A Controlled Trial of the Effect of a Prepaid Group Practice on the Utilization of Medical Services

Willard G. Manning, Arleen Leibowitz,
George A. Goldberg, William H. Rogers,
Joseph P. Newhouse

Rand HEALTH INSURANCE EXPERIMENT SERIES

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

ISBN: 0-8330-0679-7

Published by The Rand Corporation

R-3029-HHS

A Controlled Trial of the Effect of a Prepaid Group Practice on the Utilization of Medical Services

Willard G. Manning, Arleen Leibowitz, George A. Goldberg, William H. Rogers, Joseph P. Newhouse

September 1985

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

Rand
1700 MAIN STREET
P.O. BOX 2138
SANTA MONICA, CA 90406-2138

PREFACE

This report compares people's utilization of health services under a prepaid group practice plan with that under fee-for-service. The work was undertaken as a part of the Rand Health Insurance Experiment, a large-scale social experiment designed to investigate the effects of health insurance plans on utilization of health services and health status. Other Rand reports based on this experiment also deal with the demand for health services: J. P. Newhouse, W. G. Manning, C. N. Morris, et al., *Some Interim Results from a Controlled Trial of Cost Sharing in Health Insurance*, R-2847-HHS, January 1982; N. Duan, W. G. Manning, C. N. Morris, and J. P. Newhouse, *A Comparison of Alternative Models for the Demand for Medical Care*, R-2754-HHS, January 1982; E. B. Keeler, J. E. Rolph, N. Duan, et al., *The Demand for Episodes of Medical Treatment: Interim Results from the Health Insurance Experiment*, R-2829-HHS, December 1982. These reports discuss the initial fee-for-service results of the study and discuss the statistical problems to be faced in estimating the demand for medical care.

The present report should be of interest to persons studying the use of medical services and the organization of medical care. An abridged version of this report was published in the *New England Journal of Medicine*, June 7, 1984.

SUMMARY

Does a prepaid group practice deliver less care than the fee-for-service system when both serve comparable populations with comparable benefits? To answer this question, we randomly assigned a group of 1580 persons to receive care free of charge either from a fee-for-service physician of their choice (431 persons) or at the Group Health Cooperative of Puget Sound, or GHC (1149 persons). Another 733 prior enrollees of the GHC were studied as a control. Additionally, 782 persons who shared in the cost of their fee-for-service care, but were otherwise comparable to the first two groups, were studied to observe the effects of cost sharing.

The rate of hospital admissions for both groups at GHC was about 40 percent less than in the free-care fee-for-service group ($p < 0.01$), although ambulatory visit rates were similar. The calculated expenditure rate for all services was about 25 percent less in the two GHC groups when compared with the free-care fee-for-service group ($p < 0.01$ for the experimental group, $p < 0.05$ for the control group). Preventive visits were higher in the prepaid groups, but cannot explain the reduced hospitalization. The similarity of utilization between the two prepaid groups implies that the mix of health-risks at GHC was similar to that in the fee-for-service system.

Cost sharing, if sufficiently large, could reduce the total use of care in fee-for-service to the level of that observed at GHC, but the pattern of use was quite different. Cost sharing reduced both ambulatory visits and hospital admissions, whereas GHC did not reduce visits, but reduced admissions an insignificantly greater amount.

The lower rate of use that we observed at GHC, along with comparable reductions found in noncontrolled studies by others, suggests that the style of medicine at prepaid group practices is markedly less hospital intensive and consequently less expensive.

ACKNOWLEDGMENTS

We would like to thank our colleague Susan Hosek for a careful review of the manuscript. We would also like to thank our Rand colleagues Rae Archibald, Robert Brook, Marie Brown, Maureen Carney, Allyson Davies, Naihua Duan, Emmett Keeler, and Kenneth Krug; our former colleagues Carl Morris and Marshall Rockwell, Jr.; William Schwartz of Tufts University; and our past and present Health and Human Services project officers, Larry Orr and James Schuttinga, for helpful advice and comments and for assistance over the years in implementing the project. We also thank Glen Slaughter, Cliff Wingo, Marilyn Hecox, Lauron Lindstrom, Judi Wilson, Tom Weston, and their colleagues at Glen Slaughter and Associates for processing claims and maintaining the status of the sample during the experiment, and Bernadette Benjamin and Jack Seinfeld for their meticulous programming and data management. Finally, we are indebted to the Group Health Cooperative of Puget Sound for agreeing to participate in the study, and especially to its Director of Research during the experimental period, Richard Handschin, for his assistance in the trial's implementation and his comments on this manuscript. Neither the Cooperative nor any of the above-named people necessarily agree with or endorse the findings reported here.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS vii

