient psychotherapy. The predictions are for the enrollment population carried forward through each year of the study. The standard errors are corrected for intertemporal and intrafamily correlation. The t-statistics for the contrasts with the free plan are −6.69, −6.33, −11.57, and −10.69 for the last four rows of the first column, respectively; −2.74, −3.57, −4.80, and −1.28 for the last four rows of the second column, respectively; and −4.05, −4.91, −6.74, and −3.78 for the last four rows of the third column, respectively. These t-statistics are larger than those one would compute from the standard errors shown in the table because use of the standard errors ignores the positive covariance between the two predicted plan means from the shared X β term. The differences in expenses between the 25 and 50 percent plans are significant at the 5 percent level (t = 1.97), and between the 50 and 95 percent plans are significant at the 6 percent level (t = 1.93). The parameter estimates underlying these predictions are available in Appendix A.

more robust estimates also indicate that the largest response to plan occurs between free care and the 25 percent plan, with smaller decreases thereafter.

Not surprisingly, given the approximate orthogonality of plan and covariates, adding covariates does not change the estimated probability

dent variable instead of the four-part model would lead to a much larger estimate of plan response, one that would be biased upward. (See Duan et al., 1982, 1983.)

of any use of medical services—87 percent of the free plan participants are predicted to use any service during the course of the year, whereas only 68 percent of the 95 percent plan participants are. These differences in the likelihood of receiving any care account for over three-fifths of the overall response to cost sharing. Virtually all the remaining response is attributable to the effect of cost sharing on hospital admissions.

Cost sharing for outpatient services only (the individual deductible plan) produces a different pattern of utilization than cost sharing for all services. Outpatient-only cost sharing reduces expenditures relative to free care ($p < 0.0001$), largely by reducing the likelihood of any use. Outpatient-only cost sharing also reduces inpatient use, but by an insignificant amount ($p = 0.20$ for the probability of any inpatient use). This last result is the only important change from the previously published analysis of the first 40 percent of the data (Newhouse et al., 1981). In that analysis one could reject at the 5 percent level the hypothesis that the free plan and individual deductible plan means for inpatient use were the same. This difference may have occurred because inflation in the late 1970s reduced the real value of the deductible, which was kept fixed at $150 (i.e., in nominal dollars), or may have simply been due to chance.

USE BY SUBGROUPS

An important goal of the HIE was to study how the response to cost sharing varied across subgroups. These included differences in responses across income groups, differences between adults and children, differences between the sickly and healthy, as well as differences across time (e.g., any transitory surges in use as insurance changed), and differences across medical markets (e.g., urban versus rural).

Across Income Groups

Different aspects of the use of medical services exhibit different responses to income (Table 4.2).[5] In Table 4.3 we observe differences in use that are due to both income directly and the effects of variables correlated with income; i.e., these are not partial effects.

[5]Recall that the income measure comes from the first partial year of enrollment. The division into thirds is site-specific (e.g., the lowest third is the lowest third of each site's income distribution), because (1) expenses are not corrected for cross-sectional differences in prices, and (2) we did not want to confound income and site; the sites were chosen to represent a spectrum of medical market characteristics. See Appendix D, Table D.1, for the ANOVA estimates by plan by income group (as well as by other subgroups).

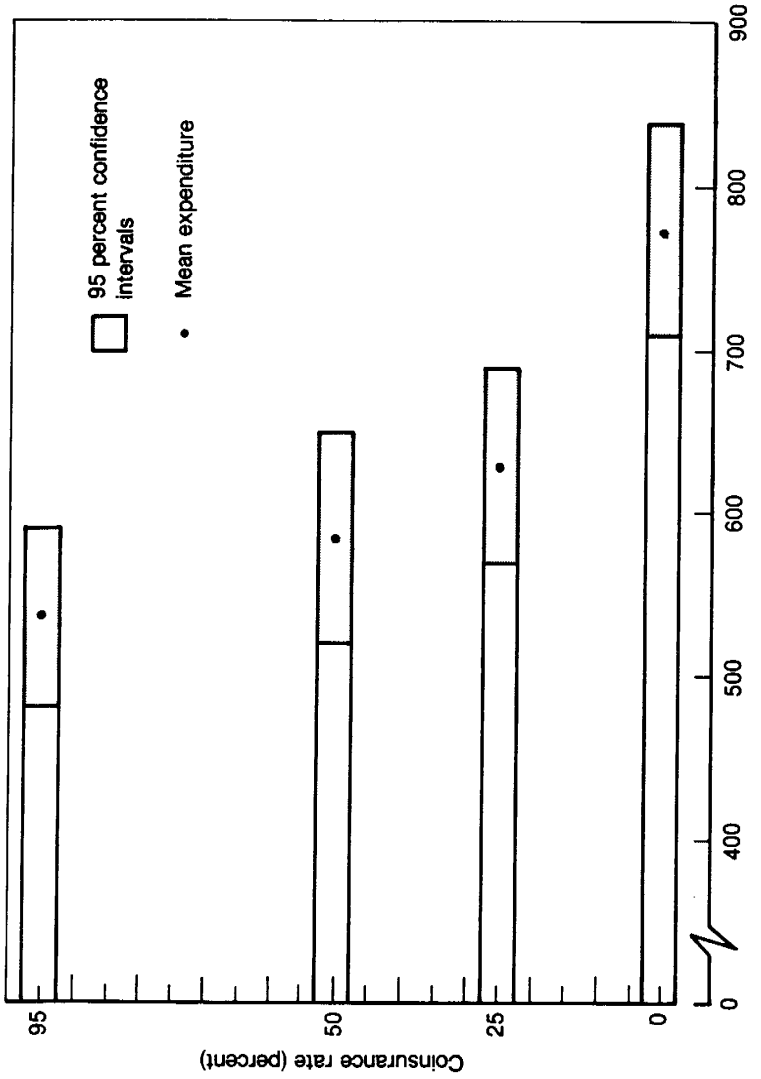

Fig. 4.1—Demand and 95 percent confidence intervals, by coinsurance rate

Table 4.3

VARIOUS MEASURES OF PREDICTED ANNUAL USE OF MEDICAL SERVICES, BY INCOME GROUP

Plan	Income Group			Significance Test	
	Lowest Third Mean	Middle Third Mean	Highest Third Mean	t on Contrast of Middle Third with Lowest Third[a]	t on Contrast of Highest Third with Lowest Third[a]
Likelihood of Any Use (%)					
Free	82.8	87.4	90.1	4.91	5.90
Family pay					
25 percent	71.8	80.1	84.8	5.45	6.28
50 percent	64.7	76.2	82.3	4.35	4.86
95 percent	61.7	68.9	73.8	3.96	4.64
Individual deductible	65.3	73.9	79.1	6.09	7.09
Likelihood of One or More Admissions (%)					
Free	10.63	10.14	10.35	−0.91	−0.35
Family pay					
25 percent	10.03	8.44	7.97	−2.95	−2.75
50 percent	9.08	8.06	7.77	−1.78	−1.66
95 percent	8.77	7.38	7.07	−2.79	−2.46
Individual deductible	9.26	9.44	9.88	0.31	0.68
Expenses (1984 $)					
Free	788	736	809	−1.78	0.53
Family pay					
25 percent	680	588	623	−3.17	−1.47
50 percent	610	550	590	−1.89	−0.49
95 percent	581	494	527	−3.09	−1.41
Individual deductible	609	594	670	−0.57	1.38

NOTES: Excludes dental and outpatient psychotherapy. Predictions for enrollment population carried forward for all years of the study.

[a]The t-statistics are corrected for intertemporal and intrafamily correlation. The statistics test the null hypothesis that the mean of middle (highest) third equals the mean of lowest third; e.g., the 4.91 figure implies we can reject at the 0.001 level the hypothesis that in the free plan the likelihoods of any use for the lowest and middle thirds of the income distribution are equal.

Within each of the five plans, the probability of any use of medical services increases with income, with larger increases for the family pay (25, 50, and 95 percent) and individual deductible plans than the free plan.[6] In contrast, the (unconditional) probability of any use of inpatient services declines with income for the family pay plans but is not significantly different across income groups for the two plans with free inpatient care (the free and individual deductible plans). Because of these two conflicting effects of income—positive on outpatient use but negative on inpatient use—the net result on total expenditure is a shallow U-shaped response.

Our estimate of the differences by income group within the family pay plans is influenced by the income-related upper limit on out-of-pocket expenses. The observed response is a combination of the direct response to income and the fact that families with lower incomes are more likely to exceed their (lower) limit and receive free care for part of the year.[7] If medical care is a normal good, then any positive direct effect of greater income would be reduced by the decreased likelihood of going over the limit. In the case of the positive effect of income on the probability of any use, the direct income effect is probably more important, and in the case of the negative effect on the probability of any inpatient use, the limit has relatively more influence.[8]

The individual deductible plan provides a cleaner test of the differences by income group of use of medical services, because the deductible in that plan is not income related. We observe an insignificant 10 percent increase in medical expenses between the bottom and top third of the income distribution. The effect of income is limited to an increased likelihood of using outpatient services, probably because inpatient services are free on this plan.

Thus far we have compared response among income groups rather than examining the partial effect of income. Although income has a statistically significant positive partial effect on use of service, the magnitude is small enough to be swamped by other factors correlated with income (see Appendix A, Tables A.2–A.4 and A.6).[9]

---

[6]Note that this is not a *ceteris paribus* statement, so there is no contradiction with standard theory, which would suggest no income effect in the free plan.

[7]See Appendix B for data on the proportion exceeding the upper limit on out-of-pocket expenses.

[8]Some may argue that income is endogenous with respect to inpatient expenditure. This may well be true but is not likely to account for our result because only a few months of data are "tainted."

[9]Income has a moderately significant (at $p < 0.10$) and positive partial effect on use in all but the inpatient expenditure equation; in the level of outpatient-only expenditures, however, the income coefficients are of mixed sign. The probabilities with which we can reject the null hypothesis that the income coefficients are zero are: $p < 0.001$ for

Across Age Groups

We observed about the same outpatient response to insurance plan for children (ages under 18) as for adults, but children are less plan-responsive for inpatient care (Table 4.4).[10] As we observed with a subset of these data (Newhouse et al., 1981, 1982; Leibowitz et al., 1985), we cannot reject the hypothesis that admission rates for children show no response to insurance coverage.[11] By contrast, adults have significantly lower use of inpatient services on the family pay plans than they do on the free plan.[12] For outpatient services, we observe a very similar pattern of plan responses for children and adults.

Other Subgroups

Health Status. Although health status was a strong predictor of expenditure levels, we observed no differential response to health insurance coverage between the healthy and the sickly. Using ANOVA methods, we found no significant plan interactions with health status, defined as site-specific thirds of the distribution of the general health index (GHINDX), our best summary health measure ($\chi^2(8) = 5.18$ for total expenditures, $p > 0.70$). We obtained similar results for the other measures of use of medical services, as well as with the four-equation model (see Table D.3 in Appendix D for the ANOVA estimates).

This null result is striking because of the upper limit feature. If anything, the presence of an upper limit on out-of-pocket expenses would lead to less plan response for the sickly; all other things equal, sicker individuals are more likely to exceed their upper limit and receive some free care—especially on the 95 percent plan, where care is

---

any use of medical services, $p < 0.10$ for the probability of any inpatient use given any medical use, $p < 0.001$ for the (log) level of outpatient-only use, and $p > 0.10$ for (log) level of medical expenditure if any inpatient use. The test statistics include plan income interactions and missing value replacement dummy variables.

[10]Recall that children were overrepresented in the study relative to the population of our sites. Hence, our estimates understate (modestly) the population responsiveness in our sites.

[11]χ^2 (4) = 5.19 using ANOVA estimates for the probability of any inpatient use, and χ^2 (4) = 5.36 for the admission rate. Another possible hypothesis is no differential plan response for children relative to adults. We can reject this hypothesis; the test statistics are χ^2 (4) = 16.49 for the probability of any inpatient use and χ^2 (4) = 14.08 for total admissions. Hence, it appears that children and adults respond differently and that children do not respond to cost sharing for inpatient care.

[12]χ^2 (3) = 24.22 for the probability of any inpatient use and 16.31 for the admission rate. By contrast, there are no significant differences among the family pay plans for adults. χ^2 (2) = 1.69 for expenditures, 0.73 for total admissions, and 1.39 for the probability of any inpatient use, again based on ANOVA (see Table D.2, Appendix D, for the ANOVA estimates).

Table 4.4

VARIOUS MEANS OF PREDICTED ANNUAL USE OF MEDICAL SERVICES, BY AGE GROUP AND PLAN

Plan	Likelihood of Any Use (%)	Likelihood of One or More Admissions (%)	Medical Expenses (1984 $)
Children			
Free	84.0	5.33	346
Family pay			
25 percent	75.1	4.98	287
50 percent	70.3	4.62	279
95 percent	63.5	4.23	236
Individual deductible	68.5	5.86	299
Adults			
Free	88.6	13.9	1080
Family pay			
25 percent	81.4	11.5	872
50 percent	77.1	10.9	797
95 percent	71.2	10.2	744
Individual deductible	75.6	12.1	852

NOTES: Excludes dental and outpatient psychotherapy services. The eight t-statistics for the contrasts between the free plan and the pay plans for the likelihood of any use all exceed 6. For one or more admissions, the t-statistics for children for contrasts with the free plan (rows 2 through 5 of the table) are 0.55, 1.13, 1.81, and −0.63, respectively, and for adults are 2.92, 3.64, 4.69, and 1.89, respectively (e.g., the t-statistic on the difference between 13.9 and 12.1 is 1.89). For medical expenses the t-statistics on contrasts with the free plan for children are 2.16, 2.20, 4.10, and 1.42, respectively, and for adults are 3.70, 4.80, 6.07, and 3.63, respectively.

free after gross expenditures of $1,050 or more. Furthermore, some might expect the sickly to be less responsive to insurance coverage than the healthy, on the supposition that their use of services is less discretionary. If, in fact, there is no interaction between plan and health status, one can infer that the opposite is true at the margin; i.e., at the margin the sickly may have more discretion.

Sites. The six sites in the HIE were selected to reflect a spectrum of city sizes, waiting times to appointment, and physician to population ratios (Newhouse, 1974).[13] Our concern was that the response to

[13]For example, city sizes in 1970 ranged from 34,000 (Georgetown County) to 1.2 million (Seattle), waiting times for nonemergent care in 1973–1974 ranged from 4.1 days (Seattle) to 25.0 days (Fitchburg), and physicians per capita in 1972 numbered from 30 per 100,000 (Fitchburg) to 59 per 100,000 (Seattle).

insurance coverage could vary according to the complexity of the medical market or to the excess demand in the medical delivery system. Yet we found no differences among the sites in the response to insurance coverage, $\chi^2(19) = 14.96$ ($p > 0.50$). For example, with longer delays to appointment, there might be a lower plan response because self-limiting illnesses (e.g., colds) are less likely to be treated by a physician. Yet we found no differences among the sites in the response to insurance coverage; using ANOVA methods on total expenditures, $\chi^2(19) = 14.96$ ($p > 0.50$); see Table D.4, Appendix D, for the ANOVA estimates.

Nevertheless, we did find differences among the sites in levels of use. Table 4.5 provides predictions of the use of services on the free plan for the six sites. We observe no significant differences among the three northern urban sites (Dayton, Seattle, and Fitchburg), but both of the rural sites (Franklin County, Massachusetts, and Georgetown County, South Carolina) and both of the southern sites have lower

Table 4.5

PREDICTED ANNUAL USE OF MEDICAL SERVICES, BY SITE

(Standard error in parentheses)

	Likelihood of Any Medical Use (%)		Likelihood of Any Inpatient Use (%)		Medical Expenses (1984 $)	
Site	Mean	t vs. Seattle	Mean	t vs. Seattle	Mean	t vs. Seattle
Seattle	89.2 (0.90)	—	10.5 (0.68)	—	863 (46.0)	—
Dayton	89.9 (0.88)	+0.65	10.6 (0.72)	+0.12	914 (53.1)	+0.94
Fitchburg	91.4 (1.04)	+1.88	11.0 (0.88)	+0.56	908 (65.6)	+0.68
Franklin County	92.0 (0.82)	+2.80	8.9 (0.66)	−1.91	644 (41.0)	−4.31
Charleston	77.2 (1.56)	−7.62	10.0 (0.72)	−0.46	693 (46.3)	−3.19
Georgetown County	79.1 (1.28)	−7.46	11.1 (0.69)	+0.67	637 (37.1)	−4.79

NOTE: Predictions for the standard population as if enrolled on the free plan for all years of the study.

expenditures than the northern urban sites. Part of the difference between the southern and northern sites is a lower propensity to obtain care in the South, even when care is free.

Interestingly, the site with the longest delay to appointment and lowest physician to population ratio (Fitchburg) had the second highest probability of any use, the second highest expenditures per enrollee, and the highest probability of any inpatient use. (The latter two phenomena may represent substitution of inpatient for outpatient care (McCombs, 1984), but the first is difficult to explain).[14]

Period of Enrollment. As noted above, we enrolled families for three or five years to see if the response to insurance changed over time and if the duration of enrollment mattered. The free plan might generate transitorily high demand; the 95 percent plan might generate postponement of demand at the end of the experiment (Arrow, 1975; Metcalf, 1973). Neither effect was found.[15] Nor did duration of enrollment matter to either the absolute level of spending or the responsiveness to plan.

To test for the presence of such plan-related surges, we compared the use of medical services during the middle years of the study with the first and last year of the study.[16] We found no significant differences in the response to plan across the years; using ANOVA techniques on medical expenses, $\chi^2\,(8) = 9.49, p > 0.30$, with similar results for the other measures of use of medical services; see Table D.5, Appendix D, for the ANOVA estimates.[17]

We did find some upward drift in medical expenses over time, after using the medical component of the Consumer Price Index to control for medical inflation. Table 4.6 presents the predicted mean expenses for the standard population for the free plan. We observed no such trend for median expense; Appendix A contains the order statistics and other summaries by plan and by year for total medical expenditures.

One explanation of this trend is the cost of new (and expensive) procedures and technology; clearly we cannot assume constant technology over the five years of the experiment. Another explanation we considered, but which is not likely to be the principal explanation, is an

---

[14]Length of waiting time to an appointment with a primary care physician is associated positively with the use of emergency rooms (O'Grady et al., 1985; Long et al., 1986).

[15]A transitory effect was found for dental services; see Manning et al. (1985a, 1986a) for details.

[16]For this test, the middle years of the study are Year 2 for the group enrolled for three years, and Years 2–4 for the group enrolled for five years.

[17]Keeler et al. (1982) and Manning et al. (1985a) did find that there was surge in the use of well care and dental care, respectively, at the beginning of the study.

Table 4.6

PREDICTED ANNUAL USE OF MEDICAL SERVICES, BY YEAR

(Standard error in parentheses)

	Likelihood of Any Medical Use (%)		Likelihood of Any Inpatient Use (%)		Medical Expenses (1984 $)	
Year	Mean	t vs. Year 1	Mean	t vs. Year 1	Mean	t vs. Year 1
1	87.5	—	10.1	—	711	—
	(0.71)	(0.52)	(34.1)			
2	85.9	−3.17	10.1	+0.08	760	+1.39
	(0.78)	(0.52)	(39.2)			
3	86.0	−2.94	10.8	+1.33	830	+3.25
	(0.77)	(0.56)	(41.2)			
4	87.6	+0.04	10.5	+0.50	820	+1.77
	(0.96)	(0.87)	(64.5)			
5	88.1	+0.60	10.6	+0.64	924	+3.32
	(0.97)	(0.85)	(68.7)			

NOTE: Predictions for the standard population as if enrolled on the free plan.

interaction of inflation and the upper limit, which was kept fixed in nominal terms. As medical prices rose between 1974 and 1982, individuals were more likely to exceed the upper limit on the pay plans; see Table B.3 in Appendix B. Thus, over time, more people received free care for part of the year on the pay plans. But such an explanation would generate a plan/year interaction, which we have not detected. Also this explanation would predict no drift in the free plan expenses, but we observe such a drift, albeit a smaller one than occurs on the pay plans.