TABLES xi

Section

I. INTRODUCTION 1

II. METHODS 3
- Design of the Trial 3
- Analytical Methods 10
- Measurement of Use 11

III. RESULTS 13

IV. DISCUSSION 18
- The Literature on Prepaid Group Practice 18
- The Lower Hospitalization Rate at GHC 19
- How Many Dollars Would Be Saved if HMO Enrollment Increased? 19
- Policy Implications 21

Appendix: STATISTICAL METHODS 25

REFERENCES 31

TABLES

1. Insurance Plans 3
2. Average Characteristics at Baseline of Those Who Accepted and Refused Offer to Enroll, by Plan 6
3. Comparison of Means Among Study Populations at Enrollment 8
4. Reasons for Completion of Study in the Seattle Area by Plan 9
5. Comparison of Likelihood of Using Any Service, Likelihood of Hospitalization, and Imputed Annual Expenditure 14
6. Admission and Face-to-Face Visits, Annual Rates 16
7. Annual Use Outside GHC 17
8. Percent of Sample with One or More Admissions per Year for Children and Adults 20

I. INTRODUCTION

Health Maintenance Organizations (HMOs) have been advocated for many years as an important innovation in medical care delivery; indeed, for a decade, federal legislation and subsidies have encouraged their formation. Previous studies played a large role in persuading the Congress and the Executive Branch to promote enrollment in HMOs. They indicated that the prepaid group practice (PGP) variant of HMOs has ambulatory visit rates similar to those in fee-for-service medicine, but reduces hospital admission rates by as much as 40 percent,[1] a prospect with wide appeal.

But a cloud of suspicion has lingered over the ability of HMOs to deliver on the promise of reduced cost. Virtually all comparisons of HMOs with fee-for-service medicine have used a self-selected sample; that is, they compared people who had voluntarily chosen HMOs with those who either had opted for fee-for-service medicine or who had no choice. If those people choosing HMOs were healthier than those choosing to receive care from the fee-for-service system, the observed reduced use at HMOs could simply be an artifact. Indeed, the single previous randomized study found that one PGP did lower hospital use but experienced sufficiently greater ambulatory use to make it significantly more expensive overall (Perkoff, Kahn, and Haas, 1976). This finding, however, has not dimmed enthusiasm for PGPs, perhaps because the particular PGP studied was small and newly established.

To isolate the relationship between prepayment and use of services, we have conducted a controlled trial in one well-established PGP. We sought to answer two questions:

(1) When people who previously received care from fee-for-service physicians are randomly assigned to receive care at a PGP, how does their use differ from that of similar people who remain with fee-for-service physicians?

(2) When people who previously received care from fee-for-service providers are randomly assigned to receive care at a PGP, how does their use differ from that of people who are already enrolled in the PGP?

The prepaid group practice where we conducted the study, Group Health Cooperative of Puget Sound (GHC), is located in Seattle, Washington, was established in 1947, and currently has an enrollment

[1]See Gaus, Cooper, and Hirschman (1976); Luft (1978, 1980, 1981).

of 324,000 people, about 15 percent of the Seattle area population. Its history can be found in MacColl (1966). In 1976, at the beginning of our study, GHC owned its own hospital; in 1977, it opened a second hospital.

II. METHODS

DESIGN OF THE TRIAL

We compared the groups shown in Table 1. All but the last were sampled from the Seattle area population who were *not* enrolled in GHC in 1976 but who were otherwise eligible for the trial. Those ineligible included people over 62 at the time of enrollment; Seattle area families whose incomes exceeded $56,000 (1983 dollars; this excluded 1 percent of the families contacted); the institutionalized; the military and their dependents; veterans with service-connected disabilities; and those eligible for disability Medicare and the end-stage renal dialysis programs.

Participants in the plans shown in the second and third rows of Table 1 were assigned to plans that covered virtually all health services from fee-for-service physicians and ancillary personnel such as speech therapists. In one group, the services were provided at no cost to the participant; this plan is referred to as the Free Fee-for-Service plan. Its benefits are identical to those of the GHC Experimental group, except for the restriction to GHC providers that applies to the GHC

Table 1

INSURANCE PLANS

Plan	Number of Persons
Experimentals[a]	
Group Health Cooperative	1149
Fee-for-service	
Free	431
Combined pay plans[b]	782
Controls: Group Health Cooperative[c]	733

[a]Sampled from Seattle population not enrolled in Group Health Cooperative at the beginning of the study.

[b]All plans with out-of-pocket costs.

[c]Sampled from Group Health Cooperative membership with at least one year of enrollment at the beginning of the study.

group. In the Experimental plans shown in the third row of Table 1, participants had to share in the cost of their medical services. They paid 25 or 95 percent of their medical bills, subject in most cases to a limit on out-of-pocket expenditure of up to $1000 per family (less for the poor). In one pay fee-for-service plan, the Individual Deductible plan, participants paid 95 percent of outpatient bills up to $150 per person or $450 per family per year; all inpatient services were free. The participants in these fee-for-service plans formed part of the sample we previously used to assess how cost sharing affects utilization and health outcomes.[1]

Participants in the GHC Experimental plan received free services at GHC. If GHC did not provide the service (e.g., chiropractic), the plan fully covered services sought outside GHC. But if the GHC Experimentals on their own sought care outside GHC for a service that GHC provided, they were reimbursed only 5 percent of the cost. By contrast, referrals that GHC made to fee-for-service providers, as well as emergency out-of-area care, were fully covered. Except for the restriction to GHC providers and facilities, the benefits of the GHC Experimentals were identical to those of the Free Fee-for-Service group.

Although there was an element of randomization in the assignment of families to the fee-for-service and the GHC Experimental plans, a statistical method analogous to stratification was used to obtain greater comparability among the three groups than would be expected if simple random assignment were used (Morris, 1979).

The final group used in our analysis, shown in the last row of Table 1, was a random sample of the membership of GHC in 1976 that otherwise met the eligibility requirements described above and had been enrolled at GHC for at least one year. Hereafter we refer to this group as the GHC Controls. Participants in this group remained on whatever plan of benefits they had at the start of the experiment. Although Control participants received most services free of charge, some services involved rather modest cost sharing.[2]

---

[1]See Newhouse, Manning, Morris, et al. (1982) for utilization results and Brook, Ware, Rogers, et al. (1983) for adult health status results.

[2]Control participants typically paid a copayment of $8 to $10 per outpatient psychotherapy visit in excess of 10 free visits per year. There was also some cost sharing for drug, vision, and hearing benefits, and inpatient mental health benefits were minimal. The overall amount of cost sharing is sufficiently small, however, that it should little affect the comparisons of aggregate medical use made in this report. Partly because of the cost-sharing differences, however, the expenditure of all four groups on drugs, supplies, and outpatient psychotherapy services is excluded from our analysis. These excluded categories accounted for around a fifth of all medical expenditure (excluding dental expenditure) in the Seattle fee-for-service sample. Dental services, which GHC does not provide, are also excluded from our analysis.

At enrollment, each group was composed as follows:

- The GHC Experimental group, 1149 persons (448 families)
- The Free Fee-for-Service group, 431 persons (162 families)
- The combined Pay Fee-for-Service group, 782 persons (319 families)
- The GHC Control group, 733 persons (301 families)

The GHC Controls were enrolled for five years; half the GHC Experimentals were enrolled for five years and the remainder for three years; 25 percent of the fee-for-service sample was enrolled for five years and the remainder for three years. Assignment to three- or five-year participation was made at random. Because we found that use rates did not differ significantly between the three- and five-year groups, we have combined them in the analyses presented here.[3] The sample used in our analyses consists of those originally enrolled participants while they remained in the experiment and in the Seattle area.

Most people who refused to participate in the trial did so before we made assignments to experimental treatments. In all, 29 percent of those originally contacted refused to participate in preliminary interviews. We have not yet completed our analysis of the characteristics of those who refused preliminary interviews in Seattle, but those who refused such interviews in another site, Dayton, Ohio, did not differ on a number of dimensions from those who accepted.

Of those people who were assigned to a plan and to whom an offer was made to participate, 5 percent refused to participate in the Free Fee-for-Service group, 21 percent refused to participate in the GHC Experimental group, 16 percent refused in the GHC Control group, and 19 percent refused the cost-sharing plans. (Those who refused were not reassigned to another plan.)

If these refusals occurred at random, our results would not be biased, but the differences in the refusal rates among plans suggest that bias may have been introduced. However, any such bias appears to be minimal; see Table 2. Although some differences appear, they are quantitatively small. For reasons of parsimony, we have not reported results that correct these differences; such correction leaves our qualitative findings unaffected.

We have also compared those who accepted the offer across plans. Health status, prior use patterns, and demographic variables did not differ significantly among those who enrolled in the fee-for-service

[3]The test statistic for no plan versus period of enrollment interactions is $F(4,3084) = 1.20$ for total expenses. For any effect (main or interacted with plan) the test statistic is $F(5,3084) = 1.15$. Both are insignificant at conventional levels.