Subexperiments. As described above, the HIE contained a number of subexperiments. None of the subexperiments to study methods effects had a measurable effect on expenditure. Those individuals who took the entry physical examination averaged $44 less a year in medical expenses (in 1984 dollars) than those without the entry examination.[18] The difference is statistically insignificant ($t = -1.31$). If we

[18]This result is for the standard population, assuming everyone had free medical care; the year is the first year of the study. We observed no significant plan interactions with the entry physical examination.

limit the analysis to the first year after the examination (to allow the exam to act as a substitute for planned well care), we find a positive but very insignificant effect on expenditures or visits (all t's less than 0.50) using ANOVA. There were no significant interactions between the entry examination and insurance plan, using ANOVA; the largest χ^2 statistic for the seven use measures was χ^2 (4) = 3.40 for the probability of any medical use.

Individuals who were initially assigned not to file health diaries had \$15 higher expenditures than those who filed weekly, but \$6 lower medical expenditures than those who filed biweekly.[19] The differences have t-statistics of 0.25 and −0.10, respectively. Using ANOVA, we could find no significant differences in the plan response between the no, weekly, and biweekly health report groups; the largest χ^2 statistic for the seven use measures was $\chi^2(8) = 11.36$ for inpatient dollars.

The annual expenditures for the three- and five-year groups differed by an insignificant \$4 per person per year ($t = 0.11$).[20] Using ANOVA on data from the first three years of the study, we could find no significant differences in the plan response between the three- and five-year groups; the largest χ^2 statistic for the seven use measures was $\chi^2(4) = 2.43$ for total admissions.

The additional lump-sum payment (unanticipated by participants) in the next to last year of the study did not have a statistically significant or appreciable effect on utilization. This subexperiment was limited to the 95 percent and individual deductible plan. The increase was a quarter or a third of their expected annual lump sum, up to a maximum of \$250. To estimate the effect of the additional payment, we subtracted the third to the last year's utilization from the second to the last (i.e., Year 1 from Year 2 or Year 3 from Year 4). This let each person act as his own control. We then compared the difference for the groups with and without the extra lump-sum payment. The group with the extra payment had 0.20 visits and 0.024 admissions more and \$6.50 lower expenditures than the group without the payment; the t's are +0.71, +0.86, and −0.05, respectively.

The results on the effects of these subexperiments suggest that our data collection methods did not have a significant or appreciable effect on our results.

---

[19]This result is for the standard population, assuming everyone had free medical care, for all years of the study. This standardization was required because different sites had different treatments.

[20]This result is for the standard population, assuming everyone had free medical care, for the first three years of the study. We standardize to the first three years of the study, to avoid confounding three- and five-year estimates with the drift in expenses mentioned above.

DENTAL AND HMO RESULTS

Here we only summarize these results, which are reported in greater detail elsewhere (Manning et al., 1984a, 1985a, 1985b, 1986a). Dental services do show greater responsiveness to plan in Year 1 than in subsequent years ($p < 0.001$). (This would be expected if dental services were more durable than other medical services, as is plausible.) Table 4.7 shows demand for dental services by plan in the first two years; the responsiveness by plan in Year 2, which is typical of the middle years, is of the same general magnitude as that for other medical services.

We also randomized a group of participants into an HMO, the Group Health Cooperative of Puget Sound in Seattle.[21] This group, whom we call the HMO experimentals, was given a plan of benefits

Table 4.7

USE OF DENTAL SERVICES BY DENTAL PLAN: SAMPLE MEANS

(Standard Errors in Parentheses)

	Year 1 of Dental Coverage			Year 2 of Dental Coverage		
Insurance Plan	Probability (%)	Visits per Enrollee	Expenses per Enrollee ($)	Probability (%)	Visits per Enrollee	Expenses per Enrollee ($)
Free	68.7	2.50	380	66.8	1.93	261
	(1.19)	(0.065)	(18.0)	(1.18)	(0.049)	(12.5)
25%	53.6	1.73	224	52.6	1.51	190
	(3.39)	(0.138)	(32.8)	(3.34)	(0.111)	(28.0)
50%	54.1	1.80	219	53.0	1.50	177
	(2.41)	(0.118)	(31.3)	(2.55)	(0.103)	(32.3)
95%	47.1	1.39	147	48.3	1.44	179
	(2.59)	(0.098)	(18.7)	(2.62)	(0.099)	(24.9)
Individual deductible	48.9	1.70	242	48.1	1.33	158
	(2.12)	(0.104)	(24.1)	(2.12)	(0.080)	(20.4)

NOTES: Expenses were converted to January 1984 dollars using the dental fee component of the Consumer Price Index. There has been no adjustment for regional differences in prices, or differences in population characteristics across plans and years. Standard errors are corrected for intrafamily and intertemporal correlation.

[21]An HMO is reimbursed a fixed amount per month, in return for which it agrees to provide medical care. Thus, unlike fee-for-service medicine, the approximate marginal

identical to the free fee-for-service plan. In addition, we enrolled a random sample of existing HMO enrollees, the HMO controls. Thus, a comparison of the experimentals and the free fee-for-service plan establishes the "pure" HMO effect on use; a comparison of the experimentals and controls establishes the extent, if any, of selection with respect to the HMO.[22]

Our results (Table 4.8) show no evidence of selection in the single HMO that we studied; those previously enrolled at the HMO (the controls) used services at approximately the same rate as those who were not previously enrolled (the experimentals). By contrast, the percentage of experimental plan participants with one or more hospital admissions was only two-thirds as great as the percentage on the free fee-for-service plan. Because outpatient use was approximately similar on the two plans, the expenditure difference between the HMO experimentals and free fee-for-service participants was somewhat narrower; expenditures per person among the HMO experimentals were only 72 percent of expenditures on the free fee-for-service plan.

These findings demonstrate that a markedly less hospital-intensive style of medicine than is commonly practiced in the fee-for-service system is technically feasible. Whether the technical style will be attractive to consumers, and if it is, whether a market of competing HMOs is economically feasible—or whether adverse selection problems will prove insurmountable (Rothschild and Stiglitz, 1976)—are still somewhat open questions, although the size and history of large HMOs such as Group Health Cooperative of Puget Sound suggest that the style is attractive to some consumers.

In projecting the effect of the growing HMO market share on hospital admissions and medical expenditure, one must keep in mind that the above comparisons have been made against the free care plan. Because virtually all private fee-for-service health insurance plans include some cost sharing, one should compare the reduction in hospital admissions at the HMO, some 35 percent, with the reduction caused by cost sharing, some 15 to 25 percent depending on plan. The values presented above, however, do represent the *ceteris paribus* HMO effect; if an HMO were to use cost sharing, its observed rates of use might be even lower.

Consumers contemplating enrollment in an HMO will weigh the cost savings against any effect of the reduction in services upon health

---

revenue from delivering additional services is zero (of course, there are market constraints on the HMO's product because it competes with fee-for-service medicine for patients).

[22]The fee-for-service sample in this comparison is from Seattle, to keep the population sampled the same between groups.

status and consumer satisfaction. Our findings on health status of the HMO are analogous to those in the free fee-for-service system; the mean person in the fee-for-service plan appeared to derive few or no benefits from the additional hospital services (Ware et al., 1986; Sloss et al., 1987). Those who are both in poor health and of low income who were in the HMO exhibited a higher rate of bed-days and serious symptoms (relative to those in the free fee-for-service plan). There is thus some suggestive evidence that special programs to facilitate access for Medicaid enrollees in HMOs may be worthwhile, but we caution that this result comes only from one HMO (albeit a well-established and well-regarded HMO) and that the precision with which we could measure results among the poor, sick group makes this result less than definitive, even in the case of this HMO.

Those who had self-selected the HMO (the controls) were on average as satisfied with their care as those in the fee-for-service system (Davies et al., 1986). Theory would suggest that the marginal person would be equally satisfied in both systems, and it is not surprising that we detected no difference for the average person. By contrast, the HMO experimentals were less satisfied overall with their care than those in the fee-for-service system, although on certain dimensions they were as satisfied or even more satisfied.

HEALTH STATUS OUTCOME RESULTS

These results also are only briefly summarized here; they have been reported in greater detail elsewhere (Brook et al., 1983, 1984; Valdez et al., 1985; Valdez, 1986; Bailit et al., 1985; Ware et al., 1986). For the person with mean characteristics, we can rule out clinically significant benefits from the additional services in the fee-for-service free plan relative to either the cost sharing plans or the HMO experimental group. For poor adults (the lowest 20 percent of the income distribution) who began the experiment with high blood pressure (specifically, who were in the upper 20 percent of the diastolic blood pressure distribution), there was a clinically significant reduction in blood pressure in the free fee-for-service plan compared with the plans with cost sharing. Epidemiologic data imply that the magnitude of this reduction would lower mortality about 10 percent each year among this group. (The sample size is much too small to test this prediction with actual mortality among the experimental population.) For poor adults who began the experiment with vision problems that were correctable with eyeglasses, there was a modest improvement in corrected vision. For individuals between the ages of 12 and 35 there was some improvement in

Table 4.8

ANNUAL USE OF MEDICAL SERVICES PER CAPITA, SEATTLE SAMPLE, BY HMO AND FFS STATUS[a]

(Standard error in parentheses)

Plan	Likelihood of Any Use (%)	One or More Admissions (%)	Imputed Expenditures ANOVA[b] (1983 dollars)	Imputed Expenditures with Age-Sex Covariates[b] (1983 dollars)	Person Years
HMO experimental	87.0	7.1	434	426	3687
	(1.0)	(0.50)	(28)	(23)	
HMO control	91.1	6.4	432	465	2596
	(0.8)	(0.55)	(34)	(47)	
Fee-for-service free	85.3	11.2	640	612	1221
	(1.6)	(1.17)	(81)	(66)	
t-statistic on free-experimental difference[c]	−0.88	3.24	2.44	2.69	
p-value for t-statistic, 2 tail	n.s.	0.0012	0.016	0.007	

[a]The sample includes participants while they remained in the Seattle area. The sample excludes children born into the study and excludes partial years except for deaths. For HMO controls and experimentals, the data include both in-plan and out-of-plan use. The standard errors are corrected for intertemporal and intrafamily correlation. The numbers differ slightly from those in Manning et al. (1984a), because of minor corrections in the data, as well as the use of a less precise, but more robust method of calculating standard errors, namely, the method described in the previous section. The method is the same as that described in Table 4.2.

[b]See Manning et al. (1984a) for details of imputation method.

[c]Testing null hypothesis of no difference between HMO experimental and free fee-for-service plan.

oral health; principally, caries (decayed teeth) were more likely to be filled, but there was also a modest improvement in the health of the gums.

The specific gains in health just described were all for relatively prevalent chronic problems (of course, we had difficulty detecting effects for rare problems) that are relatively inexpensive to diagnose and remedy. One can infer that programs targeted at these problems would be much more cost effective in achieving these gains in health than free care for all services. For example, more than half the benefit of free care for high blood pressure (and presumably for risk of dying) was available from a one-time screening examination, whose cost is a small fraction of free care for all services (Keeler et al., 1985).

V. CONCLUSIONS

OUR ESTIMATES OF DEMAND COMPARED WITH THOSE IN THE LITERATURE

Our results leave little doubt that demand elasticities for medical care are nonzero and indeed that the response to cost sharing is nontrivial. How do our estimates compare with those in the nonexperimental literature?

This question is difficult to answer, because most prior empirical work has parameterized cost sharing as a constant coinsurance rate (e.g., Feldstein, 1971, 1977) or has examined particular changes in insurance plans (e.g., an imposition of a $3 per visit copayment: Scitovsky and Snyder, 1972; Phelps and Newhouse, 1972; Scitovsky and McCall, 1977). By contrast, experimental policies were from a two-parameter family (coinsurance rate and maximum dollar expenditure). We make no apologies for this intentional noncomparability; a constant coinsurance rate, although convenient for obtaining comparative statics results, is not what theory suggests would be an optimal insurance policy assuming risk aversion (Arrow, 1963, 1971, 1973). Indeed, an optimal policy would almost certainly contain a stop-loss feature, exactly as the experimental plans did.[1]

One could, of course, attempt to estimate the functional response of demand to variation in the two parameters; one can view the values presented above as selected points in the response surface generated by varying coinsurance at given maximum dollar expenditure levels. To compare our results with those in the literature, however, we must extrapolate to another part of the response surface, namely, the response to coinsurance variation when there is no maximum dollar expenditure. Although any such extrapolation is hazardous (and of little relevance given the considerable departure from optimality of such an insurance policy), we have undertaken such an extrapolation rather than forgo entirely any comparison with the literature. Specifically, we have used three different methods to estimate a price elasticity comparable to those estimates in the literature:

[1]A stop-loss feature means there is a maximum out-of-pocket loss that the insured can sustain. In addition to its risk reduction properties, without a stop-loss feature, no worst case payment would have been possible and hence selection effects might have been introduced.

1. One can estimate a pure coinsurance elasticity by analyzing variation in the demand for episodes of care rather than annual expenditure per person (Keeler et al., 1982). The theory of demand suggests that individuals who have not yet exceeded the upper limit on out-of-pocket expenses, when making a marginal medical consumption decision, will discount the nominal price by the probability of exceeding the limit (because with that probability the true price is zero) (Keeler, et al., 1977a; Ellis, 1986).[2] We therefore examine demand for episodes of treatment by individuals who are more than $400 from their limit. This gives an approximation of the pure price effect if such people treat the true probability of exceeding their limit as nearly zero.[3] The estimation method controls for unobserved propensities to have episodes, as well as other observed covariates, by looking at experience before and after the MDE is exceeded; see Keeler et al. (1982) for a description of the methodology. We have computed arc elasticities for the 0–25 and 25–95 percent ranges of coinsurance; those elasticities are shown in Table 5.1.
2. A second estimate comes from using an indirect utility function and applying it to total expenditure in the 25–95 percent range. This estimate is very close to the first, -0.18 (Manning, 1988).
3. A third estimate comes from a similar calculation to those in the literature, i.e., it used average coinsurance rates (Table 5.2). The usual proof of an upward bias in the elasticity estimate from using the average coinsurance rate (Newhouse et al., 1980a) does not apply here because of the balance across plans. The amount of bias, if any, depends on two effects that work in opposite directions. For small expenditures the experimental plans will exhibit smaller expenditure than would a pure coinsurance rate plan of 16 or 31 percent (because the effective coinsurance rate is likely to be higher); for large expenditures exceeding the MDE the opposite will be true (because the marginal coinsurance rate will be zero, not positive). Which effect predominates is an empirical question the

[2]The specific result requires risk neutrality and separability of the utility function in health and money, but the qualitative result does not.

[3]Because there was no appreciable difference between demand for outpatient episodes when the MDE remaining was between $1 and $400 and when it was more than $400, this assumption seems reasonable for outpatient episodes. It may cause some bias in the estimated hospital elasticity; if the true MDE were, say $10,000 rather than $1,000, we might observe fewer hospitalizations.

Table 5.1

ARC ELASTICITIES FOR VARIOUS TYPES OF CARE CALCULATED FROM EPISODES

(Standard error in parentheses)[a]

Range of Nominal Coinsurance Variation	Outpatient Care[b]			Hospital Care[c]	Outpatient Care[c]	All
	Acute	Chronic	Well			
0–25	0.16	0.20	0.14	0.17	0.17	0.17
	(0.02)	(0.04)	(0.02)	(0.04)	(0.02)	(0.02)
25–95	0.32	0.23	0.43	0.14	0.31	0.22
	(0.05)	(0.07)	(0.05)	(0.10)	(0.04)	(0.06)

[a]The method of calculating standard errors is described in Keeler et al. (1988).

[b]Acute conditions are unforeseen and treatment opportunities are nondeferrable. Chronic episodes comprise foreseen and continuing expenditure; treatment is designed to ameliorate the consequences of the disease rather than cure. Flare-up of chronic conditions, which are unforeseen, we treat as acute. Well care episodes are medically deferrable without great loss and can occur when the patient is not considered sick.

[c]Estimate derived by weighting elasticities for various types of care by budget shares.

experimental data cannot resolve; nonetheless, this method yields values that are somewhat lower but still close to those of the other two methods. (The lower value suggests that the first bias predominates.)

These three methods suggest that price elasticities are in the −0.1 to −0.2 range, values that are consistent with those in the lower range of the nonexperimental literature, which vary from −0.1 to −2.1.

AN EXPLANATION OF THE SUSTAINED RISE IN MEDICAL EXPENDITURE

At first blush our estimates of demand response imply that the spread of health insurance can account for only a modest portion of the postwar rise in medical expenditure, contrary to the commonly held view described in the introduction. Between 1950 and 1984 real medi-

Table 5.2

ARC ELASTICITIES FOR VARIOUS TYPES OF CARE CALCULATED FROM AVERAGE COINSURANCE RATES

Range of Nominal Coinsurance Variation	Range of Average Coinsurance Variation	Outpatient Care	All Care
0–25	0–16	0.13	0.10
25–95	16–31	0.21	0.14

SOURCE: Calculated from data in Table 4.1 (outpatient) and 4.2 (total). For those who wish to calculate arc elasticities with the 50 percent plan, from the data in Tables 4.1 or 4.2, the average coinsurance rate in the 50 percent plan is 24 percent.

cal expenditure rose by a factor of seven,[4] but our estimates of price elasticity do not begin to imply this degree of increase. For example, although the average coinsurance rate is an imperfect measure of the generosity of insurance, it is a gross measure of how much insurance changed over the post-1950 period and therefore indicative of the role insurance might have played in this increase. Table 5.3 shows the average coinsurance rate by type of service (see Table 5.2 for comparable values from the 25, 50, and 95 percent plans). Although the figures by service are based on an arbitrary accounting convention, they suggest that the change in insurance in the postwar period was of roughly the same magnitude as the difference between the 95 percent coinsurance and free care plans.[5]

Because the free plan demand was only around 1.5 times that of the 95 percent plan, it appears that the change in insurance can explain only a small part, perhaps a tenth, of the factor of seven change in health expenditure.

Nor can changes in real income (around a factor of three during this period) directly account for much of the rise. Income elasticities estimated from the experimental data (the partial response, not the

[4]Nominal expenditure data from Levit et al. (1985), deflated by the GNP deflator.

[5]The accounting convention used by the Health Care Financing Administration allocates a common deductible to services in proportion to gross expenditure. We have followed the same convention in calculating comparable figures from the experimental data.

Table 5.3

CHANGE IN AVERAGE COINSURANCE RATE, 1950–1983, BY TYPE OF SERVICE

Year	Hospital	Physician	Other	Total
1950	0.30	0.83	0.86	0.66
1984	0.09	0.28	0.56	0.28

SOURCE: Levit et al. (1985).

one shown in Table 5.1) are at most 0.2—much too small to account for anything like a factor of seven change.[6]

Thus, we still must account for the bulk of the expenditure increase. The rather obvious "accounting" explanation of the expenditure increase is technological change; there are a host of new medical products and procedures today that did not exist in 1950. For example, those with kidney failure are now treated with renal dialysis or kidney transplantation; in 1950 these individuals died rather quickly. This merely pushes the puzzle back one stage, however; what role, if any, did insurance (and income growth) play in inducing the technological change? Unfortunately that question cannot be answered at all from our experimental data.[7]

Thus, if insurance is playing a role in inducing a welfare loss, the bulk of that loss must come about from inducing innovation for which unsubsidized consumers would not be willing to pay.[8] Given that most countries in the world have also experienced a long-term sustained increase in expenditure despite widely varying institutional arrangements, it is at least arguable that consumers would be willing to pay for much of the increase, but there clearly has been no pure market test (Newhouse, 1977, 1984).

[6]Real Gross National Product increased between 1950 and 1983 by a factor of 2.9. Even allowing for the usual downward bias from using measured income to estimate income elasticities, it is clear that changes in income can explain only a modest portion of the expenditure increase.

[7]Because most consumers have been insured for inpatient services throughout the relevant time period, it is an extremely difficult question to answer. Moreover, one cannot buy an insurance policy that will not cover new procedures. Hence, there is no straightforward test of willingness to pay for new technology. Although virtually all policies do not cover "experimental" procedures, once efficacy and "safety" are demonstrated, insurance plans tend to cover all procedures.

[8]The willingness-to-pay calculation should include any willingness to pay for others' care.