Table 2

AVERAGE CHARACTERISTICS AT BASELINE OF THOSE WHO ACCEPTED AND REFUSED OFFER TO ENROLL, BY PLAN
(Standard errors in parentheses)

Variable	Free Fee-for-Service Accept	Free Fee-for-Service Refuse	Pay Fee-for-Service Accept	Pay Fee-for-Service Refuse	GHC Experimentals Accept	GHC Experimentals Refuse	GHC Controls Accept	GHC Controls Refuse	F-Statistic[a]
Family variables									
No. of families	154	10	311	77	434	111	128	43	
Family size (no.)	2.6	2.3	2.4	2.4	2.5	2.8	2.4	2.7	1.0
	(0.1)	(0.5)	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)	(0.2)	
Income in 1974	10.7	14.6	10.4	10.8	10.3	13.6	12.2	12.2	7.9
($1000)	(0.5)	(2.1)	(0.3)	(0.8)	(0.3)	(0.6)	(0.4)	(1.1)	
Person variables									
No. of persons	431	23	782	191	1149	310	733	145	
Age in years	25.2	34.5	25.8	26.8	24.4	28.7	25.6	26.7	5.0
	(1.1)	(4.7)	(0.7)	(1.6)	(0.6)	(1.3)	(0.8)	(2.0)	
Female (%)	51.1	47.8	50.4	57.8	51.0	50.0	52.8	52.6	0.7
	(2.6)	(10.6)	(1.9)	(3.6)	(1.6)	(2.9)	(2.1)	(4.7)	
Black (%)	1.7	0.0	2.9	1.7	1.6	3.3	4.0	5.9	1.0
	(1.1)	(0.0)	(0.8)	(1.2)	(0.7)	(1.3)	(1.2)	(4.8)	
Education of persons 21 years	13.0	13.2	13.1	12.3	13.1	12.8	13.9	14.3	1.8
and over (years)	(0.2)	(0.7)	(0.1)	(0.3)	(0.1)	(0.2)	(0.2)	(0.4)	
Self-rated health	1.5	1.7	1.5	1.5	1.5	1.5	1.4	1.5	0.8
status[b,c]	(0.0)	(0.2)	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)	(0.1)	
Amount of pain	3.3	3.1	3.3	3.4	3.3	3.3	3.2	3.2	1.5
due to health[d]	(0.0)	(0.3)	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)	(0.1)	
Amount of worry	3.2	3.5	3.3	3.2	3.2	3.4	3.2	3.1	1.7
due to health[d]	(0.1)	(0.2)	(0.0)	(0.1)	(0.0)	(0.1)	(0.0)	(0.1)	
Physician visits	3.9	2.5	3.5	4.1	4.0	3.4	4.5	5.4	1.7
in previous year	(0.4)	(0.4)	(0.2)	(0.5)	(0.2)	(0.3)	(0.3)	(0.9)	
Percentage hospitalized in	10.6	4.3	9.8	14.3	10.4	8.4	9.3	11.4	1.1
previous year	(1.6)	(10.0)	(1.4)	(3.1)	(1.1)	(1.7)	(1.3)	(3.7)	
Dental visits in	2.1	2.4	2.0	1.7	1.8	2.0	2.2	1.9	0.9
previous year	(0.2)	(0.4)	(0.2)	(0.2)	(0.1)	(0.2)	(0.2)	(0.3)	
Percent on AFDC	12.0	0.0	6.1	11.0	5.2	3.9	2.2	1.8	0.7
program	(2.0)	(0.0)	(1.2)	(3.1)	(1.0)	(1.5)	(0.8)	(2.4)	

Table 2—continued

Variable	Free Fee-for-Service Accept	Free Fee-for-Service Refuse	Pay Fee-for-Service Accept	Pay Fee-for-Service Refuse	GHC Experimentals Accept	GHC Experimentals Refuse	GHC Controls Accept	GHC Controls Refuse	F-Statistic[a]
Percent with group insurance	77.6 (3.4)	78.3 (13.3)	77.8 (2.4)	80.8 (4.5)	78.7 (2.1)	88.0 (2.8)	73.4 (2.9)	60.5 (7.1)	3.8
Percent with non-group insurance	10.4 (2.7)	17.4 (10.0)	14.0 (2.0)	9.3 (4.0)	12.0 (1.6)	10.3 (2.7)	37.7 (3.1)	51.8 (7.5)	0.4
Percent with public insurance	9.6 (1.8)	0.0 (0.0)	5.1 (1.1)	15.5 (3.7)	4.7 (1.0)	3.9 (1.5)	2.4 (0.8)	0.9 (0.4)	3.0

NOTE: The values in this table are based on baseline variables and baseline family compositions for those who accepted (enrolled) and refused. For those who enrolled, missing values could have been imputed using additional data collected at enrollment. Such imputations are included in Table 2 but are excluded here in order to keep the comparison between the accepted and refused groups free of methods effect. Family income for calendar 1976 (shown in Table 2) was only available for enrollees; to maintain comparability, family income reported in this table is for the preexperimental period. For all but the Control group, missing values comprised about 5 percent of the data; the figure is about 20 percent for the Control group. The remaining discrepancies between Tables 1 and 2 are due to changes in family composition between baseline and enrollment (three to nine months later).

[a] *F*-statistics are for Accept vs. Refuse (4 degrees of freedom). An *F* of 2.9 is significant at the 5 percent level.

[b] Significant at $p < 0.05$.

[c] Response to How would you rate your health? 1=Excellent; 2=Good; 3=Fair; 4=Poor.

[d] Responses to How much pain (worry) does your health cause you? 1=Lots; 2=Some; 3=A little; 4=None.

groups and the GHC Experimental group (see Table 3). Family size and income, race and Aid to Families with Dependent Children (AFDC) status also did not differ significantly among those enrolled in the two fee-for-service groups and the GHC Experimental group.[4] Therefore, the refusal rate differential does not appear to affect comparisons among these three (randomized) groups.

By contrast, the GHC Control group was somewhat older, had higher income, and had more education; the GHC Controls differ significantly from the GHC Experimentals in both individual and family

[4] For those with complete data, we cannot reject the hypothesis that the GHC Experimentals had the same characteristics as the fee-for-service plans. The test statistics are $F(18,19359) = 0.90$ for personal characteristics (e.g., health status) and $F(8,4236) = 0.98$ for family characteristics (e.g., income and family size).

Table 3

COMPARISON OF MEANS AMONG STUDY POPULATIONS AT ENROLLMENT

(Standard errors in parentheses)

	Free Fee-for-Service		GHC Experimentals		GHC Controls		Pay Fee-for-Service	
Variable	Mean	Std Error	Mean	Std Error	Mean	Std Error	Mean	Std Error
Family variables								
AFDC (%)	6.0	(1.9)	6.1	(1.2)	3.6	(1.2)	4.7	(1.3)
Black (%)	1.9	(1.1)	2.3	(0.7)	4.0	(1.1)	3.2	(1.0)
Family size (number)	2.7	(0.13)	2.6	(0.07)	2.4	(0.08)	2.5	(0.08)
Income (1983 $)	25300	(1200)	22400	(600)	27700	(940)	22900	(740)
Person variables								
Age (years)	25.2	(0.79)	24.6	(0.45)	26.6	(0.60)	26.1	(0.59)
Female (%)	49.7	(2.4)	51.0	(1.5)	52.7	(1.8)	51.0	(1.79)
Number of chronic complaints[a]	6.5	(0.35)	6.9	(0.22)	7.5	(0.26)	7.0	(0.26)
General Health Index[b,c]	72.9	(0.77)	73.5	(0.47)	72.8	(0.57)	74.5	(0.55)
Mental Health Index[b,d]	75.8	(0.68)	75.8	(0.41)	75.2	(0.51)	75.9	(0.52)
Percent with physical or role limitation[e]	16.8	(1.8)	15.9	(1.1)	14.6	(1.3)	14.1	(1.24)
MD visits[f]	3.8	(0.28)	4.0	(0.17)	4.5	(0.21)	3.6	(0.21)
Hospitalized in previous year (%)[f]	10.3	(1.48)	10.5	(0.92)	9.4	(1.09)	10.0	(1.08)
Education (years)[g]	12.7	(0.12)	12.8	(0.07)	13.8	(0.09)	12.8	(0.08)
Number enrolled	431		1149		733		782	

NOTE: Means and standard errors calculated with the family as unit of observation for family variables and the person as unit of observation for person variables.

[a]Applies to individuals 14+ at enrollment. See Manning, Newhouse, and Ware (1982) for details.

[b]A higher value reflects better health.

[c]See Brook, Ware, Rogers et al. (1983) for details.

[d]Applies to individuals 5+ at enrollment. See Brook, Ware, Rogers et al. (1983) for details.

[e]Physical limitations are called personal limitations in Brook, Ware, Rogers et al. (1983).

[f]Year prior to the beginning of the Health Insurance Study. We have no definitive explanation for the seemingly high GHC Control hospitalization rate.

[g]Own education if age 18 or older, otherwise education of female (if present, otherwise male) head of household.

characteristics ($p < .001$).[5] Because the Control group was not randomly assigned, however, these differences almost certainly reflect true differences between people who select GHC and the remainder of the Seattle population.

Our sample includes 3095 people with an average of 3.3 years of participation. People who moved from the Seattle area could, as a practical matter, no longer receive services at GHC. To maintain comparability with the two GHC groups, those in the fee-for-service groups who moved from the Seattle area were also omitted from our analyses starting at the time of their move. Moves from the Seattle area account for about half of the partial years of participation; see Table 4. The difference in normal completions among the experimental plans is entirely attributable to differences in the mix of three- and five-year enrollment periods.

Table 4

REASONS FOR COMPLETION OF STUDY IN THE SEATTLE AREA BY PLAN

(In percentages)

Reason	GHC Experimental	GHC Control	Free Fee-for-Service	Pay Fee-for-Service
Voluntarily withdrew	4.4	11.9	0.2	9.6
Terminated because of failure to meet study obligations	2.5	20.6[a]	3.0	3.2
Died	0.3	0.3	0.2	0.8
Moved from Seattle area	22.3[b]	10.5	20.7[b]	12.9[b]
Other	0.5	0.8	0.2	0
Completing normally	70.1	55.9	75.6	73.5

[a]Loss of eligibility for GHC (for example, because of employment change) is included in this value.