ON THE MAGNITUDE OF WELFARE LOSS FROM HEALTH INSURANCE

Setting aside the issue of possible welfare loss from induced technological change, one can estimate the welfare loss in the usual static framework. Under a number of strong assumptions (including that gross medical care prices are competitive and there are no externalities), our estimates imply a nontrivial welfare loss from first dollar health insurance coverage. An approximation to the loss from moving from a universal 95 percent plan (with a $1,000 MDE) to the free care plan is $37 to $60 billion, as against an expenditure around $200 billion on these services in 1984 by the under 65 population.[9] From the $37–$60 billion figure must be deducted some amount for the reduced risk in the free plan relative to the 95 percent plan. Usual values for risk aversion, however, would suggest the deduction is small in the presence of a $1,000 cap (Feldstein, 1973; Keeler et al., 1977b). We caution, however, that the $37–$60 billion figure ignores any welfare loss from induced technological change, and that figure could be potentially large relative to $37–$60 billion.[10]

---

[9]The $37 and $60 billion figures are calculated in the usual Harberger fashion by taking the $325 per capita difference in spending between the 95 percent and free plans from Tables 4.1 and 4.4 (Year 2 values) and adding $19 for mental health services (Wells et al., 1982, inflated by the change in the CPI Medical Services price index between 1977 and 1984). We then multiply by 207,000,000 the number of resident civilians under 65. This yields a figure of $71 billion. One then multiplies by 0.525 or 0.845. Both fractions are larger than the usual 0.5 because we do not start at an unsubsidized point. Our 95 percent $1,000 MDE plan had an average coinsurance rate of 0.31. An upper bound on the welfare loss comes from assuming that individuals valued the last dollar at 0.31. A lower bound on the welfare loss comes from assuming that the extra spending is all from individuals who valued the last dollar of spending at 0.95, the nominal coinsurance rate. The 0.525 figure equals 1 – 0.95/2, and the 0.845 figure equals 1 – 0.31/2.

The $200 billion figure can be estimated in two ways: (1) Data from Levit et al. (1985) show expenditure on personal health care services of $342 billion in 1984. Waldo and Lazenby (1984, Table 4.8) estimate that $120 billion of this is for the over 65 population, leaving $222 billion for those under 65. Some of this, however, is for noncovered services, such as nonprescription drugs, and some other part is for ineligible populations, such as the institutionalized. Adjusting for these noncomparabilities is necessarily somewhat imprecise but would probably leave a final figure around $200 billion. (2) Data from Tables 4.1 and 4.4 (Year 2 values) plus data on outpatient mental health spending from Wells et al. (1982) inflated to 1984 and scaled up by 207 million population imply an expenditure of $224 billion on the free care plan in our sites and $178 billion on the 25 percent coinsurance plans. Adjusting for price and usage levels in our sites relative to the nation is necessarily imprecise, but these two values probably bracket the true national figure.

[10]The induced technological change is clearly only a welfare loss if patent protection is at the level to induce the appropriate investment in new products in an unsubsidized market. If there is not enough patent protection, there is no necessary welfare loss from insurance.

Feldstein (1973) attempted to adjust for the willingness of consumers to pay for "higher quality care." There is no empirical way to do this, however, so the magnitude of

EXISTING INSURANCE COVERAGE OF VARIOUS MEDICAL SERVICES

One can find several economic reasons for the traditionally more generous coverage of inpatient services relative to outpatient services (Table 5.3). Loading charges (as a percentage of premium) are less, and the risk of a large loss is greater. For children, price elasticities for inpatient services are not measurably different from zero, and hence for them there is no measurable moral hazard.

This structure of more extensive insurance for inpatient services has been attacked as misguided, however (Roemer et al., 1975), on the grounds that lack of insurance for outpatient services deters ignorant individuals from seeking care at a time in their illness when they can be treated relatively cheaply. Others have also asserted that the more generous coverage of inpatient services leads physicians to hospitalize patients who could be treated on an outpatient basis, thereby minimizing private but increasing social expenditure.

Analysis of a natural (not randomized) experiment supported the claim that more complete coverage of outpatient expenditure reduced total expenditure (Roemer et al., 1975; Helms et al., 1978), but a prior controlled experimental study testing this hypothesis rejected it (Lewis and Keairnes, 1970; Hill and Veney, 1980). At issue is whether outpatient and inpatient services are substitutes or complements.

Our findings decisively reject the hypothesis that increased coverage of outpatient services, holding constant the coverage of inpatient services, will reduce expenditure. As Table 4.1 shows, the mean expenditure on the individual deductible plan (free inpatient, costly outpatient care) is 20 percent less than the mean on the free care plan (free inpatient, free outpatient care), and the difference is statistically significant ($p < 0.001$).[11] Disaggregation shows that the outpatient deductible not only reduces outpatient expenditure (Table 2.3) but, if anything,

---

the true welfare loss is highly problematic. Feldstein's method, although not explicit on the point, in effect ignores true technological change. He implicitly assumes that consumers in earlier years could have purchased "higher quality" medical care, but they chose not to because they faced a higher coinsurance rate or had lower incomes. (Alternatively, physician norms of care were "lower" because of the higher coinsurance rate and lower income.) As the renal dialysis example makes clear, however, consumers were simply unable to purchase some medical services in earlier years because they did not exist. In many cases their subsequent existence depended on fundamental scientific advance such as the discovery of DNA and would not have occurred without that advance, despite lower coinsurance or higher incomes. Whether consumers in the 1950s and early 1960s would have purchased such services if they had existed then can obviously not be answered from actual expenditure data. Feldstein's method also yields an upper bound for the same reason our $60 billion estimate is an upper bound.

[11]In the ANOVA results (Table 2.3) the estimated reduction is 19 percent and the t-statistic is 2.34 ($p < 0.02$, two-tailed test).

decreases hospital admissions for adults as well (Table 4.3). The (possibly) decreased admissions for adults suggests that outpatient and inpatient services are, if anything, complements, not substitutes.

In the interests of brevity we summarize four other implications for health insurance coverage:

- There appears to be little justification for the common practice of group insurance policies' treating emergency room services more generously than physician office visits, because emergency room services are as responsive to plan as physician office visits.[12]
- There is no support for the so-called offset hypothesis, namely, that more complete coverage of psychotherapy services will reduce total medical costs (or at least not increase them) (Follette and Cummings, 1967, 1968). The experimental data, however, are not very precise on this question.
- We have weak support for the lesser coverage of outpatient mental health care. Although the estimated plan response is substantially larger than that for outpatient medical care, the difference is statistically insignificant.[13]
- Well care services are about as price responsive as other medical services. Although there are other reasons for the common practice of not covering well care services as generously as other outpatient services (primarily little or no uncertainty), greater price responsiveness is not a reason.

---

[12]The discussion assumes that a presumed lower response to insurance is the reason for greater coverage of emergency room services. The alternative explanations, differential loading charges and asymmetric information, are not particularly plausible as an explanation of the better coverage of emergency room services. Asymmetric information is not very relevant to a single insurance plan offered in a group setting unless the service that is not well covered is deferrable and costly enough to motivate an employment change (which might apply to psychotherapy or certain costly dental services such as orthodontia). Routine office visits do not match this description. Moreover, asymmetric information may apply to both office and emergency room services. An individual may know that his use of office visits differs from average (whereas the insurer does not) but may also know that his likelihood of an accident differs from average, and the insurer may not.

[13]The estimated ratio of the free to 95 plan expenditures is 233 percent, which differs from a 169 percent estimate for medical outpatient care (Manning et al., 1986b).

WAS IT WORTH IT?

One question frequently raised about social experimentation is whether its benefits are worth its costs (e.g., Ashenfelter, 1986; Haveman, 1986). Because the question concerns the value of information, and because the links from this type of information to actual behavior are generally impossible to establish with any rigor, the question admits of no easy answer (save for the trivial case in which the experiment was so poorly designed or conducted that it produced no information). In other words, any attempt to justify the cost of an experiment is necessarily speculative.

Despite the circumstantial nature of the evidence, we believe that the benefits of this particular experiment greatly exceeded the (current dollar, undiscounted) costs of a little over $80 million ($136 million if put in 1984 dollars, and brought forward to 1984 using a 3 percent real discount rate). Between 1982 and 1984 there was a remarkable increase in initial cost sharing in the United States, at least for hospital services. For example, the number of major companies with first-dollar charges for hospital care rose from 30 percent to 63 percent in those two years, and the number of such firms with an annual deductible of $200 per person or more rose from 4 percent to 21 percent (Goldsmith, 1984). Although it is impossible to know how much of this change can be attributed to the experimental results, the initial findings of the experiment were published in December 1981 (Newhouse et al., 1981) and December 1983 (Brook et al., 1983) and given wide publicity in both the general and trade press. In certain instances a direct link between changes in cost sharing and the experimental results can be made.[14]

According to the experimental results, this increase in cost sharing should have decreased demand. Hospital days among the population under 65 decreased by 19.3 million days, or 12.8 percent (discharges decreased by 7.8 percent (American Hospital Association, 1985)). We estimate the cost saving from this reduced use to be around $7.25 billion. Physician visits among the population under 65 fell 27 million

[14]The average cost per hospital day in 1984 was $417. This uses the 1983 $368 figure from the American Hospital Association (1985) inflated by 13.3 percent, the change in per day inpatient costs from 1983 to 1984 (American Hospital Association, 1985). Friedman and Pauly (1981, 1983) have argued that the marginal cost/average cost ratio for hospital services is near one. Hence a *ceteris paribus* estimate of the savings from decreased use, assuming a marginal cost/average cost ratio of 0.9, is around $7.25 billion (19.33 million × 417 × 0.9). The American Hospital Association cost per day figure includes the over 65 population; however, cost per day is not very different for those over 65.

during these two years, but to be conservative we have not taken account of this change in estimating the cost savings.[15]

If all the changes in patient-days were attributable to the increased cost sharing, and if all the increase in cost sharing were due to the publication of experimental results, and if the benefits of the forgone use were negligible, as our results suggest, the experiment paid for itself in about a week (0.136/7.25)(912)![16] It is clear that these assumptions overstate the benefits of the experiment, yet it is equally clear that their assumptions can be greatly relaxed and still yield the result that the experiment was worth it. Moreover, we have ignored any benefits to countries other than the United States, and any benefits from the decrease in physician visits or changes in dental or mental health coverage or emergency room coverage. We have also ignored any benefits from the results of the HMO portion of the experiment, although HMOs' market share has been expanding rapidly from a period just before and after our first article describing the HMO results (Manning et al., 1984a). Finally, we have ignored the value of the public use files to future research efforts.[17]

Implicit in our conclusions is the assumption that one could not reduce uncertainty with nonexperimental data to the satisfaction of those making decisions about cost sharing. We believe that this is likely to be true, because of the wide range of nonexperimental estimates of insurance elasticity cited in the introduction, the difficulty of inferring health status effects from nonexperimental data, and the temporal proximity of the changes in cost sharing to the publication of the experimental results (many of the nonexperimental results had been in the literature for a decade, during which time cost sharing had, if anything, decreased). Thus, we think it highly plausible that the benefits of this endeavor were indeed worth its costs.

---

[15]In part, we do not account for such a change because the physician visit rate rose in 1985 to its 1982 value. Thus, the decrease from 1982 to 1984 could have been attributable to chance; alternatively the continued decrease in hospital care in 1985 (another 7.1 percent decrease in patient-days) may have led to a substitution of outpatient use. Data on physician visits are from the *National Health Interview Survey* (USDHHS, 1984, 1985).

[16]The negligible benefits assumption relies on the observation that cost sharing for hospital services was near zero in 1982 and that there were no measurable health benefits outside the dental area for the middle-class employees who would have been the dominant group for whom the cost sharing changed.

[17]The public use files can be ordered from Publications Department, The RAND Corporation, 1700 Main Street, Santa Monica, California, 90406–2138.

ON EXPERIMENTATION IN ECONOMICS

Econometric and economics texts often have a statement near the beginning that experimentation is not nearly as possible in economics as it is in the physical sciences. Perhaps the degree of difference is not as great as many think. Well designed and executed field and laboratory experiments are feasible and can add substantially to the body of knowledge (Heller, 1975; Plott, 1982).[18] We hope this example, from which public use tapes are now available, will encourage others to ask whether an experiment is practical or feasible when approaching empirical questions.

---

[18]For other views of field experiments see Hausman and Wise (1985) and Ferber and Hirsch (1978).

Appendix A

PARAMETER ESTIMATES AND SUMMARY STATISTICS ON THE DISTRIBUTION OF MEDICAL EXPENSES

Table A.1

INDEPENDENT VARIABLES

INDICATOR VARIABLES (if not otherwise defined, variables equal 1 if right-hand-side condition holds, otherwise zero)	
Insurance Plan[a]	
P25	Family coinsurance = 25 percent
P50	Family coinsurance = 50 percent
P95	Family coinsurance >= 95 percent
IDP	Individual deductible
PAY	P25 + P50 + P95
Other Subexperiments	
TERM3	Three-year enrollment[b]
TOOKPHYS	Entry physical examination[c]
NOHR	Did not file health diary[d]
WKLY	Filed health diary weekly[d]
Year and Site[e]	
DAY	Enrolled in Dayton
FIT	Enrolled in Fitchburg
FRA	Enrolled in Franklin Co., Mass.
CHA	Enrolled in Charleston
GEO	Enrolled in Georgetown Co., S.C.
YR2	Year 2
YR3	Year 3
YR4	Year 4
YR5	Year 5
Other Variables	
BLACK	Black
CHILD	Age < 18 on first day of year
AFDC	Aid to Families with Dependent Children
FEMALE	Female
AFAIRNM	Income data not missing
PHYSLM	Physically or role limited
GHINNM	General health measure not missing
INF	Health data from infant form (ages <= 4)[f]
PED	Health data from pediatric form
Continuous Variables	
MA1	(1 - FEMALE)*(AGE - 30)
MA2	(1 - FEMALE)*(LN(AGE) - LN(30))
FA1	FEMALE * (AGE - 30)
FA2	FEMALE * (AGE - 30)**2
FA3	FEMALE * (AGE - 30)**3

Table A.1 (continued)

XGHI	Predicted part of General Health Index[g]
XGHI2	Predicted part of (General Health Index squared)[g]
GHINDX	GHINNM * (Residual part of General Health Index)[g]
GHINDX2	GHINDX squared[g]
DISEA	Count of chronic diseases and health problems[h]
MHI	Mental Health Index[i]
XLINC	Predicted part of log family income[j]
LINC	AFAIRNM * (Residual part of log family income)[j]
LINC2	LINC squared[j]
XAFDC,AFDC	Received Aid to Families with Dependent Children[j]
LFAM	Log of family size[k]
Interactions	
P25INC	P25 * log of family income
P50INC	P50 * log of family income
P95INC	P95 * log of family income
IDPINC	IDP * log of family income
PAYINC	PAY * log of family income
CHP25	P25 * CHILD
CHP50	P50 * CHILD
CHP95	P95 * CHILD
ADIDP	IDP * (1 – CHILD)
ADPAY	PAY * (1 – CHILD)
CHIDP	IDP * CHILD
CHPAY	PAY * CHILD
T3YR2	TERM3 * YR2
T3YR3	TERM3 * YR3

[a]The free care plan (coinsurance rate = 0) is the omitted group.

[b]The five-year enrollment group is the omitted group.

[c]The no entry examination group is the omitted group.

[d]The group that filed health diaries biweekly is the omitted group.

[e]Seattle and the first year of the study are the omitted groups.

[f]The adult (ages 14+) version of the health questionnaire is the omitted group.

[g]Because of a missing General Health Index for all Dayton participants at enrollment, we imputed missing value replacements for all participants. For those with GHINNM = 1, we regressed the General Health Index (or its square) on socioeconomic and demographic variables, and preexperimental health as measured by Excellent/Good/Fair/Poor, Pain, and Worry, which were available in all sites. We used the predicted part for everyone and the residuals for the non-Dayton sites as our general health measure. A prefix of X indicates the predicted part; the absence of a prefix of X indicates the residual part (for those with complete data, 0 otherwise).

[h]Available only for those completing the adult form of the health questionnaire.

[i]Not available for the infant form (ages $\leq$ 4) of the health questionnaire.

[j]Because of missing income data, we imputed missing values for income and AFDC status by the same method used for the General Health Index in footnote g. Results in 1967 dollars.

[k]Count of the number of eligible family members present during the year.

Table A.2

PROBIT REGRESSION FOR ANY MEDICAL USE

Variable	Coefficient	SD (Coef.)	T
INTERCEP	1.6686E+00	1.298E+00	1.29
DAY	−8.9943E−02	1.232E−01	−0.73
FIT	2.2278E−01	7.702E−02	2.89
FRA	2.3024E−01	7.042E−02	3.27
CHA	−1.1470E−01	8.534E−02	−1.34
GEO	−5.6277E−02	8.062E−02	−0.70
P25	−1.2094E+00	5.683E−01	−2.13
P50	−2.1455E+00	8.321E−01	−2.58
P95	−1.0878E+00	5.922E−01	−1.84
IDP	−1.3591E+00	5.129E−01	−2.65
YR2	−8.8565E−02	2.771E−02	−3.20
YR3	−8.4453E−02	2.852E−02	−2.96
YR4	−6.7576E−02	4.709E−02	−1.43
YR5	−3.9220E−02	4.907E−02	−0.80
BLACK	−4.1825E−01	6.935E−02	−6.03
MA1	1.4593E−02	3.018E−03	4.84
MA2	−4.5263E−01	6.834E−02	−6.62
FEMALE	5.5474E−01	4.412E−02	12.57
FA1	−6.8812E−03	5.083E−03	−1.35
FA2	−1.3336E−03	2.653E−04	−5.03
FA3	6.7793E−05	2.106E−05	3.22
FA4	1.8315E−06	3.158E−07	5.80
FA5	−8.1798E−08	1.916E−08	−4.27
LFAM	−9.6001E−02	4.186E−02	−2.29
XLINC	3.0193E−01	5.410E−02	5.58
LINC	4.5628E−02	4.822E−02	0.95
AFAIRNM	1.8371E−01	6.570E−02	2.80
XGHI	−8.8564E−02	3.369E−02	−2.63
GHINNM	−4.6108E−02	9.854E−02	−0.47
GHINDX	−1.2152E−02	1.155E−02	−1.05
PHYSLM	2.2101E−01	5.781E−02	3.82
DISEA	2.2657E−02	3.117E−03	7.27
INF	−1.1787E−01	1.028E−01	−1.15
PED	−1.3954E−01	6.779E−02	−2.06
XXGHI2	5.4156E−04	2.367E−04	2.29
XGHI2	1.9097E−04	1.589E−04	1.20
GHINDX2	1.0616E−04	6.084E−05	1.75
TERM3	−3.4842E−02	4.726E−02	−0.74
TOOKPHYS	8.3052E−02	4.178E−02	1.99
NOHR	−4.2327E−02	6.453E−02	−0.66
WKLY	1.6997E−01	9.042E−02	1.88
P25INC	9.8601E−02	6.496E−02	1.52
P50INC	1.8684E−01	9.421E−02	1.98
P95INC	4.0739E−02	6.796E−02	0.60
IDPINC	8.9514E−02	5.973E−02	1.50

Table A.3

PROBIT REGRESSION FOR ANY INPATIENT USE—
GIVEN ANY MEDICAL USE

Variable	Coefficient	SD (Coef.)	T
INTERCEP	−4.5359E−01	4.128E−01	−1.10
DAY	8.2948E−02	1.103E−01	0.75
FIT	7.7392E−02	5.787E−02	1.34
FRA	−5.5344E−02	5.800E−02	−0.95
CHA	2.5761E−02	7.073E−02	0.36
GEO	6.2717E−02	6.848E−02	0.92
ADPAY	1.0016E+00	4.162E−01	2.41
ADIDP	5.0808E−02	4.752E−01	0.11
CHPAY	1.1006E+00	4.183E−01	2.63
CHIDP	2.0705E−01	4.805E−01	0.43
NOHR	3.4991E−03	5.464E−02	0.06
WKLY	−1.5569E−01	7.081E−02	−2.20
TERM3	−1.3247E−02	4.057E−02	−0.33
TOOKPHYS	−8.4536E−03	3.355E−02	−0.25
CHILD	−2.2495E−01	8.176E−02	−2.75
BLACK	4.3863E−02	6.336E−02	0.69
MA1	1.2811E−02	3.071E−03	4.17
MA2	−2.2558E−01	5.378E−02	−4.19
FA1	−1.9014E−02	4.935E−03	−3.85
FA2	−1.9020E−03	2.935E−04	−6.48
FA3	8.4682E−05	2.100E−05	4.03
FA4	1.8074E−06	3.564E−07	5.07
FA5	−7.5474E−08	1.987E−08	−3.80
FEMALE	4.5161E−01	4.850E−02	9.31
LINC	3.6082E−02	4.089E−02	0.88
LFAM	−3.9865E−02	3.635E−02	−1.10
XLINC	6.8142E−02	4.507E−02	1.51
GHINDX	−5.9229E−03	1.310E−03	−4.52
DISEA	−1.1342E−04	2.388E−03	−0.05
PHYSLM	1.6453E−01	4.807E−02	3.42
XGHI	−1.8474E−02	2.384E−03	−7.75
AFAIRNM	−9.5692E−02	5.563E−02	−1.72
GHINNM	−1.6275E−02	9.184E−02	−0.18
YR2	1.3142E−02	3.453E−02	0.38
YR3	5.6648E−02	3.486E−02	1.63
YR4	2.0928E−02	5.669E−02	0.37
YR5	2.3471E−02	5.287E−02	0.44
PAYINC	−1.2289E−01	4.715E−02	−2.61
IDPINC	−5.4188E−03	5.431E−02	−0.10

Table A.4

ORDINARY LEAST SQUARES FOR LOG MEDICAL EXPENSES—
IF OUTPATIENT EXPENSES ONLY

Variable	Coefficient	SD (Coef.)	T
INTERCEP	3.5957E+00	7.859E−01	4.57
DAY	−1.4081E−01	8.489E−02	−1.66
FIT	−8.9658E−02	5.346E−02	−1.68
FRA	−1.8637E−01	4.812E−02	−3.87
CHA	−1.8027E−01	5.927E−02	−3.04
GEO	−1.2592E−01	6.359E−02	−1.98
TERM3	6.0232E−02	4.172E−02	1.44
TOOKPHYS	4.0898E−02	2.940E−02	1.39
NOHR	−5.4108E−02	4.415E−02	−1.23
WKLY	3.0624E−04	5.919E−02	0.01
P25	−2.7876E−01	4.267E−02	−6.53
P50	−4.3917E−01	6.325E−02	−6.94
P95	−5.1655E−01	4.864E−02	−10.62
IDP	−3.2371E−01	4.773E−02	−6.78
YR2	1.1921E−02	4.041E−02	0.29
YR3	−6.8498E−02	4.244E−02	−1.61
YR4	1.1690E−02	4.051E−02	0.29
YR5	1.0673E−01	4.083E−02	2.61
BLACK	−2.4046E−01	5.736E−02	−4.19
LFAM	−2.3055E−01	3.747E−02	−6.15
XLINC	2.5482E−01	4.550E−02	5.60
LINC	3.5858E−02	2.936E−02	1.22
XAFDC	2.1750E−01	1.553E−01	1.40
AFDC	9.5526E−02	8.491E−02	1.13
AFAIRNM	−4.5663E−03	5.714E−02	−0.08
XGHI	−2.6836E−02	2.025E−02	−1.33
GHINNM	−1.4818E−01	6.784E−02	−2.18
GHINDX	−2.2222E−02	8.248E−03	−2.69
DISEA	1.3821E−02	2.051E−03	6.74
PHYSLM	1.3987E−01	4.133E−02	3.38
INF	−1.2443E−01	8.334E−02	−1.49
PED	−6.2863E−02	6.095E−02	−1.03
MA1	2.1724E−02	2.473E−03	8.79
MA2	−2.6668E−01	4.916E−02	−5.42
FEMALE	3.5118E−01	3.558E−02	9.87
FA1	5.4449E−03	3.817E−03	1.43
FA2	−9.9838E−04	2.214E−04	−4.51
FA3	4.7362E−05	1.592E−05	2.97
FA4	1.1147E−06	2.543E−07	4.38
FA5	−5.5444E−08	1.446E−08	−3.83
T3YR2	−8.0548E−02	4.867E−02	−1.66
T3YR3	1.4705E−01	5.032E−02	2.92
XXGHI2	5.7985E−05	1.468E−04	0.39
XGHI2	2.4635E−04	1.126E−04	2.19

Table A.4 (continued)

Variable	Coefficient	SD (Coef.)	T
CHILD	−4.4560E−03	6.181E−02	−0.07
CHP25	5.3261E−02	6.246E−02	0.85
CHP50	2.8772E−01	9.293E−02	3.10
CHPFD	1.5244E−01	6.954E−02	2.19
CHPDP	2.8422E−02	7.008E−02	0.41

NOTE: In 1967 dollars; add log (Medical CPI/100) to intercept to put into present dollars.

Table A.5

SMEARING FACTOR FOR OUTPATIENT EXPENSES ONLY

(Ordinary Least Squares)

Variable	Coefficient	SD (Coef.)	T
INTERCEP	1.4760E+00	4.247E−02	34.76
P25	1.2735E−01	7.814E−02	1.63
P50	1.3838E−01	1.335E−01	1.04
P95	2.2665E−01	1.084E−01	2.09
IDP	3.0484E−01	9.407E−02	3.24
YR2	1.2483E−01	5.952E−02	2.10
YR3	1.5945E−01	6.069E−02	2.63
YR4	3.7500E−02	8.162E−02	0.46
YR5	1.5015E−01	9.123E−02	1.65
P25YR2	−8.6331E−02	9.932E−02	−0.87
P50YR2	−9.9127E−02	1.538E−01	−0.64
P95YR2	1.3519E−01	1.321E−01	1.02
IDPYR2	−2.8266E−01	1.121E−01	−2.52
P25YR3	−1.2595E−01	1.036E−01	−1.22
P50YR3	−2.5311E−01	1.314E−01	−1.93
P95YR3	−1.3097E−01	1.456E−01	−0.90
IDPYR3	−6.0827E−02	1.238E−01	−0.49
P25YR4	5.8408E−03	1.743E−01	0.03
P50YR4	3.3043E−01	2.272E−01	1.45
P95YR4	−5.4672E−02	2.312E−01	−0.24
IDPYR4	−2.7122E−02	2.101E−01	−0.13
P25YR5	−2.3857E−01	1.485E−01	−1.61
P50YR5	−9.0410E−02	1.835E−01	−0.49
P95YR5	−2.4708E−02	2.098E−01	−0.12
IDPYR5	−2.5651E−01	2.102E−01	−1.22

NOTES: Interactions are plan x year. In 1967 dollars; multiply by Medical CPI/100 to put into present dollars.

Table A.6

ORDINARY LEAST SQUARES FOR LOG MEDICAL EXPENSES—IF ANY INPATIENT USE

Variable	Coefficient	SD (Coef.)	T
INTERCEP	6.8069E+00	4.145E−01	16.42
DAY	−7.9430E−03	1.455E−01	−0.05
FIT	1.8631E−01	7.058E−02	2.64
FRA	−5.7147E−02	7.973E−02	−0.72
CHA	−2.4039E−02	8.133E−02	−0.30
GEO	−2.6530E−01	7.947E−02	−3.34
ADPAY	−2.2435E−02	4.942E−02	−0.45
ADIDP	−9.8320E−02	5.800E−02	−1.70
CHPAY	−6.8029E−02	9.498E−02	−0.72
CHIDP	−1.0277E−01	1.072E−01	−0.96
NOHR	4.4623E−02	6.532E−02	0.68
WKLY	3.6569E−02	7.780E−02	0.47
TERM3	4.0170E−03	4.667E−02	0.09
TOOKPHYS	−1.1635E−01	4.042E−02	−2.88
CHILD	6.9015E−02	1.211E−01	0.57
BLACK	−5.3306E−02	6.547E−02	−0.81
MA1	4.0765E−03	4.690E−03	0.87
MA2	1.8517E−01	7.687E−02	2.41
FA1	2.3041E−02	5.544E−03	4.16
FA2	7.6302E−05	4.103E−04	0.19
FA3	−3.8807E−05	2.586E−05	−1.50
FA4	−1.8776E−07	5.486E−07	−0.34
FA5	3.4964E−08	2.656E−08	1.32
FEMALE	−1.0370E−02	6.701E−02	−0.15
LINC	−3.7398E−02	3.526E−02	−1.06
LFAM	−3.5739E−02	4.340E−02	−0.82
XLINC	9.1097E−02	4.323E−02	2.11
GHINDX	−1.5404E−04	1.521E−03	−0.10
DISEA	1.9583E−03	2.676E−03	0.73
PHYSLM	2.1274E−01	5.048E−02	4.21
XGHI	−9.5000E−03	2.649E−03	−3.59
AFAIRNM	−1.5085E−01	7.006E−02	−2.15
GHINNM	−7.0544E−02	1.295E−01	−0.54
YR2	8.5359E−02	4.775E−02	1.79
YR3	1.0632E−01	4.688E−02	2.27
YR4	1.7341E−01	8.118E−02	2.14
YR5	2.6077E−01	7.937E−02	3.29

NOTES: In 1967 dollars; add log (Medical CPI/100) to intercept to put into present dollars.

Table A.7

SMEARING FACTOR IF ANY INPATIENT USE

(Ordinary Least Squares)

Variable	Coefficient	SD(Coef.)	T
INTERCEP	1.4112E+00	4.891E−02	28.85

NOTES: In 1967 dollars. Multiply by Medical CPI/100 to put into present dollars.

Table A.8

STATISTICS FOR MEDICAL EXPENSES, BY PLAN

(Constant 1984 dollars)

Quantile	Free	25	50	95	Individual Deductible
0.10	0	0	0	0	0
0.20	37	0	0	0	0
0.30	84	40	33	0	18
0.40	139	79	72	33	50
0.50	201	128	123	70	98
0.60	289	187	178	122	166
0.70	419	278	249	204	273
0.80	665	467	397	358	499
0.90	1,570	1,160	909	946	1,370
0.95	3,290	2,740	2,030	2,410	2,730
0.99	9,090	9,240	6,750	8,030	6,860
Max	72,800	48,000	148,000	50,700	70,500
Mean	749	634	674	518	608
Standard deviation	2,390	2,350	5,210	2,054	2,466
Skewness	12.9	10.6	22.2	11.8	15.2
Kurtosis	270	152	555	198	321

Table A.9

STATISTICS FOR MEDICAL EXPENSES, BY YEAR

(Constant 1984 dollars)

Quantile	1	2	3	4	5
0.10	0	0	0	0	0
0.20	0	0	0	0	0
0.30	40	32	33	44	47
0.40	82	69	76	86	96
0.50	139	122	135	133	149
0.60	205	186	209	194	219
0.70	309	294	319	303	327
0.80	510	490	546	484	568
0.90	1,190	1,250	1,400	1,230	1,430
0.95	2,560	2,770	3,050	2,740	3,220
0.99	6,720	8,430	8,200	12,500	11,100
Max.	95,600	148,000	59,100	50,700	44,700
Mean	581	641	656	735	784
Standard deviation	2,410	3,030	2,240	2,890	2,886
Skewness	22.1	27.0	11.8	10.25	9.06
Kurtosis	704	1,110	211	135.7	100.0

Table A.10

PERCENTAGE OF TOTAL EXPENSES PAID BY TOP PERCENTAGE OF PEOPLE, BY PLAN

Top % of People	Free	25	50	95	Individual Deductible
90	100.0	100.0	100.0	100.0	100.0
80	99.8	100.0	100.0	100.0	100.0
70	99.0	99.7	99.8	100.0	100.0(–)
60	97.5	98.7	99.0	99.7	99.4
50	95.2	97.1	97.6	98.7	98.2
40	92.0	94.6	95.3	96.9	96.1
30	87.4	91.0	92.2	93.8	92.6
20	80.3	85.3	87.6	88.6	86.5
10	67.0	74.0	79.0	77.8	73.1
5	51.7	59.9	68.6	63.2	57.3
1	23.8	30.3	46.5	31.0	30.3

Table A.11

PERCENTAGE OF TOTAL EXPENSES PAID BY TOP PERCENTAGE OF PEOPLE, BY YEAR

Top % of People	Free	25	50	95	Individual Deductible
90	100.0	100.0	100.0	100.0	100.0
80	100.0	100.0	100.0	100.0	100.0
70	99.6	99.8	99.8	99.7	99.6
60	98.6	99.0	99.0	98.8	98.7
50	96.7	97.5	97.4	97.3	97.2
40	93.7	95.2	94.8	95.1	94.9
30	89.3	91.5	90.8	91.8	91.4
20	82.5	85.5	84.5	86.5	85.8
10	69.4	73.6	71.3	76.1	74.6
5	54.3	59.3	55.4	63.5	60.5
1	27.2	30.6	26.1	32.1	31.5

Appendix B

EXCEEDING THE UPPER LIMIT ON OUT-OF-POCKET EXPENSES

FAMILY COINSURANCE PLANS

We have analyzed the data for families not on the individual deductible plan to determine the percentage of families who exceeded the limit on out-of-pocket expenses during the HIE. Dayton Year 1 was excluded because outpatient psychiatric did not count toward the upper limit; all other years and sites are included.[1]

A fifth of families on the 25 and 50 percent plans exceeded the upper limit, whereas about a third of those on the 95 percent plan did; see Table B.1.

If a family was not pinned at the maximum upper limit of $1,000 ($750 in the 25 percent plan in some site-years), it was more likely to exceed the upper limit; see Table B.2. A regression equation on exceeding the upper limit has the following t-statistics: dummy for 50, −1.58; dummy for 95, 2.17; dummy for pinned, −7.89; dummy for 50 × pinned, 2.72; dummy for 95 × pinned, 1.99; the unpinned 25 percent plan is the omitted group. The Wald test for plan/pinned interactions is $\chi^2(2) = 8.66, p < 0.01$.

The proportion of families exceeding the upper limit tended to creep up through time as inflation worked against the fixed $1,000 (or $750) maximum; see Table B.3. The year is the calendar year of the first day

Table B.1

PERCENTAGE OF FAMILIES EXCEEDING UPPER LIMIT

Coinsurance Rate	Percentage	t vs. 25%
25 percent	20.8	—
50 percent	21.5	0.24
95 percent	35.0	6.46

[1]The plans with a medical coinsurance rate of 25 percent and a dental/mental coinsurance rate of 50 percent are grouped with the 25 plans.

Table B.2

PERCENTAGE OF FAMILIES EXCEEDING UPPER LIMIT

Medical Coinsurance Rate	Percentage
25 percent — below maximum	33.8
25 percent — pinned	11.2
50 percent — below maximum	25.9
50 percent — pinned	18.4
95 percent — below maximum	41.4
95 percent — pinned	27.3

of the accounting year. Years 1975 and 1976, and 1980+ were grouped because of small sample sizes. In a regression equation with main effects for plan and year and plan/year interactions, the year main

Table B.3

PERCENTAGE OF FAMILIES EXCEEDING MDE, BY YEAR

Coinsurance Rate and Year	Percentage Exceeding
25 percent	
1975/1976	14.8
1977	18.9
1978	22.6
1979	27.0
1980+	24.4
50 percent	
1975/1976	18.8
1977	22.2
1978	20.9
1979	24.1
1980+	25.5
95 percent	
1975/1976	25.8
1977	34.6
1978	35.7
1979	37.2
1980+	39.3

effects versus 1976 are all significant at 1 percent except for 1977 (t = 1.57). A test of significance on the plan-year interactions has them less significant than they would be at random ($\chi^2(8) = 2.58$).

We also tested whether "the poor" were more likely to exceed their upper limit. Families were grouped according to whether their upper limit was 5 percent of their income or less (about 40 percent of the families fell in this group); see Table B.4. In a regression with plan dummies, a dummy for 5 percent of income or less, and plan × 5 percent of income dummy interactions, the main effect of the income dummy was very significant (t = 5.69), and the interaction terms were marginally significant ($\chi^2(2) = 3.89$).

Table B.4

PERCENTAGE OF FAMILIES EXCEEDING MDE, BY INCOME

Coinsurance Rate and Income Group	Percentage Exceeding
25 percent — MDE < 5% of income	28.0
25 percent — MDE ≥ 5% of income	13.4
50 percent — MDE < 5% of income	36.7
50 percent — MDE ≥ 5% of income	14.6
95 percent — MDE < 5% of income	48.6
95 percent — MDE ≥ 5% of income	28.1

Table B.5

PERCENTAGE OF FAMILIES EXCEEDING MDE, BY INCOME GROUP

Coinsurance Rate and Income Group	Percentage Exceeding
25 percent	
Lower half	25.8
Upper half	15.4
50 percent	
Lower half	21.2
Upper half	21.8
95 percent	
Lower half	37.3
Upper half	32.7

Finally, we looked at the effect of being in the upper half of each site's income distribution; see Table B.5. The Wald test for the plan income interaction is $\chi^2(2) = 3.89$.

INDIVIDUAL DEDUCTIBLE

We also looked at individuals exceeding their deductible on the individual deductible plan. Dayton Year 1 was deleted because of differences in coverage. Over the course of the study, 45.1 percent (t = 25.32) exceeded this deductible annually. In contrast to the family coinsurance plans, there was no significant year effect ($\chi^2(4) = 5.52$). The year-to-year pattern was not consistently monotonic. Low-income people went over 35.6 versus 51.6 percent for high income (t = 4.86) using site-specific median income to classify by high and low income. If we create a dummy variable for having an upper limit/income ratio in excess of the median (2.8 percent), we find that those below the median exceed 56.0 percent of the time, compared with the 37.6 percent for those above (t = 5.46). Here, the MDE is defined as the maximum *family* out of pocket.

All of these inference statistics were corrected for intertemporal correlation.

FAMILY HEALTH EXPENSES

Finally, we examined the distribution of total family health care expenditures. These include outpatient dental and mental health expenses as well as all medical expenses. Table B.6 provides summary statistics by plan in nominal dollars, not adjusted or corrected for inflation; note that the dollar amounts in the text and in Appendix A are in June 1984 dollars. The sample excludes Year 1 for all sites.

If there were no uncertainty about illness, then we would expect that the upper limit would have little effect on very high users. At the margin, these individuals on the pay plan would have free care. They would differ from those on the free plan by only the income effect of paying the inframarginal cost sharing. Since this amount is at most $1,000 on the family pay plans, and $450 on the individual deductible, the income effects should be minimal.

The numbers on Table B.6 suggest that the effect of the plan persists well beyond the gross dollar amount corresponding to the upper limit on out-of-pocket expenses ($\leq$ $3,000 on the 25 percent, $\leq$ $2,000 on the 50 percent, $\leq$ $1,053 on the 95 percent, and $\leq$ $450 on the individual deductible plan).

Table B.6

STATISTICS FOR FAMILY HEALTH EXPENSES, YEARS 2–5

Quantile	Free	25%	50%	95%	Individual Deductible
0.10	99	35	16	0	0
0.20	275	134	79	35	67
0.30	470	245	165	109	160
0.40	702	380	305	185	300
0.50	1010	581	443	326	499
0.60	1430	802	591	519	835
0.70	1970	1190	919	881	1330
0.80	2840	1880	1610	1680	2020
0.90	4550	3780	3070	3220	3250
0.95	6830	5610	4940	5360	4950
0.99	12700	14000	9870	12700	12900
Max	37200	34300	72700	34900	43000
Mean	1894	1454	1290	1218	1425
Standard deviation	2790	2790	4190	2630	3150
Skewness	4.71	5.23	13.70	5.85	7.29
Kurtosis	36.71	39.21	225	50.77	71.58

The probability that total expenses exceed \$4,000 (nominal) is lower for each of the family coinsurance plans (significant at $p \leq 0.01$) for the 50, 95, and individual deductible plans, and $p = -0.05$ for the 25 percent plan. For \$8,000, the set of pay plans is significantly different (and lower) at $p = 0.10$. By \$12,000 the plan differences are no longer significant, but this corresponds to the upper 1 percent of total family expenses.

Because we do not expect major income effects from the copayment, likely explanations are that: (1) individuals can take advantage of going over the upper limit only if they are sick and (2) individuals do not know what their sickness will be. They respond to expected price and are risk averse. For a detailed theoretical discussion of this issue, see Keeler et al. (1977a).