[b]These individuals were kept in the experiment; the GHC Experimentals however, were switched to the Free Fee-for-Service plan once they moved from the Seattle area. For reasons explained in the text, this group is included in the analysis only for the period of time they lived in Seattle. Once Control families moved away from the Seattle area, they were dropped.

[5]For those with complete data, the test statistics are $F(4,4236) = 9.68$ for family characteristics and $F(9,19359) = 7.55$ for personal characteristics.

ANALYTICAL METHODS

With the exception of the expenditure figures, we calculated sample means (analysis of variance) for each experimental treatment. Had we instead adjusted for participant characteristics, our results would not have materially changed.

In the case of expenditure, however, we can gain substantial precision by accounting for age and sex variation.[6] We have therefore presented results that come from a multiple regression model using plan, age, and sex as covariates. This model is discussed in the appendix.

In the analysis of visit and admission rates, we have weighted observations to correct for the length of an individual's participation in the study in the Seattle area; see the appendix. This method would be biased if the participants who left the study early did not come from the same distribution as the stayers. A number of tests we have run persuade us that the assumption that they came from the same population is valid.[7]

When we calculate imputed expenditure and the percent using one or more services in a year, we use only participants who completed a full year of participation in the Seattle area plus decedents. In the case of the percentage using a service, it is difficult to interpret data on people who completed only part of a year. In the case of expenditure, statistical methods we used did not permit including participants who did not complete a full year.[8] The expenditure values we show come

---

[6]If age and sex are not included as covariates, precision for expenditures would be reduced. It is further reduced if data are included on individuals who complete only partial years. Making both these changes leaves the estimated expenditure largely unaffected, but sufficiently degrades precision to the point that one can reject the null hypothesis of no difference in expenditure between the Experimental group and the Free Care plan with only a probability of .08 (two-tail test).

[7]The test statistic for a differential expenditure response to plan by early departure is $F(5,3083) = 0.81$. The test statistic for any difference (main effect or plan interaction) is $F(6,3083) = 0.78$. The average person who left early spent \$43 more per person per year than those who stayed until the end of the study, with a standard error of \$55. These tests are based on the generalized least squares rather than four-part model estimates (see footnote 8 and the appendix). We also compared admission and visit rates and found no significant differences for admissions ($\chi^2(5) = 2.90$) but significant differences for visits $\chi^2(5) = 13.60$, largely limited to the 95-percent plan; both of these tests excluded decedents, who had much higher use rates, but who were few enough in number that our estimates are not sensitive to them. Had we corrected for sample loss, none of the entries in Table 5 would change by more than 5 percent.

[8]The model used in the expenditure analyses requires equal time periods for each observation, because it does not allow convolution of observations. Thus, people who participated for only part of a year could appear to be different when their underlying behavior was the same. Although the part-year participants have more physical limitations and are more likely to be enrolled in less generous plans, we found no evidence that the part-year people behave differently (see footnote 7).

from averaging, within each group, the predicted values from regression equations using the actual age and sex of each individual in that group.

MEASUREMENT OF USE

Data on use at GHC came from abstracting GHC records (Goldberg, 1983). Data on out-of-plan use in the case of the GHC groups, as well as all data on use by the fee-for-service participants, came from claim forms filed with the experiment, which functioned as the participants' insurance company.

We compared the number of visits and the number of admissions, but such a partial comparison does not detect any differential intensity of service per visit or per admission between GHC and fee-for-service participants. Because actual expenditures are not available at GHC, we constructed a measure of intensity, which we call imputed expenditure. Our method for calculating imputed expenditure differed for hospital services and physician services. For admissions at the GHC hospital(s), we used the dollar figure that GHC would have charged, had it billed the case to a payer outside GHC. (GHC does bill for some admissions; two common instances are emergency admissions of a non-enrollee at a GHC hospital and Workman's Compensation cases.) For admissions at fee-for-service hospitals, we used the hospital's actual billed charges.

In the case of physician services, we compared the number of California Relative Value Studies (CRVS) units that GHC and fee-for-service physicians delivered (California Medical Association, 1975). To arrive at an imputed expenditure figure, we valued units in both systems at the same dollar figure.

To calculate dollar figures for physician services (both inpatient and outpatient), as well as for services delivered by other providers such as speech therapists, we coded the procedure(s) from the GHC medical record and the fee-for-service claim form. Each procedure was associated with a number of CRVS units from the CRVS (California Medical Association, 1975). Because the units are not commensurate across different types of services (e.g, medical, surgical), one cannot simply tally the units. Hence, we established a common denominator in order to add different types of CRVS units. GHC has established charges to other payers for each type of CRVS unit; this charge per unit was applied to each unit delivered by both GHC and fee-for-service providers, and total physician charges were calculated. The estimated total charges for the fee-for-service system did not equal the actual total; we therefore multiplied both the estimated GHC and fee-for-service total

charges by the same factor of proportionality, which was set so as to equate the estimated and actual fee-for-service charges.