Appendix C

SPLIT-SAMPLE ANALYSIS

by Naihua Duan and Willard Manning

In estimating how insurance plans affect the use of health services, we could choose among several statistical models. The choice of model is important, as we show below, because different modeling assumptions yield appreciably different estimates of the effects of insurance plans. In the course of the analysis, we considered analysis of variance (ANOVA) models, analysis of covariance (ANOCOVA) models, transformed (e.g., log) expenditure models, Adjusted Tobit models, and multipart models such as the four-part model described above. The traditional ANOVA and ANOCOVA models yielded consistent but imprecise results. Some of the more complicated models led to greater precision, but yielded inconsistent predictions of the plan response. Finally, we developed the four-part model to eliminate this inconsistency, without a substantial loss in precision.

However, by fitting such an elaborate model as the four-part model, we run the risk of overfitting the data. The additional complexity of the model may be simply fitting the noise in the data. If so, then our forecasts for other populations would be unreliable. To test for this possibility, we have employed a split-sample technique. We estimate the parameters of each model on a random half of the families in the sample, and then make forecasts for each model to the other half of the families in the sample. The competing models are evaluated in terms of mean squared forecast error and mean forecast bias on the forecast sample. A model can behave poorly on the forecast sample if it is either imprecise or inconsistent because of overfitting the estimation sample.

In earlier work (Duan et al., 1982, 1983, and 1984), we conducted such an analysis using the first two-fifths of the sample. We repeat that analysis here to see if any of our earlier conclusions would be changed by the addition of more observations (especially the addition of data from the two South Carolina sites), or by the use of better

health and income covariates than were available earlier. Because both the ANOCOVA model for untransformed expenses and the model with a Box-Cox two parameter transformation of expenses behaved so poorly in the original analysis (Duan et al., 1982, 1983), we do not examine those models here. Instead, we focus our attention on the following models: ANOVA, a two-part model with two separate equations for the probability of any use and the level of (log) positive expenses; the Adjusted Tobit model, a similar two-equation model which assumes that the two error terms are drawn from a bivariate normal distribution; and the four-part model.

METHODOLOGY

We have limited this analysis to annual expenditures on all health services, excluding outpatient mental health and dental care. The deterministic specification for the two-part and Adjusted Tobit model follow those of the four-part model: The probability of any use specification is the same as that in Table A.2; the (log) level of positive expenditures is the same as that in Table A.4. The ANOVA specification includes only main effects for each of the five insurance plans.

The Alternative Models

Analysis of Variance on Untransformed Expenses. The simplest model to predict expenses as a function of insurance plan is the one-way ANOVA model. ANOVA yields unbiased forecasts irrespective of the error distribution as long as the sample is assigned to each plan randomly. By the design of the HIE, this condition is met.

However, ANOVA estimates can be very sensitive to extreme values, that is, to large expenses. The distribution of the medical expenses is highly skewed toward the positive side. As a result, the sample average does not provide an efficient estimate of the plan mean. As noted above, one case accounts for 16 percent of the 50 percent plan mean, and is responsible for the plan reversal in Table 4.1.

Moreover, we cannot estimate the effects of covariates on medical expenses by using the ANOVA model, but the effects of certain covariates (e.g., income) and their possible interactions with plan are of great interest. This could be rectified, of course, using ANOCOVA, but our earlier split-sample analysis showed ANOVA performed measurably better than ANOCOVA in modeling plan response (Duan et al., 1983). Even with only standard demographic variables, ANOCOVA overfits the data.

Two-Part Model. The essence of the two-part model is to decompose one observed random variable, medical expenditures (*MED*), into two observed random variables—"$MED > 0$" and "$MED \mid MED > 0$". We then use two equations to model these two random variables. The first equation is a probit equation for the probability that a person will receive any medical service during a year. The second equation is a linear regression for the logarithm of total annual medical expenses of users.[1] The log transformation of annual expenses for the group of users reduces dramatically, but does not eliminate, the undesirable skewness in the distribution of expenses among users. We therefore expect the estimates from this model to be more robust than those that might be obtained from ANOVA and ANOCOVA models on untransformed expenses.

More formally, the first equation is a probit equation for the dichotomous event of zero versus positive medical expense:

$$I_{1i} = x_i\delta_i + \eta_{1i},$$

$$(\eta_{1i} \mid x_i) \sim N(0, 1) \tag{C.1}$$

where medical expense is positive if $I_{1i} \geq 0$, 0 otherwise; and x_i is a row vector of given individual characteristics (e.g., plan and age).

The second equation is a linear model on the log scale for positive medical expenses *if* the person receives any medical services:

$$ln(MED_i \mid I_{1i} \geq 0, x_i) = x_i\delta_2 + \eta_{2i}, \tag{C.2}$$

where $E(\eta_{2i} \mid x_i, I_{1i} \geq 0) = 0$, x_i is a row vector of given individual characteristics, and η_{2i} is i.i.d. For the last equation, the error is not assumed to be normally distributed.

As with the four-part model, the likelihood function for this model is multiplicatively separable because of the way the conditional densities are calculated; we therefore estimate the two equations separately.

If the error term η_2 in the (log) expense equation were normally distributed, then the expected medical expense would be

$$E(MED_i) = p_i \exp(x_i\delta_2 + \sigma^2/2) \tag{C.3}$$

where $p_i = Prob_i(MED_i > 0) = \Phi(x_i\delta_1)$,
Φ = normal c.d.f., and
σ^2 = var (η_2).

[1]Note that the four-part model differs from the two-part model in that the four-part model splits the group of users into outpatient-only and any-inpatient users and models the expenses of these two user groups separately.

and where the factor $\exp(\sigma^2/2)$ is the adjustment in the mean for retransformation in the second (or conditional) equation if η_2 were normally distributed with variance σ^2. However, the normal assumption for η_2 is not satisfied for the medical expense data, because the residual distribution is still skewed. As a result of this nonnormality, the factor $\exp(\sigma^2/2)$ is not the correct adjustment in the mean for the retransformation from the logarithmic scale to the untransformed dollar scale and would lead to statistically inconsistent predictions of the mean expenditure. In the case of medical expenses, the normal retransformation estimates would be biased downward.

As an alternative to the normal retransformation, we use the smearing estimate developed by Duan (1983). The smearing estimate, a nonparametric estimate of the retransformation factor $\varphi = E[exp(\eta_2)]$, is the sample average of the exponentiated least squares residuals. The smearing estimate is statistically consistent for the retransformation factor if the error distribution does not depend on the characteristics x.

A consistent estimate of the expected expense for medical services for the two-part model is given by

$$E(MED_i) = p_i \exp(x_i\delta_2)\varphi \qquad \text{(C.4)}$$

where $\varphi = \Sigma\, exp(ln\ Y_i - x_i\delta_2)/n$.

We have examined both homoscedastic (φ is a constant) and heteroscedastic (φ varies by subgroup) versions of the smearing estimate. We found little evidence to support a preference for a heteroscedastic retransformation; see below.

Adjusted Tobit Model. The econometric literature provides an additional class of models for continuous but limited dependent variables. These models include the Tobit model (Tobin, 1958), the Adjusted Tobit model (van de Ven and van Praag, 1981b), and sample selection models (Maddala, 1983). Like our two-part model, these models are two-equation models, with one equation (typically a probit) for whether there is a positive amount, and another equation for the level of the positive amount. These models differ from ours in that they model the *potential* level of use directly instead of the *actual* level of use (Poirier and Ruud, 1981; Duan et al., 1984). They do so by explicitly estimating the correlation between the probability of any use and the potential level of use. For example, the Adjusted Tobit model is

$$I_i = x_i\delta_1 + \eta_{1i} \tag{C.5a}$$

$$w_i = x_i\beta_2 + \epsilon_{2i} \tag{C.5b}$$

$$\begin{aligned} ln(MED) &= w_i \qquad \text{if } I_i > 0 \\ &= -\infty \qquad \text{if } I_i <= 0 \end{aligned} \tag{C.5c}$$

where

$$(\eta_{1i}, \epsilon_{2i}) = N(0, \Sigma), \qquad \text{i.i.d.} \tag{C.5d}$$

$$\Sigma = \begin{bmatrix} 1 & \rho\tau \\ \rho\tau & \tau^2 \end{bmatrix} \tag{C.5e}$$

and where w_i is the potential use of medical services, if everyone used medical services. Equation (C.2) differs from (C.5c), in that (C.2) refers to observed, rather than potential, use of services.

The expected expense for medical expenses for the Adjusted Tobit model is

$$E(MED_i) = \Phi(x_i\delta_1 + \rho\tau)\exp(x_i\beta_2 + \tau^2/2) \tag{C.6}$$

following van de Ven and van Praag (1981a,b).

We employ two different estimation methods for the Adjusted Tobit model: the Mill's ratio limited-information method (LIML) and the full-information maximum-likelihood (FIML).

With LIML, we use probit regression to estimate Eq. (B.5a); then use ordinary least squares to estimate the following equation:

$$E(w_i \mid I_i > 0) = x_i\beta_2 + \rho\tau\lambda_i \tag{C.7a}$$

where

$$\lambda_i = \varphi(x_i\delta_1)/\Phi(x_i\delta_1) \tag{C.7b}$$

φ and Φ are the normal p.d.f. and c.d.f. We then use Heckman's (1979, p. 157) consistent estimate for τ^2 :

$$\hat{\tau}^2 = n_1^{-1}\Sigma[\hat{e}_i^2 + \hat{c}^2\lambda_i(x_i\delta_1 + \lambda_i)] \tag{C.7c}$$

where the summation extends over all i's for which $I_i > 0$, n_1 = number of such i's, $\hat{e}_i$ = empirical residual in Eq. (C.7a), and $\hat{c}$ = the regression coefficient for λ_i in Eq. (C.7a). (Heckman's original formula did not square c, which apparently was a typographical error.) We then estimate the correlation $\rho\tau$ by

$$\hat{\rho} = \hat{c} / \hat{\tau}. \qquad \text{C.7d)}$$

For FIML, we use a RAND-compatible version of Bronwyn Hall's MAXLIK version 1.0, using the Mill's ratio estimate as the starting values.

Four-Part Model. The four-part model was described in Sec. III.

Split-Sample Methodology

The split-sample analysis can be viewed as an application of the classical cross-validation technique (see, e.g., McCarthy, 1976). Each site-year of data is randomly split into two subsamples—an estimation subsample and a forecast subsample. From the estimation subsample, we derive estimates of the parameters (regression coefficients, variances, and smearing coefficients) for each of the models. We then forecast the expenditures for each person in the forecast subsample, using the models fitted on the estimation subsample. The forecasts are then compared with the actual medical expenditures observed.

This approach guards against overfitting the data. Some models are more complex than others. If overfitting occurs, the forecasts to a new data set will perform poorly.

The specific criteria that we use to evaluate the forecasts are based on the mean squared forecast error (MSFE)

$$MSFE = [\Sigma(\hat{MED}_i - MED_i)^2] / M \qquad \text{(C.8a)}$$

and mean forecast bias (MFB)

$$MFB = [\Sigma(\hat{MED}_i - MED_i)] / M \qquad \text{(C.8b)}$$

where the summation extends over the M individuals in the forecast subsample, MED_i is the forecast for the ith individual, and MED is his actual expense.

The measure MSFE can be reexpressed as follows:

$$\begin{aligned} MSFE &= \Sigma[\hat{MED}_i - E(MED_i)]^2 / M \\ &+ \Sigma[MED_i - E(MED_i)]^2 / M \\ &- 2\Sigma[\hat{MED}_i - E(MED_i)] \\ &\times [MED_i - E(MED_i)] / M. \end{aligned} \qquad \text{(C.9)}$$

As Eq. (C.8) indicates, the measure can be expressed as a sum of the

deficiency in the fitted model (the first term on the right side), the measurement error (the second term on the right side), and the cross-product of the two. Given the estimation subsample, the conditional expectation of the measure is:

$$E(MSFE) = [\Sigma[\hat{MED}_i - E(MED_i)]^2 + \Sigma\, var(MED_i)]/M \quad \text{(C.10)}$$

Thus, MSFE is a biased estimate of the deficiency in the fitted model, because of measurement error variance—the second term on the right side of (C.9).

The actual measure that we use to compare the models is the *difference* in the MSFE. For two competing models—say, Models 1 and 2—the conditional expectation of the difference is

$$E[MSFE(1) - MSFE(2)] = \Sigma\,[MED_i(1) - E(MED_i)]^2/M$$

$$- \Sigma\,[MED_i(2) - E(MED_i)]^2/M \quad \text{(C.11)}$$

In contrast to the measures for each model separately, the MSFE difference is an unbiased estimate of the corresponding difference in the deficiency in the fitted models. Taking the differences removes the measurement error variance term from the MSFE.

We used a subpopulation sign test to detect consistent patterns in MSFE. If one model forecasts expenditures appreciably better than another, we expect the pattern to hold consistently across the subpopulations. Therefore, the subpopulation sign test counts the number of subpopulations for which one model performs better than the other in terms of MSFE. Conditioned on the estimation subsample (and therefore conditioned on the fitted models), and under the null hypothesis of no difference between the two fitted models, the count follows a binomial distribution with probability 0.5 and sample size equal to the number of subpopulations. (Conditioned on the estimation sample, the counts in distinct subpopulations are stochastically independent, from which the binomial null distribution follows. Since the conditioned null distribution is always the same, it follows that the unconditional null distribution is also the same binomial distribution.) Significantly high or low counts indicate the existence of a consistent pattern, which we take as evidence that one model is significantly better or worse than the other. (Strictly speaking, the subpopulation sign test tests the null hypothesis that median [MSFE(1) − MSFE(2)] = 0, instead of the hypothesis E[MSFE(1) − MSFE(2)] = 0. However, under the stronger null hypothesis that MSFE(1) and MSFE(2) follow the same distribution, the difference is symmetrically distributed, and the median coincides with the expectation.)

For this analysis, the data can be naturally grouped into 145 subpopulations, one for each site, year, and plan combination.

For mean forecast bias (MFB), we used an χ^2 test for differences in plan comparisons (free versus other plans), and for bias by plan.

To guard against the possibility that our results were due to the luck of the draw, we present results for two independently drawn split samples.

EMPIRICAL RESULTS

The different models produce quite different estimates of the plan response. Table C.1 contains the plan means among the forecast sample and the plan predictions for each of the models; the ANOVA predictions are the estimation sample means. The ANOVA estimates on untransformed expenditures are noisy and lack the monotonic pattern one would expect between insurance plan and expenditure. For example, in the estimation sample, the 50 percent plan has the highest plan mean, whereas in the prediction sample, it has the lowest plan mean for both split samples. The other models yield predictions where

Table C.1

FORECAST SAMPLE MEAN EXPENDITURES AND PREDICTIONS

	First Split Sample					Second Split Sample				
	Free	25	50	95	Ind. Ded.	Free	25	50	95	Ind. Ded.
Forecast sample mean expenditure	752	692	430	604	624	763	619	457	578	616
Prediction based on										
ANOVA[a]	746	580	870	446	586	733	648	911	473	599
TWONHOM	700	522	426	378	425	763	475	461	368	475
TWOSHOM	836	623	509	452	567	936	582	566	452	583
FOURSHET	776	652	593	549	584	809	606	564	525	610
LIML	698	520	424	375	472	785	474	463	365	475
FIML	798	592	453	415	543	1191	670	556	494	720

[a]Means of estimation sample.

TWONHOM = two-part model with normal theory homoscedastic retransformation.

TWOSHOM = two-part model with smearing homoscedastic retransformation.

FOURSHET = Four-part model with smearing retransformation heteroscedastic for outpatient only equation, and homoscedastic for inpatient equation.

LIML = limited information version of Adjusted Tobit (selection) model.

FIML = full information version of Adjusted Tobit (selection) model.

predicted expenditures fall monotonically as the coinsurance rate falls. But these estimates differ appreciably. In particular, the two variants of the Adjusted Tobit model differ by about $70 per enrollee per year in their forecasts in the first split sample and by $260 in the second.

Forecast Bias

Of the models considered, only the four-part model yielded forecasts of the response to insurance plans that were neither appreciably nor statistically significantly biased; see Table C.2. For each split sample, the first column presents the Wald test for a bias in the forecasted differences among the plans, by model.[2] The second column presents

Table C.2

FORECAST SAMPLE BIAS

(Standard error in parentheses)

	First Split Sample				Second Split Sample			
	No Plan Difference[a]	All Plan Biases=0[b]	Average Bias		No Plan Difference[a]	All Plan Biases=0[b]	Average Bias	
Model	χ^2 (4)	χ^2 (5)	Mean	t	χ^2 (4)	χ^2 (5)	Mean	t
JOVA	56.60[c]	60.94[c]	31.80 (32.19)	0.99	59.04[c]	68.33[d]	−7.90 (29.64)	−0.27
VONHOM	7.91[e]	23.52[c]	123.40 (29.62)	4.17[c]	12.02[d]	25.51[c]	93.00 (27.35)	3.40[c]
VOSHOM	10.54[d]	10.65[e]	18.20 (29.59)	0.61	19.16[c]	19.86[c]	−33.50 (28.04)	−1.20
)URSHET	7.32	7.41	8.20 (29.25)	0.28	4.88	5.46	−11.60 (27.00)	−0.43
ML	8.01[e]	24.39[c]	126.00 (29.62)	4.25[c]	13.92[d]	26.11[c]	85.50 (27.58)	3.10[c]
ML	9.33[e]	11.88[d]	52.30 (29.77)	1.76[e]	36.78[c]	51.48[c]	−176.40 (82.17)	−5.48[c]

[a]Null hypothesis is no different in bias in forecasts between free and other four plans.
[b]Null hypothesis is no different in bias in forecasts for all five plans.
[c]Significant at 1 percent level.
[d]Significant at 5 percent level.
[e]Significant at 10 percent level.

[2]The forecast bias was regressed on an intercept (for the free plan) and indicators for the 25, 50, 95, and individual deductible plans. The test is for the coefficients of the last four variables to be zero.

the χ^2 statistic for a bias in the forecasts by plan.[3] The third column gives the grand mean of the forecast bias across all plans.

Even the ANOVA forecasts performed poorly, in large part because of the largest expenditure in the study falling in the 50 percent estimation sample of both split samples. ANOVA was not significantly biased for the other four plans. All of the two-equation models—the normal two-part model, the smearing two-part model, and the LIML and FIML estimates of the Adjusted Tobit models—yielded significantly biased estimates of the plan response in both split samples. The normal theory two-equation models also underforecasted the grand mean by $52 to $126 per person per year in the first sample. In the second split sample, the two-equation models both under- and overpredicted the grand mean, by substantial amounts.

In contrast, the smearing two-part model did not significantly misforecast the grand mean in either sample. The four-part model was off by $8 per person per year for the first split and $12 per person per year in the second. ANOVA was off by $32 in the first split sample and $8 in the second. Neither ANOVA nor the four-part model was significantly different from zero for either split sample.

We have also compared the models in terms of the subpopulation sign tests for MSFE and MFB; see Tables C.3 and C.4. For each split sample, we provide weighted comparisons of the proportion of site/year/plans where one model has lower MSFE or MFB than another. If the fraction is greater than 50 percent, Model 1 has lower MSFE, whereas if the fraction is less than 50 percent, Model 2 has lower MSFE.

We performed a weighted comparison because some of the plans have very small samples for some sites and years. For example, the 50 percent plan subsamples in the Massachusetts and South Carolina sites are in the teens in Years 4 and 5 of the study. These small cells should be much less reliable than the larger cells. The weighted approach makes the comparisons population based. In Duan et al. (1983, 1984) we provided unweighted comparisons. Because the cell sizes in those analyses differed little, the results were insensitive to weighting. But these results include the last two years of the study, which do have small cell sizes.

The subpopulation sign test results are similar to our earlier results using the first two-fifths of this data (Duan et al., 1983, 1984). With one notable exception, the more complicated two-part, Adjusted Tobit, and four-part models perform significantly better (at the 1 percent

[3]The test is for all five coefficients to be zero in footnote 2.

Table C.3

WEIGHTED SUBPOPULATION SIGN TESTS FOR MEAN SQUARE FORECAST ERROR (MSFE)

		Fraction in Which Model 1 Has Lower MSFE	
Model 1	Model 2	First Split Sample	Second Split Sample
FOURSHET	TWONHOM	50.1	47.9
	TWOSHOM	52.5	60.5[a]
	LIML	47.7	51.0
	FIML	48.8	78.7[b]
	ANOVA	72.9[a]	64.0[b]
TWONHOM	TWOSHOM	55.0	76.0[b]
	LIML	61.1[a]	71.6[b]
	FIML	53.1	79.5[b]
	ANOVA	78.3[b]	68.2[b]
TWOSHOM	LIML	46.9	30.2[b]
	FIML	55.0	76.8[b]
	ANOVA	69.5[b]	54.3
LIML	FIML	48.7	79.9[b]
	ANOVA	78.7[b]	67.0[b]
FIML	ANOVA	72.8[b]	33.8[b]

[a]Significant at 5 percent level.
[b]Significant at 1 percent level.

level) than ANOVA in terms of MSFE. The FIML estimate was significantly worse in MSFE than ANOVA in the second split sample.

The comparisons tend to favor the four-part model over the two variants of the two-part model. In the case of the smearing variant in the second split sample, the comparison was significant. The inconsistency in the two-part model (by not modeling the inpatient response directly) is not large enough to make the two-part model significantly worse than the four-part model in terms of the subpopulation sign test for MSFE. There is no evidence that the additional complexity of the four-part model overfits the estimation sample. The results in Table C.2 indicate that both variants of the two-part model are appreciably, and sometimes significantly, biased but that the four-part model is not.

The outcome of the comparison of the four-part and the Adjusted Tobit models depends on the luck of the draw in the split samples. The four-part model performs weakly worse than both variants of the

Table C.4

WEIGHTED SUBPOPULATION SIGN TESTS FOR MEAN FORECAST BIAS

		Fraction in Which Model 1 Has Lower MFB	
Model 1	Model 2	First Split Sample	Second Split Sample
FOURSHET	TWONHOM	59.2	43.8
	TWOSHOM	48.0	52.8
	LIML	57.5	46.0
	FIML	51.7	73.2
	ANOVA	59.2	52.3
TWONHOM	TWOSHOM	46.4	58.3
	LIML	58.2	65.2
	FIML	44.2	66.2
	ANOVA	56.1	54.2
TWOSHOM	LIML	55.2	41.7
	FIML	50.9	66.0
	ANOVA	54.8	50.6
LIML	FIML	47.2	65.4
	ANOVA	55.5	55.3
FIML	ANOVA	55.6	39.4

Adjusted Tobit model in the first split sample, significantly better than FIML in the second split sample, and weakly better than LIML in the second sample comparison.

The normal theory variant of the two-part model performs better than either variant of the Adjusted Tobit model. All of the comparisons are significant in the second split sample, and half are in the first split sample. The normal theory variant of the two-part model also outperforms the smearing variant of the two-part model, significantly so in the second split sample.

The outcome of the comparison of the smearing variant of the two-part model with the Adjusted Tobit model depends on whether LIML or FIML is used for the Adjusted Tobit. The smearing two-part model loses to LIML, significantly so in the second split sample. The smearing two-part model outperforms FIML, significantly so in the second split sample.

The LIML variant of the Adjusted Tobit model performs significantly better in terms of the subpopulation sign test than does the FIML variant in one split sample and insignificantly worse in the other. In the second split sample, LIML is significantly better. This is somewhat surprising for two reasons. First, the LIML version of the selection models (of which the Adjusted Tobit is a special case) has a reputation for being badly behaved in practice; see Duan et al. (1984) for an example. Second, the starting values for the FIML procedure are the LIML estimates.

All of these comparisons, with the exception of ANOVA (noted above), seem to be relatively insensitive to outliers in expenditure. We repeated the estimation and subpopulation sign tests for the first split sample after deleting the largest spender in the study—a person with $148,000 in annual expense, who happened to fall into the estimation sample of both split samples on the 50 percent plan. The pattern of results in Table C.3 for the first split sample was largely unchanged.

Although the sign test is not valid for mean forecast bias (Duan et al., 1982), we provide such a comparison in Table C.4, for the reader's information. The results are mixed, depending on the split sample, for comparisons of the four-part, two-part, and the LIML variant of the Adjusted Tobit model. The FIML variant performs poorly against the four-part, two-part smearing, and LIML version of the Adjusted Tobit model. As we noted above, only the four-part model provides plan forecasts that are not significantly biased. ANOVA, the two-part smearing, and the four-part models are the only models that do not provide significantly biased forecasts of the grand mean.

Retransformation Method

In the preceding discussion, we have limited ourselves to the homoscedastic normal and smearing retransformations for the two-part model and the heteroscedastic smearing for the four-part model. We examined several retransformation approaches in our model selection. These included all four combinations of smearing and normal methods, and homoscedastic and heteroscedastic methods for both the two-part and four-part models. In the case of the normal theory models, we use homoscedastic and heteroscedastic to mean the use of a common or different estimate of the variance (on the log scale) for the level of use, given any use, across plans and years. In the case of the smearing models, we use homoscedastic and heteroscedastic to mean the same or different distributions (in any sense, scale, or shape) across different plans and years.

In the four-part model, we always use a homoscedastic smearing retransformation method for the positive inpatient sample; our sample is inadequate for a precise analysis of heteroscedasticity in the inpatient use equation. The four alternative methods apply to the expenses of those with outpatient use only.

The MSFE results from both split samples do not distinguish between heteroscedastic and homoscedastic variants. The smearing variants perform significantly better than the normal theory variants for the four-part model, whereas the normal theory variants outperform the smearing variants for the two-part model.

CONCLUSION

Multipart Models versus ANOVA

We chose to emphasize results from the multi-equation models instead of ANOVA for two reasons. First, there was an intrinsic interest in the effect of other variables (e.g., income and health status) on the use of health services, and how the plan response varied across different income, age, and health groups. Second, we were concerned that ANOVA was too imprecise, because of its sensitivity to the skewness in expenditures. Our split-sample analysis indicates that all of the multi-equation models performed better than ANOVA in terms of the subpopulation sign test for MSFE.[4] The bias results for the four-part model indicate that it was the only multi-equation model to obtain this increase in precision without running the risk of either biasing plan comparisons or the estimate of the grand mean.

Hence, this split-sample analysis vindicates our choice of the four-part model over the more conventional ANOVA approach.

Four-Part versus Two-Part Models

We chose the four-part model over the two-part model because early results indicated that the two-part model yielded biased estimates of the plan response (Duan et al., 1982, 1983). There are two sources of bias in the two-part estimates. First, the distribution of positive expenditures is longer in the right tail than one would expect from a log normal variate. Second, medical expenditures are the sum of outpatient and inpatient expenditures. Inpatient expenditures are much rarer than outpatient use, have a different mean and variance on the log scale, and have a different response to plan and other covariates.

---

[4]With the exception of the FIML variant of the Adjusted Tobit model, in the second split sample.

If the error term in the conditional expenditure equation were homoscedastic, then the smearing retransformation would solve the bias issue. However, in the presence of the second problem, the smearing estimate is inconsistent (Duan, 1983).

The split-sample analysis lends some support to our choice of the four-part model over the two-part model. The two-part models do yield biased plan forecasts. The normal theory variant yields significantly biased plan comparisons and significantly underforecasts the grand mean. The smearing version yields significantly biased plan comparisons and misforecasts the grand mean. The overall bias in the smearing variant of the two-part model seems to be of indeterminate sign. In contrast, the four-part model does not yield significantly biased forecasts.

In terms of mean square forecast error, there does not appear to be any pattern to the four- and two-part comparisons, other than that the normal two-part model beats the smearing two-part model.

Four-Part versus Adjusted Tobit Models

In principle, the Adjusted Tobit model shares the same theoretical liabilities that the two-part models have: skewness in the conditional expenditures because of inpatient expenses, and a failure to model the heteroscedasticity because of the different inpatient and outpatient response.

The split-sample analysis supports our choice of the four-part model over the Adjusted Tobit model. The Adjusted Tobit model yields significantly biased plan forecasts, just as the other two-part normal theory model does. Both the LIML and FIML variants yield significantly biased plan comparisons. In both split samples, the LIML version underforecasts the grand mean. The FIML version significantly misforecasts the grand mean. Given these results, the overall bias in the Adjusted Tobit model appears to be of indeterminate sign. In contrast, the four-part model does not yield significantly biased plan forecasts.

In terms of mean square forecast error, there does not appear to be any pattern to the four-part and Adjusted Tobit model comparisons other than the fact that the four-part model is no worse in weighted comparisons and is significantly better than FIML in one split sample.

The erratic behavior of the FIML variant of the Adjusted Tobit model is distressing. The magnitude of the bias in one split sample is appreciable, whereas in the other it is of the opposite sign and very large. In one split sample, the estimate of the interequation correlation is 0.04; in the other it is 0.62. These differences occurred in estimation

sample sizes of about 10,000 observations. If the true model were an Adjusted Tobit model, then the estimates would be consistent but could be imprecise in finite samples. However, one would expect better behavior in samples of the size that we used.

Two-Part versus Adjusted Tobit Models

The split-sample analysis lends weak support to the two-part model over the Adjusted Tobit model. Both models, with normal theory variants, yield biased forecasts. In our two split samples, there was no apparent pattern to which model had the larger bias. In terms of mean square forecast error, the normal theory variant of the two-part model outperformed the Adjusted Tobit model. The smearing variant of the two-part model did better than FIML but not better than LIML.

In our original modeling, we chose not to use selection models, such as the Adjusted Tobit, for several reasons. First, there is no conceptual reason why it is necessary to model the correlation between the likelihood of having any use and the level of use equations. In this case, the correlation coefficient is a nuisance parameter. There is ample precedent for excluding such parameters in conditional likelihood estimation (McCullagh and Nelder, 1983; and Duan et al., 1985). Second, we have shown elsewhere that our model is consistent in the presence of correlated errors across equations (Duan et al., 1984). Third, our earlier analysis showed that the LIML estimates of the interequation correlation are unstable, in large part because the Mills ratio used in LIML is highly collinear with the independent variables in the level of use equation (Duan et al., 1984). Fourth, this collinearity raises the problem that one cannot distinguish the LIML version of the Adjusted Tobit from a two-part model, especially if there is some true nonlinearity (e.g., a quadratic term) or heteroscedasticity in the conditional (level of use) equation.

Fifth, we rejected the Adjusted Tobit model because it relies on a strong set of distributional assumptions. If those assumptions are not met, the estimates of the insurance plan effect will be inconsistent. For example, the FIML variant is known to be extremely sensitive to minor departures from the underlying assumptions. Unfortunately, one cannot test the full set of distributional (in our example, normality and i.i.d.) assumptions embedded in the Adjusted Tobit models. To do so would require observing use by those who do not use or knowing a priori that some variables are excluded from some equations.

Finally, the concern of the Adjusted Tobit model with correlation generates a misplaced concern with the wrong tail of the distribution. In an application such as ours, we would expect a positive correlation

between the equations, if one exists. The resulting selection effect is to remove cases from the left tail of users. But the left tail does not matter that much; its "correction" adds very little to the grand mean. What really matters is the extra skewness (relative to a normal or log normal variate) in the right tail resulting from inpatient use. Inpatient users account for 69 percent of all medical expenses. The bottom 50 percent of the population (users and nonusers combined) account for less than 5 percent of total expenditures (see Table A.10). Thus, by concentrating on the left rather than the right tail, the Adjusted Tobit model is truly penny wise and pound foolish.

For a fuller discussion of some of these issues, see Duan et al. (1982, 1984, 1985), Poirier and Ruud (1981), and Maddala (1983, 1985).

Appendix D

ANOVA ESTIMATES OF PLAN RESPONSES, BY SUBGROUPS

Table D.1

MEAN ANNUAL USE OF MEDICAL SERVICES, BY INCOME GROUP AND PLAN

(Standard error in parentheses)

Plan	Probability of Any Medical Use (%)	Probability of Any Inpatient Use (%)	Admissions	Visits	Inpatient Expenses ($)	Ambulatory Expenses ($)	Medical Expenses ($)
Lower Third							
Free	83.67	10.51	0.1322	4.4971	401.48	327.39	728.89
	(01.49)	(00.76)	(0.0126)	(0.3555)	(59.84	(20.31)	(71.79)
25 percent	72.30	10.23	0.1315	3.4704	516.43	271.87	788.30
	(02.60)	(01.10)	(0.0187)	(0.3628)	(95.28)	(27.72)	(112.90)
50 percent	68.30	07.99	0.1031	2.7603	998.59	213.95	1212.50
	(04.13)	(01.64)	(0.0263)	(0.3664)	(481.20)	(31.04)	(492.20)
95 percent	58.35	09.04	0.1206	2.2195	377.25	186.66	563.90
	(02.93)	(00.93)	(0.0153)	(0.2375)	(82.60)	(20.92)	(96.99)
Individual deductible	65.00	08.92	0.1054	2.4269	294.29	198.33	492.62
	(02.59)	(00.89)	(0.0116)	(0.2307)	(46.24)	(19.72)	(55.50)
Middle Third							
Free	87.40	10.47	0.1343	4.3229	449.26	324.03	773.29
	(01.39)	(00.80)	(0.0125)	(0.2638)	(60.23)	(18.50)	(72.82)
25 percent	79.07	07.57	0.0926	2.7327	279.90	212.06	491.96
	(02.35)	(01.06)	(0.0137)	(0.2474)	(48.65)	(21.34)	(64.50)
50 percent	73.66	05.35	0.0673	2.6020	198.94	180.94	379.88
	(04.17)	(01.12)	(0.0177)	(0.2836)	(71.12)	(18.27)	(83.13)
95 percent	70.46	08.50	0.1052	2.9283	318.76	206.84	525.59
	(03.14)	(01.04)	(0.0136)	(43.15)	(49.25)	(19.10)	(63.10)
Individual deductible	71.51	09.75	0.1198	2.6603	433.18	198.30	631.48
	(02.53)	(00.89)	(0.0127)	(0.2205)	(83.78)	(14.14)	(87.22)
Upper Third							
Free	89.19	09.81	0.1176	4.8229	375.81	367.03	742.84
	(01.31)	(00.77)	(0.0112)	(0.2427)	(45.59)	(16.97)	(55.61)
25 percent	86.08	07.08	0.0865	3.8756	312.65	302.49	615.14
	(01.44)	(00.90)	(0.0126)	(0.3608)	(67.65)	(26.42)	(86.49)
50 percent	87.60	08.46	0.1083	3.6732	279.82	274.06	553.88
	(02.37)	(01.25)	(0.0169)	(0.4329)	(55.21)	(32.62)	(71.38)
95 percent	75.13	06.00	0.0678	3.0730	239.87	216.87	456.76
	(02.57)	(00.81)	(0.0099)	(0.3288)	(50.41)	(22.34)	(65.32)
Individual deductible	80.40	10.03	0.1187	4.0322	379.20	315.74	694.94
	(2.50)	(1.05)	(0.0151)	(0.3988)	(72.40)	(26.26)	(86.18)
Test for plan interaction chi-square[a]	18.7453	13.6515	14.5444	10.7602	13.4042	11.1950	15.0452

NOTES: Thirds of the income distribution are site-specific for the enrollment population. Estimates are by ANOVA.

[a]Degrees of freedom = 8.

Table D.2

MEAN ANNUAL USE OF MEDICAL SERVICES, BY AGE GROUP AND PLAN

(Standard error in parentheses)

Plan	Probability of Any Medical Use (%)	Probability of Any Inpatient Use (%)	Admissions	Visits	Inpatient Expenses 6$)	Ambulatory Expenses ($)	Medical Expenses ($)
dren (age<18)							
·ee	84.01	05.08	0.0580	3.5758	146.08	196.77	342.85
	(01.33)	(00.49)	(0.0064)	(0.1712)	(32.12)	(8.26)	(34.14)
) percent	75.34	04.38	0.0473	2.5571	144.45	146.50	290.95
	(02.21)	(00.59)	(0.0070)	(0.2044)	(36.81)	(10.41)	(41.33)
) percent	73.94	05.67	0.0727	3.0426	295.59	175.05	470.64
	(03.86)	(00.97)	(0.0154)	(0.3770)	(167.50)	(20.52)	(170.20)
) percent	63.80	03.84	0.0443	2.2217	79.01	116.56	195.57
	(02.63)	(00.57)	(0.0073)	(0.2032)	(17.86)	(9.39)	(23.72)
dividual	67.12	05.38	0.0622	2.3769	111.96	137.71	249.67
leductible	(02.53)	(00.68)	(0.0087)	(0.2239)	(17.02)	(11.57)	(23.80)
lt (age≥18)							
·ee	88.70	13.82	0.1760	5.2161	589.37	437.73	1027.10
	(00.77)	(00.65)	(0.0105)	(0.2307)	(48.14)	(15.14)	(57.67)
) percent	81.21	11.21	0.1453	3.8888	536.30	341.20	877.51
	(01.20)	(00.87)	(0.0137)	(0.2443)	(65.59)	(20.38)	(79.42)
) percent	79.45	08.24	0.1051	3.0287	553.59	256.73	810.32
	(02.05)	(01.02)	(0.0150)	(0.2145)	(198.50)	(19.91)	(203.50)
) percent	70.35	10.70	0.1363	3.0713	475.44	261.82	737.27
	(01.68)	(00.76)	(0.0114)	(0.2173)	(58.43)	(16.32)	(68.95)
ıdividual	75.34	12.08	0.1465	3.3985	528.80	293.11	821.91
leductible	(01.33)	(00.75)	(0.0103)	(0.2031)	(63.45)	(15.37)	(68.41)
for plan teraction ıi-square[a]	1.9231	16.4935	14.0801	14.4422	0.8133	39.3075	3.5538

NOTE: Estimates based on ANOVA.
Degrees of freedom = 4.

Table D.3

MEAN ANNUAL USE OF MEDICAL SERVICES, BY HEALTH GROUP AND PLAN

(Standard error in parentheses)

Plan	Probability of Any Medical Use (%)	Probability of Any Inpatient Use (%)	Admissions	Visits	Inpatient Expenses ($)	Ambulatory Expenses ($)	Medical Expenses ($)
Lower Third							
Free	87.72	13.62	0.1807	5.7667	618.57	453.24	1071.80
	(01.15)	(00.95)	(0.0164)	(0.3790)	(85.02)	(24.83)	(101.60)
25 percent	79.08	12.72	0.1641	3.8089	633.64	338.05	971.69
	(02.06)	(01.35)	(0.0213)	(0.3678)	(104.40)	(33.31)	(129.40)
50 percent	74.38	09.61	0.1281	3.5690	1058.30	304.18	1362.40
	(04.37)	(01.59)	(0.0240)	(0.4925)	(454.60)	(40.30)	(461.80)
95 percent	72.51	11.68	0.1564	3.6624	576.83	295.85	872.71
	(02.43)	(01.17)	(0.0186)	(0.3570)	(103.20)	(27.87)	(122.50)
Individual deductible	72.73	12.61	0.1639	3.5325	599.62	306.47	906.09
	(02.62)	(01.11)	(0.0169)	(0.3041)	(103.00)	(25.04)	(114.20)
Middle Third							
Free	85.07	09.88	0.1162	4.1429	345.61	304.10	649.71
	(01.25)	(00.73)	(0.0097)	(0.2187)	(41.38)	(14.35)	(49.44)
25 percent	76.79	07.38	0.0951	3.5563	356.38	239.46	595.84
	(01.88)	(00.88)	(0.0151)	(0.3067)	(85.39)	(20.51)	(98.69)
50 percent	74.29	05.88	0.0806	2.6906	232.77	200.11	432.88
	(03.09)	(01.04)	(0.0170)	(0.2576)	(56.15)	(20.20)	(63.78)
95 percent	65.91	06.35	0.0745	2.3482	189.96	161.64	351.60
	(02.74)	(00.71)	(0.0089)	(0.2135)	(29.51)	(12.96)	(36.55)
Individual deductible	72.16	08.66	0.0970	2.9356	307.54	210.35	517.89
	(01.96)	(00.87)	(0.0105)	(0.2621)	(60.17)	(15.52)	(64.11)
Upper Third							
Free	87.69	07.67	0.0932	3.8769	286.97	274.80	561.77
	(01.08)	(00.61)	(0.0094)	(0.1957)	(34.59)	(12.22)	(40.75)
25 percent	80.22	05.74	0.0653	2.7676	180.64	216.08	396.72
	(01.66	(00.68)	(0.0079)	(0.2138)	(26.82)	(14.31)	(33.16)
50 percent	81.90	06.53	0.0746	2.9235	174.56	183.34	357.90
	(02.62)	(00.96)	(0.0124)	(0.2963)	(38.32)	(14.42)	(46.38)
95 percent	65.05	06.04	0.0712	2.2490	202.26	159.66	361.92
	(02.38)	(00.77)	(0.0102)	(0.2185)	(36.38)	(13.14)	(43.10)
Individual deductible	71.93	07.68	0.0874	2.6148	227.07	193.17	420.24
	(01.89)	(00.87)	(0.0106)	(0.2273)	(41.21)	(13.70)	(45.81)
Test for plan interaction chi-square[a]	10.2154	6.1982	3.8816	10.3797	5.8720	4.6455	5.1821

NOTES: Thirds are site-specific textiles of the General Health Index based on enrollment population. Estimates based on ANOVA.