One problem we faced in applying this method was that not all procedures could be associated with CRVS units. In particular, some procedures were so new that their relative value had not yet been established (RNE) and in other cases the relative value was by report (BR). (BR means that the procedure is deemed too variable to be assigned a fixed unit value, and fees are justified "By Report.") For all procedures coded RNE or BR, one of the authors of this report (Goldberg) assigned a number of units. In so doing he was blind to whether the procedure was from GHC or fee-for-service. Once a value was determined, it was used for any other instances of the procedure, whether from GHC or fee-for-service. The number of units so assigned amounted to 10 percent of the total.

A test of our procedure for imputing physician charges is the degree of correlation between actual charges and imputed charges for each person receiving fee-for-service care. Therefore, we computed a correlation coefficient between the logarithm of the actual charges and the logarithm of the imputed charges for those with some use in each year of the study. For outpatient care, the lowest of the annual correlation coefficients was 0.97; for inpatient care, the lowest correlation coefficient was 0.93. Thus, the CRVS structure we employed quite accurately mirrors the actual structure of fee-for-service charges.

In addition to comparing the rates at which the participants saw physicians, we also compared the rates of preventive care visits in the various plans. Preventive care included any well-care service other than vision, hearing, and prenatal care. Well-care services were defined by the physician's diagnosis, or by the use of certain procedures (e.g., immunizations), or by a well-care treatment history code in conjunction with the patient's indicating that the reason for visit was well-care. We used the patient's reason for visit because we feared that some fee-for-service physicians might fail to label some visits as preventive (because many standard health insurance plans, although not ours, do not reimburse for preventive services). Any such failure in labeling would bias a comparison of the amounts of preventive care that the various groups received.

III. RESULTS

Although the GHC Experimental and Control groups differ little from each other in imputed expenditure on medical services, they both differ markedly from the Free Fee-for-Service group (Table 5).[1] Imputed expenditures are 28 percent less in the GHC Experimental group than in the Free Fee-for-Service group ($p < 0.01$) and about 23 percent less in the GHC Control group ($p < 0.05$).

The magnitude of the expenditure reduction at GHC is comparable to that achieved by 95 percent coinsurance within the fee-for-service system, although the means by which expenditure is reduced are considerably different. At GHC, the percentage of enrollees seeking care is comparable to or even exceeds the percentage in the Free Fee-for-Service plan; GHC reduced expenditure compared with Free Fee-for-Service because fewer enrollees were admitted to the hospital. With 95 percent coinsurance the percentage seeking care, as well as the percentage admitted to the hospital, is notably reduced from the level of the Free Fee-for-Service plan. The reduction in use appears greatest in the Individual Deductible plan, even though inpatient services are free in this plan. The differences in expenditure between the Individual Deductible plan, the 25-percent, and 95-percent coinsurance rate plans, however, are not significant and could well reflect random variation; if data from three other sites are combined with these data, people on the Individual Deductible plan spent (insignificantly) more than those on the 95-percent coinsurance plan (Newhouse, Manning, Morris, et al., 1981, 1982).[2]

Indeed, we would caution the reader about placing any weight on the statistically insignificant reversal between the Free Fee-for-Service plan and the 25-percent coinsurance rate plan as well as on the finding that the Individual Deductible plan had the lowest expense of all. Neither of these reversals is significantly different from the pattern found

[1]The GHC Controls do have a significantly higher probability of any use of medical services ($p < .01$). Although part of this is due to different characteristics (e.g., age and sex), we found that this result persists when we control for such characteristics.

[2]We have limited the fee-for-service sample in this report to Seattle in order to maintain as much comparability as possible with the GHC Experimental group, but in so doing we have increased the influence of random error relative to our earlier results, which are more precise for the fee-for-service plans. For example, the large standard error in the 25-percent coinsurance rate plan reflects both small sample size and some extremely large hospitalizations. Estimates based on larger sample sizes can be found in Newhouse, Manning, Morris, et al. (1981, 1982).