[a]Degrees of freedom = 8.

Table D.4

MEAN ANNUAL USE OF MEDICAL SERVICES, BY SITE AND PLAN

(Standard error in parentheses)

Plan	Probability of Any Medical Use (%)	Probability of Any Inpatient Use (%)	Admissions	Visits	Inpatient Expenses ($)	Ambulatory Expenses ($)	Medical Expenses ($)
Dayton							
Free	90.20	10.75	0.1194	5.4206	437.51	403.41	840.92
	(1.63)	(01.15)	(0.0151)	(0.3859)	(79.09)	(27.58)	(97.68)
25 percent	82.32	09.14	0.1218	4.3841	491.16	325.45	816.60
	(2.13)	(01.34)	(0.0229)	(0.5171)	(103.10)	(35.05)	(131.00)
50 percent	79.75	06.05	0.0745	3.5640	533.84	251.83	785.67
	(2.66)	(01.04)	(0.0145)	(0.3492)	(247.10)	(24.13)	(252.80)
95 percent	73.74	08.38	0.1024	3.1229	353.58	230.26	583.83
	(3.69)	(01.13)	(0.0144)	(0.3165)	(62.51)	(20.99)	(76.92)
Individual deductible	78.17	09.14	0.1066	3.5609	480.34	300.78	781.13
	(3.41)	(02.14)	(0.0270)	(0.6245)	(180.20)	(54.65)	(209.10)
Seattle							
Free	88.57	10.69	0.1379	4.8857	371.96	390.44	762.40
	(01.38)	(01.04)	(0.0162)	(0.2764)	(61.62)	(22.34)	(76.72)
25 percent	80.78	08.84	0.1002	4.0035	456.41	325.29	781.70
	(02.47)	(01.27)	(0.0152)	(0.3491)	(119.20)	(33.31)	(137.90)
95 percent	71.63	07.90	0.0997	3.1697	355.71	243.55	599.30
	(03.15)	(00.99)	(0.0152)	(0.3604)	(67.04)	(28.32)	(85.88)
Individual deductible	78.36	08.18	0.0958	3.9666	342.40	297.93	640.33
	(02.12)	(00.93)	(0.0126)	(0.3728)	(89.72)	(23.63)	(96.94)
Fitchburg							
Free	90.61	11.53	0.1570	4.4958	660.22	353.50	1013.70
	(02.00)	(01.46)	(0.0286)	(0.3488)	(159.70)	(28.22)	179.00)
25 percent	87.28	09.60	0.1429	3.4531	465.49	318.12	783.62
	(03.03)	(02.35)	(0.0400)	(0.4450)	(121.90)	(52.77)	(154.30)
50 percent	8. 91	07.54	0.0905	2.6432	224.11	212.23	436.33
	(03.60)	(01.82)	(0.0269)	(0.3087)	(68.98)	(33.08)	(86.75)
95 percent	78.40	09.33	0.1307	2.7973	461.10	229.31	690.41
	(03.73)	(01.89)	(0.0365)	(0.4442)	(230.70)	(49.40)	(272.10)
Individual deductible	77.31	09.77	0.1169	2.8604	481.66	245.12	726.77
	(03.10)	(01.53)	(0.0210)	(0.3302)	(164.90)	(27.77)	(170.70)
Franklin Co.							
Free	91.00	08.15	0.1033	4.3479	358.73	306.24	664.97
	(01.43)	(00.93)	(0.0140)	(0.5252)	(61.73)	(25.91)	(75.75)
25 percent	87.45	05.81	0.0618	3.0281	142.61	225.18	367.79
	(01.86)	(01.17)	(0.0133)	(0.2552)	(33.64)	(21.40)	(44.63)
50 percent	86.36	11.36	0.1591	3.3091	705.76	255.18	960.95
	(03.42)	(02.10)	(0.0365)	(0.3930)	(358.90)	(47.80)	(367.30)

Table D.4 (continued)

Plan	Probability of Any Medical Use (%)	Probability of Any Inpatient Use (%)	Admissions	Visits	Inpatient Expenses ($)	Ambulatory Expenses ($)	Medical Expenses ($)
95 percent	72.64	06.23	0.0774	2.4415	179.34	172.05	351.40
	(04.06)	(01.82)	(0.0242)	(0.3866)	(62.77)	(25.36)	(82.60)
Individual deductible	81.36	08.92	0.1145	2.8455	279.26	236.95	516.21
	(02.53)	01.09)	(0.0197)	(0.2731)	(65.76)	(22.69)	(82.12)
Charleston							
Free	76.97	09.91	0.1151	2.9765	352.24	243.34	595.59
	(02.90)	(01.05)	(0.0152)	(0.3041)	(47.82)	(23.54)	(62.07)
25 percent	65.50	07.41	0.0858	1.6082	331.13	132.92	464.05
	(04.57)	(01.52)	(0.0182)	(0.2127)	(126.90)	(15.35)	(136.20)
50 percent	48.49	04.04	0.0505	1.3131	218.36	129.64	348.01
	(12.26)	(01.96)	(0.0277)	(0.5105)	(104.40)	(49.27)	(121.30)
95 percent	56.42	07.79	0.0926	2.0337	336.95	162.65	499.60
	(03.68)	(01.41)	(0.0169)	(0.5475)	(101.60)	(33.94)	(116.50)
Individual deductible	61.23	10.00	0.1088	1.9982	390.21	175.06	565.27
	(04.80)	(01.57)	(0.0182)	(0.2731)	(100.20)	(29.16)	(112.80)
Georgetown Co.							
Free	82.44	10.44	0.1359	4.6497	339.02	305.45	644.50
	(02.14)	(00.91)	(0.0139)	(0.5196)	(52.14)	(27.20)	(70.34)
25 percent	68.89	08.52	0.1065	2.4276	249.79	169.96	419.74
	(03.91)	01.31)	(0.0175)	(0.4596)	(51.45)	(26.31)	(70.21)
50 percent	66.28	08.14	0.1047	1.9360	168.74	135.76	304.49
	(07.17)	02.34)	(0.0348)	(0.3973)	(61.84)	(23.48)	(82.64)
95 percent	46.11	07.78	0.0958	2.1018	180.84	132.83	313.67
	(04.40)	(00.93)	(0.0153)	(0.6179)	(45.46)	(28.09)	(64.44)
Individual deductible	60.33	11.27	0.1399	2.7056	356.98	175.20	532.18
	(03.73)	(01.28)	(0.0177)	(0.4704)	(59.71)	(23.95)	(67.69)
Test for plan interaction chi-square[a]	27.0365	17.4216	17.5502	14.2549	16.5054	13.254	14.9627

NOTES: Thirds of the income distribution are site-specific for the enrollment population. Estimates are by ANOVA.

[a]Degrees of freedom = 29.

Table D.5

MEAN ANNUAL USE OF MEDICAL SERVICES, BY STATUS-YEAR GROUP AND PLAN

(Standard error in parentheses)

Plan	Probability of Any Medical Use (%)	Probability of Any Inpatient Use (%)	Admissions	Visits	Inpatient Expenses ($)	Ambulatory Expenses ($)	Medical Expenses ($)
First Year							
Free	87.07	10.26	0.1223	4.2564	360.96	320.00	680.96
	(01.00)	(00.69)	(0.0091)	(0.1895)	(50.96)	(11.25)	(55.47)
25 percent	78.91	08.40	0.0983	3.1653	300.61	252.13	552.74
	(01.71)	(00.86)	(0.0106)	(0.1900)	(48.44)	(14.89)	(55.71)
50 percent	79.14	06.95	0.0802	3.0695	589.79	227.18	816.97
	(02.85)	(01.33)	(0.0168)	(0.2520)	(313.00)	(20.38)	(315.40)
95 percent	69.82	07.10	0.0946	2.7058	210.52	186.90	397.42
	(01.99)	(00.81)	(0.0125)	(0.1967)	(33.46)	(12.45)	(40.13)
Individual deductible	74.67	09.70	0.1110	3.1612	294.29	244.03	538.32
	(01.78)	(00.85)	(0.0106)	(0.1894)	(42.95)	(13.77)	(48.40)
Middle Years							
Free	86.16	10.56	0.1348	4.6045	453.12	329.62	782.74
	(00.96)	(00.62)	(0.0093)	(0.2029)	(46.14)	(12.71)	(53.16)
25 percent	79.15	08.19	0.1078	3.2796	394.86	251.54	646.40
	(01.60)	(00.78)	(0.0127)	(0.2286)	(60.96)	(18.06)	(70.41)
50 percent	75.98	07.36	0.0976	2.9354	490.62	219.46	710.07
	(02.59)	(00.98)	(0.0156)	(0.2384)	(228.50)	(18.56)	(234.50)
95 percent	67.29	(08.09	0.0978	2.6940	326.25	197.95	524.20
	(02.18)	(00.75)	(0.0096)	(0.2111)	(52.52)	(13.26	(59.83)
Individual deductible	72.06	08.60	0.1040	2.8555	337.06	216.65	553.71
	(01.86)	(00.70)	(0.0102)	(0.1986)	(58.44)	(13.76)	(62.53)
Last Year	9						
Free	87.53	09.78	0.1226	4.7459	385.27	375.08	760.37
	(01.00)	(00.68)	(0.0104)	(0.1988)	(42.65)	(15.19)	(49.62)
25 percent	77.80	08.62	0.1055	3.6028	411.20	283.27	694.47
	(01.66)	(00.91)	(0.0133)	(0.2627)	(78.75)	(18.62)	(87.70)
50 percent	77.56	07.20	0.0942	3.1801	229.18	228.51	457.68
	(02.59)	(01.56)	(0.0214)	(0.3440)	(53.73)	(20.42)	(63.31)
95 percent	66.14	08.49	0.1056	2.8026	404.59	227.79	632.41
	(02.09)	(00.89)	(0.0123)	(0.2502)	(66.43)	(17.55)	(76.42)
Individual deductible	70.09	10.93	0.1356	3.1110	508.46	253.51	761.97
	(01.88)	(00.93)	(0.0128)	(0.2134)	(93.41)	(15.08)	(98.40)
Test for plan interaction chi-square[a]	10.683	7.1819	6.2650	7.8109	10.0069	12.1174	9.4944

NOTE: Estimates based on ANOVA.
[a]Degrees of freedom = 8.

Appendix E

SAMPLE SIZES, BY YEAR OF STUDY

Table E.1

TOTAL NUMBER OF PERSONS, BY YEAR[a]

	Year 1	Year 2	Year 3	Year 4	Year 5
Start of year	5809	5734	5677	1744	1733
Departures					
Attrite	101	93	45	16	8
Terminate	68	62	78	9	12
Die	9	19	9	9	3
Become ineligible	1	1	4	2	1
Exit normally	0	0	3916	0	1741
Additions	104	118	119	25	32
End of year	5734	5677	1744	1733	0

[a]Fee-for-service plans only.

Table E.2

NUMBER OF PERSONS ON FREE PLAN, BY YEAR

	Year 1	Year 2	Year 3	Year 4	Year 5
Start of year	1893	1912	1920	592	596
Departures					
Attrite	3	1	2	0	0
Terminate	17	28	25	2	5
Die	3	2	4	3	0
Become ineligible	0	0	2	2	1
Exit normally	0	0	1333	0	605
Additions	42	39	38	11	15
End of year	1912	1920	592	596	0

Table E.3

NUMBER OF PERSONS ON 25% PLAN, BY YEAR

	Year 1	Year 2	Year 3	Year 4	Year 5
Start of year	1137	1133	1128	378	374
Departures					
Attrite	6	12	4	4	7
Terminate	12	12	22	1	2
Die	1	5	1	2	1
Become ineligible	0	0	0	0	0
Exit normally	0	0	742	0	365
Additions	15	24	19	3	1
End of year	1133	1128	378	374	0

Table E.4

NUMBER OF PERSONS ON 50% PLAN, BY YEAR

	Year 1	Year 2	Year 3	Year 4	Year 5
Start of year	383	378	378	149	143
Departures					
Attrite	9	3	8	3	0
Terminate	0	3	4	1	0
Die	0	1	0	2	0
Become ineligible	0	0	0	0	0
Exit normally	0	0	223	0	146
Additions	4	7	6	0	3
End of year	378	378	149	143	0

Table E.5

NUMBER OF PERSONS ON 95% PLAN, BY YEAR

	Year 1	Year 2	Year 3	Year 4	Year 5
Start of year	1120	1077	1041	327	318
Departures					
Attrite	49	50	15	9	0
Terminate	14	5	13	3	1
Die	0	4	2	1	1
Become ineligible	0	0	0	0	0
Exit normally	0	0	714	0	324
Additions	20	23	30	4	8
End of year	1077	1041	327	318	0

Table E.6

NUMBER OF PERSONS ON INDIVIDUAL DEDUCTIBLE PLAN, BY YEAR

	Year 1	Year 2	Year 3	Year 4	Year 5
Start of year	1276	1234	1210	298	302
Departures					
Attrite	34	27	16	0	1
Terminate	25	14	14	2	4
Die	5	7	2	1	1
Become ineligible	1	1	2	0	0
Exit normally	0	0	904	0	301
Additions	23	25	26	7	5
End of year	1234	1210	298	302	0

BIBLIOGRAPHY

Amemiya, T., "Regression Analysis When the Dependent Variable Is Truncated Normal," *Econometric*, Vol. 41, 1973, pp. 997–1017.

American Hospital Association, "1984 Hospital Cost and Utilization Trends," *Economic Trends*, Spring 1985, Vol. 1.

Arnott, R., and J. Stiglitz, *Equilibrium in Competitive Insurance Markets—The Welfare Economics of Moral Hazard: I. Basic Analytics,* Discussion Paper 465, Queens University, Kingston, Ontario, 1982.

Arrow, K. J., "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, Vol. 53, December 1963, pp. 941–973.

Arrow, K. J., *Essays in the Theory of Risk-Bearing*, Markham Publishing Co., Chicago, Illinois, 1971.

Arrow, K. J., *Optimal Insurance and Generalized Deductibles,* The RAND Corporation, R-1108-OEO, February 1973.

Arrow, K. J., *Two Notes on Inferring Long Run Behavior from Social Experiments,* The RAND Corporation, P-5546, September 1975.

Ashenfelter, O., "Book Review of the National Supported Work Demonstration," *Journal of Economic Literature*, Vol. 24, September 1986, pp. 1268–1270.

Bailit, H., J. P. Newhouse, R. Brook, et al., "Does More Generous Dental Insurance Coverage Improve Oral Health?" *Journal of the American Dental Association*, Vol. 110, May 1985, pp. 701–707.

Barer, M. L., R. B. Evans, and G. L. Stoddart, "Controlling Health Care Costs by Direct Charges to Patients: Snare or Delusion?" Occasional Paper No. 10, Ontario Economic Council, Toronto, 1979.

Baumol, W. J., and D. F. Bradford, "Optimal Departures from Marginal Cost Pricing," *American Economic Review,* Vol. 60, June 1970, pp. 265–283.

Beck, R. G., "The Effects of Co-Payment on the Poor," *Journal of Human Resources*, Vol. 9, Winter 1974, pp. 129–142.

Brook, R. H., J. E. Ware, A. Davies-Avery, et al., "Overview of Adult Health Status Measures Fielded in Rand's Health Insurance Study," *Medical Care* (Supplement), Vol. 17, 1979, pp. 1–131.

Brook, R. H., J. E. Ware, W. H. Rogers, et al., "Does Free Care Improve Adults' Health? Results from a Randomized Controlled Trial," *New England Journal of Medicine,* Vol. 309, December 8, 1983, pp. 1426–1434.

Brook, R. H., J. E. Ware, Jr., W. H. Rogers, et al., *The Effect of Coinsurance on the Health of Adults: Results from the RAND Health Insurance Experiment,* The RAND Corporation, R-3055-HHS, December 1984.

Clasquin, L. A., *Mental Health, Dental Services, and Other Coverage in the Health Insurance Study,* The RAND Corporation, R-1216-OEO, November 1973.

Clasquin, L. A., and M. E. Brown, *Rules of Operation for the RAND Health Insurance Study,* The RAND Corporation, R-1602-HEW, May 1977.

Colle, A. D., and M. Grossman, "Determinants of Pediatric Care Utilization," *Journal of Human Resources* (Supplement), Vol. 13, 1978, pp. 115–158.

Conceptualization and Measurement of Physiologic Health for Adults (series), The RAND Corporation, R-2262-HHS, various dates and authors.

Culyer, A. J., "The Nature of the Commodity Health Care and Its Efficient Allocation," *Oxford Economic Papers*, Vol. 23, 1971, pp. 189–211.

Culyer, A. J., *Need and the National Health Service*, Rowman and Littlefield, Totowa, New Jersey, 1976.

Culyer, A. J., "Measuring Health: Lessons for Ontario," Ontario Economic Council Research Studies No. 14, University of Toronto Press, Toronto, 1978.

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., et al., "Consumer Acceptance of Prepaid and Fee-for-Service Medical Care: Results from a Randomized Trial," *Health Services Research*, Vol. 21, No. 3, 1986, pp. 429–452.

Davis, K., and L. B. Russell, "The Substitution of Hospital Outpatient Care for Inpatient Care," *Review of Economics and Statistics*, Vol. 54, May 1972, pp. 109–120.

Duan, N., "Smearing Estimate: A Nonparametric Retransformation Method," *Journal of the American Statistical Association*, Vol. 78, September 1983, pp. 605–610.

Duan, N., W. G. Manning, C. N. Morris, and J. P. Newhouse, *A Comparison of Alternative Models for the Demand for Medical Care,* The RAND Corporation, R-2754-HHS, January 1982.

Duan, N., W. G. Manning, C. N. Morris, and J. P. Newhouse, "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Economic and Business Statistics,* Vol. 1, April 1983, pp. 115–126.

Duan, N., W. G. Manning, C. N. Morris, and J. P. Newhouse, "Choosing Between the Sample-Selection Model and the Multi-Part Model," *Journal of Business and Economic Statistics,* Vol. 2, July 1984, pp. 283–289.

Duan, N., W. G. Manning, C. N. Morris, and J. P. Newhouse, "Comments on Selectivity Bias," in R. M. Scheffler and L. F. Rossiter (eds.), *Advances in Health Economics and Health Services Research,* Vol. 6, *Biased Selection in Health Care Markets,* JAI Press, Inc., Greenwich, Connecticut, 1985.

Eisen, M., C. A. Donald, J. E. Ware, Jr., and R. H. Brook, *Conceptualization and Measurement of Health for Children in the Health Insurance Study,* The RAND Corporation, R-2313-HEW, May 1980.

Ellis, R. P., "Rational Behavior in the Presence of Coverage Ceilings and Deductibles," *The RAND Journal of Economics,* Vol. 17, Summer 1986, pp. 158–175.

Evans, R. G., *Strained Mercy: The Economics of Canadian Health Care,* Butterworths, Toronto, 1984.

Feldstein, M. S., "Hospital Cost Inflation: A Study of Nonprofit Price Dynamics," *American Economic Review,* Vol. 61, December 1971, pp. 853–872.

Feldstein, M. S., "The Welfare Loss of Excess Health Insurance," *Journal of Political Economy,* Vol. 81, March/April 1973, pp. 251–280.

Feldstein, M. S., "Quality Change and the Demand for Hospital Care," *Econometrica,* Vol. 45, October 1977, pp. 1681–1702.

Feldstein, M. S., and E. Allison, "Tax Subsidies of Private Health Insurance," *The Economics of Federal Subsidy Programs,* Pt. 8. Papers submitted to the Subcommittee on Priorities and Economy in Government of the Joint Economic Committee, 93rd Congress, 2nd Session, July 1974.

Feldstein, M. S., and B. Friedman, "Tax Subsidies, the Rational Demand for Insurance, and the Health Care Crisis," *Journal of Public Economics,* Vol. 7, April 1977, pp. 155–178.

Ferber, R., and W. Hirsch, "Social Experimentation and Economic Policy: A Survey," *Journal of Economic Literature,* Vol. 16, December 1978, pp. 1379–1404.

Follette, W., and N. A. Cummings, "Psychiatric Services and Medical Utilization in a Prepaid Health Plan Setting," *Medical Care,* Vol. 5, January 1967, pp. 26–35.

Follette, W., and N. A. Cummings, "Psychiatric Services and Medical Utilization in a Prepaid Health Plan Setting, Part 2," *Medical Care,* Vol. 6, January 1968, pp. 31–41.

Friedman, Bernard, and Mark Pauly, "Cost Functions for a Service Firm with Variable Quality and Stochastic Demand," *Review of Economics and Statistics,* Vol. 63, November 1981, pp. 610–624.

Friedman, Bernard, and Mark Pauly, "A New Approach to Hospital Cost Functions and Some Issues in Revenue Regulation," *Health Care Financing Review,* Vol. 4, March 1983, pp. 105–114.

Goddeeris, J., and B. Weisbrod, "What We Don't Know about Why Health Expenditures Have Soared: Interaction of Insurance and Technology," *The Mount Sinai Journal of Medicine,* Vol. 52, November 1985, pp. 685–691.

Goldman, F., and M. Grossman, "The Demand for Pediatric Care: An Hedonic Approach," *Journal of Political Economy,* Vol. 86, April 1978, pp. 259–280.

Goldsmith, Jeff, "Death of a Paradigm: The Challenge of Competition," *Health Affairs,* Vol. 3, Fall 1984, pp. 5–19.

Hausman, J., and D. Wise (eds.), *Social Experimentation,* University of Chicago Press, Chicago, 1985.

Haveman, Robert, "Book Review of Social Experimentation," *Journal of Human Resources,* Vol. 21, Fall 1986, pp. 586–605.

Hay, J. W., and R. J. Olsen, "Let Them Eat Cake: A Note on Comparing Alternative Models of the Demand for Health Care," *Journal of Business and Economic Statistics,* Vol. 2, No. 3, July 1984, pp. 279–282.

Heckman, J., "Shadow Prices, Market Wages, and Labor Supply," *Econometrica,* Vol. 42, 1974, pp. 679–694.

Heckman, J., "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables, and a Sample Estimator for Such Models," *The Annals of Economic and Social Measurement,* Vol. 5, 1976, pp. 475–592.

Heckman, J., "Sample Bias as a Specification Error," *Econometrica,* Vol. 47, 1979, pp. 153–167.

Heller, W. W., "What's Right with Economics," *American Economic Review,* Vol. 65, 1975, pp. 1–26.

Helms, L. J., J. P. Newhouse, and C. E. Phelps, "Copayments and the Demand for Medical Care: The California Medicaid Experience," *Bell Journal of Economics,* Vol. 9, Spring 1978, pp. 192–208.

Hill, D. B., and J. E. Veney, "Kansas Blue Cross/Blue Shield Outpatient Benefits Experiment," *Medical Care,* Vol. 8, March-April 1980, pp. 143–158.

Huber, P. J., "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions," *Fifth Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1, 1967, pp. 221–233.

Keeler, E. B., J. P. Newhouse, and C. E. Phelps, "Deductibles and the Demand for Medical Care Services: The Theory of a Consumer Facing a Variable Price Schedule Under Uncertainty," *Econometrica,* Vol. 45, April 1977a, pp. 641–656.

Keeler, E. B., D. Morrow, and J. P. Newhouse, "The Demand for Supplementary Health Insurance, Or Do Deductibles Matter?" *Journal of Political Economy*, Vol. 85, August 1977b, pp. 789–802.

Keeler, E. B., J. E. Rolph, N. Duan, et al., *The Demand for Episodes of Medical Treatment: Interim Results from the Health Insurance Experiment,* The RAND Corporation, R-2829-HHS, December 1982.

Keeler, E. B., R. H. Brook, G. A. Goldberg, et al., "How Free Care Reduced Hypertension of Participants in the RAND Health Insurance Experiment," *Journal of the American Medical Association*, Vol. 154, October 11, 1985, pp. 1926–1931.

Keeler, E. B., J. Buchanan, J. E. Rolph, J. Hanley, and D. Reboussin, *The Demand for Episodes of Medical Treatment in the Health Insurance Experiment,* The RAND Corporation, R-3454-HHS, March 1988.

Leibowitz, A., W. G. Manning, E. B. Keeler, et al., "Effect of Cost Sharing on the Use of Medical Services by Children: Interim Results from a Randomized Controlled Trial," *Pediatrics*, Vol. 75, May 1985, pp. 942–951.

Levit, K. R., H. Lazenby, D. R. Waldo, and L. M. Davidoff, "National Health Expenditures, 1984," *Health Care Financing Review*, Vol. 7, Fall 1985, pp. 1–35.

Lewis, C. E., and H. Keairnes, "Controlling Costs of Medical Care by Insurance Coverage," *New England Journal of Medicine*, Vol. 292, June 18, 1970, pp. 1405–1412.

Lindsay, C. M., "Medical Care and the Economics of Sharing," *Economica*, Vol. 36, November 1969, pp. 351–362.

Lohr, K. N., R. H. Brook, C. J. Kamberg, et al., "Use of Medical Care in the RAND Health Insurance Experiment: Diagnosis- and Service-Specific Analyses," *Medical Care* (Supplement), September 1986.

Long, S. H., R. F. Settle, and B. C. Stuart, "Reimbursement and Access to Physicians' Services under Medicaid," *Journal of Health Economics,* Vol. 5, September 1986, pp. 235–52.

Luft, H. S., *Health Maintenance Organizations*, John Wiley & Sons, New York, 1981.

Maddala, G. S., *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, 1983.

Maddala, G. S., "A Survey of the Literature on Selectivity Bias as It Pertains to Health Care Markets," in R. M. Scheffler and L. F. Rossiter (eds.), *Advances in Health Economics and Health Services Research: Biased Selection in Health Care Markets,* Vol. 6, JAI Press Inc., Greenwich, Connecticut, 1985, pp. 3–18.

Manning, W. G., C. N. Morris, J. P. Newhouse, et al., "A Two-Part Model of the Demand for Medical Care: Preliminary Results from the Health Insurance Experiment," in J. van der Gaag and M. Perlman (eds.), *Health, Economics, and Health Economics,* North Holland, Amsterdam, 1981.

Manning, W. G., J. P. Newhouse, and J. E. Ware, Jr., *The Status of Health in Demand Estimation: Or, Beyond Excellent, Good, Fair, and Poor,* The RAND Corporation, R-2696-1-HHS, August 1981.

Manning, W. G., J. P. Newhouse, and J. E. Ware, Jr., "The Status of Health in Demand Estimation: Or, Beyond Excellent, Good, Fair and Poor," in V. Fuchs (ed.), *Economic Aspects of Health,* National Bureau of Economic Research, University of Chicago Press, Chicago, April 1982.

Manning, W. G., A. Leibowitz, G. A. Goldberg, W. H. Rogers, and J. P. Newhouse, "A Controlled Trial of the Effect of a Prepaid Group Practice on Use of Services," *New England Journal of Medicine,* Vol. 310, June 7, 1984a, pp. 1505–1510.

Manning, W. G., K. B. Wells, N. Duan, et al., "Cost Sharing and the Demand for Ambulatory Mental Health Services," *American Psychologist,* Vol. 39, October 1984b, pp. 1090–1100.

Manning, W. G., H. J. Bailit, B. Benjamin, and J. P. Newhouse, "The Demand for Dental Care: Evidence from a Randomized Trial in Health Insurance," *The Journal of the American Dental Association,* Vol. 110, June 1985a, pp. 895–902.

Manning, W. G., A. Leibowitz, G. A. Goldberg, W. H. Rogers, and J. P. Newhouse, *A Controlled Trial of the Effect of a Prepaid Group Practice on Utilization of Medical Services,* The RAND Corporation, R-3029-HHS, September 1985b.

Manning, W. G., H. L. Bailit, B. Benjamin, and J. P. Newhouse, *The Demand for Dental Care: Evidence from a Randomized Trial in Health Insurance,* The RAND Corporation, R-3225-HHS, August 1986a.

Manning, W. G., K. B. Wells, J. E. Ware, Jr., et al., "How Cost Sharing Affects the Use of Ambulatory Mental Health Services," *Journal of the American Medical Association,* Vol. 256, October 1986b, pp. 1930–1934.

Manning, W. G., N. Duan, W. H. Rogers, "Monte Carlo Evidence on the Choice Between Sample Selection and Two-Part Models," *Journal of Econometrics,* forthcoming, 1987a.

Manning, W. G., J. P. Newhouse, N. Duan, et al., "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review,* 1987b.

Manning, W. G., *Estimating Health Demand Functions with Health Insurance Data,* The RAND Corporation, N-2729-HHS, March 1988.

McCarthy, P., "The Use of Balanced Half-Sample Replicate in Cross-Validation Studies," *Journal of the American Statistical Association,* Vol. 71, 1976, pp. 596–604.

McCombs, J. S., "Physician Treatment Decisions in a Multiple Treatment Model: The Effects of Physician Supply," *Journal of Health Economics,* Vol. 3, August 1984, pp. 155–171.

McCullagh, P., and J. A. Nelder, *Generalized Linear Models,* Chapmin and Hall, London, 1983.

Mehran, F., "Variance of the MVUE for the Log Normal Mean," *Journal of the American Statistical Association,* Vol. 68, 1973, pp. 726–727.

Metcalf, C. E., "Making Inferences from Controlled Income Maintenance Experiments," *American Economic Review,* Vol. 63, June 1973, pp. 478–483.

Morris, C. N., "A Finite Selection Model for Experimental Design of the Health Insurance Study," *Journal of Econometrics,* Vol. 11, September 1979, pp. 43–61.

Morris, C. N., *Sample Selection in the Health Insurance Experiment: Comparing the Enrolled and Nonenrolled Populations,* The RAND Corporation, N-2354-HHS, October 1985.

Morris, C. N., J. P. Newhouse, and R. W. Archibald, "On the Theory and Practice of Obtaining Unbiased and Efficient Samples in Social Surveys and Experiments," in V. L. Smith (ed.), *Experimental Economics,* Vol. 1, JAI Press, Westport, Connecticut, 1979. Also in *Evaluation Studies Review Annual,* Vol. 5, Sage Publications, Beverly Hills, California, 1980.

Musgrave, R., *The Theory of Public Finance,* McGraw Hill, New York, 1959.

Newhouse, J. P., "A Design for a Health Insurance Experiment," *Inquiry,* Vol. 11, March 1974, pp. 5–27.

Newhouse, J. P., "Medical Care Expenditure: A Cross-National Survey," *Journal of Human Resources,* Vol. 12, Winter 1977, pp. 115–125.

Newhouse, J. P., "Insurance Benefits, Out-of-Pocket Payments, and the Demand for Medical Care: A Review of the Literature," *Health and Medical Care Services Review,* Vol. 1, July-August 1978, pp. 3–15.

Newhouse, J. P., "The Demand for Medical Care Services: A Retrospect and Prospect," in J. van der Gaag and M. Perlman (eds.), *Health, Economics, and Health Economics*, North Holland, Amsterdam, 1981.

Newhouse, J. P., "Are Medical Care Costs Too High?" *Journal of Dental Education*, Vol. 48, November 1984, pp. 587–590.

Newhouse, J. P., and C. E. Phelps, "Price and Income Elasticities for Medical Care Services," in M. Perlman (ed.), *The Economics of Health and Medical Care*, John Wiley & Sons, New York, 1974.

Newhouse, J. P., and C. E. Phelps, "New Estimates of Price and Income Elasticities of Medical Care Services," in R. N. Rosett (ed.), *The Role of Health Insurance in the Health Services Sector*, National Bureau of Economic Research, New York, 1976.

Newhouse, J. P., K. H. Marquis, C. N. Morris, et al., "Design Improvements in the Second Generation of Social Experiments: The Health Insurance Study," *Journal of Econometrics*, Vol. 11, September 1979, pp. 117–129.

Newhouse, J. P., C. E. Phelps, and M. S. Marquis, "On Having Your Cake and Eating It Too: Econometric Problems in Estimating the Demand for Health Services," *Journal of Econometrics*, Vol. 13, August 1980a, pp. 365–390.

Newhouse, J. P., et al., "The Effect of Deductibles on the Demand for Medical Care Services," *Journal of the American Statistical Association*, Vol. 75, September 1980b, pp. 525–523.

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, December 17, 1981, pp. 1501–1507.

Newhouse, J. P., W. G. Manning, C. N. Morris, et al., *Some Interim Results from a Controlled Trial of Cost Sharing in Health Insurance*, The RAND Corporation, R-2847-HHS, January 1982.

Newhouse, J. P., W. G. Manning, N. Duan, et al. "The Findings of the RAND Health Insurance Experiment: A Response to Welch et al.," *Medical Care*, Vol. 25, February 1987, pp. 157–179.

O'Grady, K. F., W. G. Manning, J. P. Newhouse, et al., "The Impact of Cost Sharing on Emergency Department Use," *New England Journal of Medicine*, Vol. 313, August 22, 1985, pp. 484–490.

Pauly, M. V., "The Economics of Moral Hazard," *American Economic Review*, Vol. 58, 1968, pp. 231–237.

Pauly, Mark V., *Medical Care at Public Expenses*; Prager, New York, 1971.

Pauly, Mark V., "Taxation, Health Insurance, and Market Failure," *Journal of Economic Literature*, Vol. 24, June 1986, pp. 629–675.

Phelps, C. E., and J. P. Newhouse, "Effects of Coinsurance: A Multivariate Analysis," *Social Security Bulletin*, Vol. 35, June 1972, pp. 20–29.

Phelps, C. E., and J. P. Newhouse, "Coinsurance, the Price of Time, and the Demand for Medical Services," *Review of Economics and Statistics*, Vol. 56, August 1974a, pp. 334–342.

Phelps, C. E., and J. P. Newhouse, *Coinsurance and The Demand for Medical Services*, The RAND Corporation, R-964-1-OEO/NC, October 1974b.

Plott, C. R., "Industrial Organization Theory and Experimental Economics," *Journal of Economic Literature*, Vol. 20, December 1982, pp. 1435–1527.

Poirier, D. J., and P. A. Ruud, "On the Appropriateness of Endogenous Switching," *Journal of Econometrics*, Vol. 16, No. 2, 1981, pp. 249–256.

Ramsey, F., "A Contribution to the Theory of Taxation," *Economic Journal*, Vol. 37, March 1927, pp. 47–61.

Roemer, M. J., C. E. Hopkins, L. Carr, et al., "Copayments for Ambulatory Care: Penny-wise and Pound-foolish," *Medical Care*, Vol. 13, 1975, pp. 457–466.

Rogers, W. H., and J. P. Newhouse, "Measuring Unfiled Claims in the Health Insurance Experiment," in L. Burstein, H. E. Freeman, and P. H. Rossi (eds.), *Collecting Evaluation Data: Problems and Solutions*, Sage Publications, Beverly Hills, California, 1985, pp. 121–133.

Rosett, R. N., and L. F. Huang, "The Effect of Health Insurance on the Demand for Medical Care," *Journal of Political Economy*, Vol. 81, March/April 1973, pp. 281–305.

Rothschild, M., and J. Stiglitz, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, Vol. 90, November 1976, pp. 629–650.

Scitovsky, A. A., and N. M. Snyder, "Effect of Coinsurance on Use of Physician Services," *Social Security Bulletin*, Vol. 35, June 1972, pp. 3–19.

Scitovsky, A. A., and N. M. McCall, "Coinsurance and the Demand for Physician Services: Four Years Later," *Social Security Bulletin*, Vol. 40, May 1977, pp. 19–27.

Searle, S. R., *Linear Models*, John Wiley and Sons, New York, 1971.

Sloss, E., E. B. Keeler, R. H. Brook, et al., "Effect of a Health Maintenance Organization on Physiologic Health: Results from a Randomized Trial," *Annals of Internal Medicine*, Vol. 106, January 1987, pp. 130–138.

Stewart, A., J. E. Ware, Jr., and R. H. Brook, "The Meaning of Health: Understanding Functional Limitations," *Medical Care,* Vol. 15, No. 11, November 1977, pp. 939–952.

Stewart, A. L., J. E. Ware, Jr., R. H. Brook, and A. Davies-Avery, *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.

Stewart, A. L., J. E. Ware, Jr., and R. H. Brook, *Construction and Scoring of Aggregate Functional Status Indexes: Volume I,* The RAND Corporation, R-2551-HHS, January 1981a.

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, No. 5, May 1981b, pp. 473–488.

Taylor, L. D., "The Demand for Electricity: A Survey," *Bell Journal of Economics*, Vol. 6, No. 1, 1975, pp. 74–110.

Tobin, J., "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, Vol. 26, 1958, pp. 24–36.

U.S. Department of Health and Human Services, "Utilization of Short Stay Hospitals: United States, 1982," *Vital and Health Statistics,* Series 13, No. 78, (PHS)84-1739, Government Printing Office, Washington, D.C., 1984.

U.S. Department of Health and Human Services, "Current Estimates from the National Health Interview Survey: United States, 1982," *Vital and Health Statistics,* Series 10, No. 150, (PHS)85-1578, Government Printing Office, Washington, D.C., 1985.

Valdez, R. B., *The Effects of Cost Sharing on the Health of Children,* The RAND Corporation, R-3270-HHS, March 1986.

Valdez, R. B., R. H. Brook, J. E. Ware, et al., "The Consequences of Cost Sharing for Children's Health," *Pediatrics*, Vol. 75, May 1985, pp. 957–961.

van de Ven, W. P, and B. M. van Praag, "The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection," *Journal of Econometrics*, Vol. 17, November 1981a, pp. 229–252.

van de Ven, W. P., and B. M. van Praag, "Risk Aversion of Deductibles in Private Health Insurance: Application of an Adjusted Tobit Model to Family Health Care Expenditures," in J. van der Gaag and M. Perlman (eds.), *Health, Economics, and Health Economics*, North Holland, Amsterdam, 1981b.

Veit, C. T., and J. E. Ware, "The Structure of Psychological Distress and Well-Being in General Populations," *Journal of Consulting and Clinical Psychology,* Vol. 51, 1983, pp. 703–742.

Waldo, D., and H. C. Lazenby, "Demographic Characteristics and Health Care Use and Expenditures by the Aged in the United States, 1977- 1984," *Health Care Financing Review*, Vol. 6, Fall 1984, pp. 1–29.

Ware, J. E., "Scales for Measuring General Health Perceptions," *Health Services Research*, Vol. 11, 1976, pp. 596–415.

Ware, J. E., S. A. Johnston, A. Davies-Avery, and R. H. Brook, *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., R. H. Brook, A. Davies-Avery, et al., *Conceptualization and Measurement of Health for Adults in the Health Insurance Study: Vol. I, Model of Health and Methodology*, The RAND Corporation, R-1987/1-HEW, May 1980a.

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 1980b.

Ware, J. E., R. H. Brook, W. H. Rogers, et al., "Comparison of Health Outcomes at a Health Maintenance Organization with Those of Fee-for-Service Care," *The Lancet*, Vol. 1, No. 848, May 3, 1986, pp. 1017–1022.

Wells, K. B., W. G. Manning, Jr., N. Duan, et al., *Cost Sharing and the Demand for Ambulatory Mental Health Services*, The RAND Corporation, R-2960-HHS, September 1982.

Williams, A. W., J. E. Ware, 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, 1981, pp. 324–336.

Zeckhauser, R. J., "Medical Insurance: A Case Study of the Trade-Off between Risk Spreading and Appropriate Incentives," *Journal of Economic Theory*, Vol. 2, March 1970, pp. 10–26.

RAND/R-3476-HHS