Table 5

COMPARISON OF LIKELIHOOD OF USING ANY SERVICE, LIKELIHOOD OF HOSPITALIZATION, AND IMPUTED ANNUAL EXPENDITURE

(Standard errors in parentheses)

Plan	Percent Using Inpatient or Outpatient Service in Year	Percent with One or More Hospitalizations in Year	Imputed Annual Expenditure per Participant[a] (1983 dollars)
GHC Experimental	86.8	7.1	439
	(1.0)	(0.50)	(25)
GHC Control	91.0	6.4	469
	(0.8)	(0.55)	(44)
Fee-for-service			
Free	85.3	11.1	609
	(1.6)	(1.17)	(66)
25 percent	76.1	8.8	620
	(2.7)	(1.37)	(103)
95 percent	68.4	8.5	459
	(3.4)	(1.18)	(72)
Individual deductible	73.9	7.9	413
	(2.4)	(0.96)	(51)

NOTE: The sample consists of all participants present at enrollment while they remained in the Seattle area. Except for decedents, observations or partial years of participation are deleted.

[a]Values include both in-plan and out-of-plan use by GHC participants. The method of imputing expenditure is described in the text. The *t*-statistics to test the difference in expenditure between the GHC Experimental and the five groups below it in the table are, respectively: 0.87, 3.22, 2.22, 0.30, −0.56. Because of the inclusion of age and sex as covariates, these *t*-statistics are larger than those that would be calculated from the standard errors shown in the table.

when a larger sample is used (Newhouse, Manning, Morris, et al., 1981, 1982).

The differences between GHC and the Free Fee-for-Service plans come even more sharply into focus when we examine admissions, hospital days, and visit rates (Table 5). There are 40 percent fewer admissions ($p < 0.01$) and hospital days in the two GHC groups than in the Free Fee-for-Service plan, but face-to-face visits occur at approximately

similar rates in all three plans.[3] By contrast, participants in all the cost-sharing plans had both lower admission rates and lower visit rates than those in the free-care plan. The differences in total expenditure and admission rates between the 95-percent coinsurance plans and the two GHC groups are insignificant at the 5-percent level, but the differences in face-to-face visit rates are significant at the 5-percent level.

Although the overall face-to-face visit rates are similar between the two GHC groups and the free-care plan, the number of preventive visits is significantly higher in the two GHC groups than in the Free Fee-for-Service group; cost sharing further reduces preventive visits below the values in the Free Fee-for-Service plan (Table 6).

Outside-GHC use by the two GHC groups is relatively small (Table 7). Not surprisingly, Experimentals are more likely than Controls to seek care outside GHC. About 2 percent of the Experimental group each year sought care exclusively from ancillary providers such as chiropractors, Christian Science practitioners, and podiatrists. Half of the out-of-plan inpatient admissions were related to accidents or to psychiatric diagnoses.

---

[3]The GHC Experimental and Free Fee-for-Service plan face-to-face visit rates are not significantly different ($t = 0.22$). The rates for the GHC Controls may be higher than that for the GHC Experimentals ($t = 1.73$). The GHC Controls have an insignificantly higher visit rate than the Free Fee-for-Service Experimentals ($t = 1.46$); the difference is large but imprecisely measured due to the small Free Fee-for-Service plan sample size.

Table 6

ADMISSION AND FACE-TO-FACE VISITS, ANNUAL RATES

(Standard errors in parentheses)

Plan	Admission Rate/100 Persons[a]	Hospital Days/ 100 Persons	Face-to-Face Visits[b]	Preventive Visits[c]
GHC Experimental	8.4	49	4.3	0.55
	(0.67)	(9.6)	(0.14)	(0.02)
GHC Control	8.3	38	4.7	0.60
	(1.01)	(9.0)	(0.17)	(0.02)
Fee-for-service				
Free	13.8	83	4.2	0.41
	(1.51)	(26)	(0.25)	(0.03)
25 percent	10.0	87	3.5	0.32
	(1.43)	(28)	(0.35)	(0.03)
95 percent	10.5	46	2.9	0.29
	(1.68)	(9.9)	(0.34)	(0.04)
Individual deductible	8.8	28	3.3	0.27
	(1.20)	(5.1)	(0.33)	(0.03)

NOTE: The sample includes all participants present at enrollment while they remained in the Seattle area. For GHC Controls and Experimentals, the data include both in- and out-of-plan use.

[a]A count of all continuous periods of inpatient treatment.

[b]Includes all visits with face-to-face contact with health providers for which a separate charge would have been made in fee-for-service. Excludes radiology, pathology, and pre- and postnatal, pre- and post-operative, speech therapy, psychotherapy, dental, chiropractic, podiatry, Christian Science healer, and telephone visits.

[c]Includes well-child care, immunizations, screening examinations, routine physical and gynecological examinations, and visits with Pap smears (other than for cancer). Excludes prenatal, vision, and hearing visits. In the case of GHC, includes in-plan and out-of-plan visits.

Table 7

ANNUAL USE OUTSIDE GHC

(Standard errors in parentheses)

Type of Use	GHC Experimentals	GHC Controls
Hospital admissions per hundred[a]	0.74 (0.26)	0.21 (0.087)
Hospital days per hundred	15[b] (9)	1.4 (0.8)
Ambulatory face-to-face visits per person[a,c]	0.14 (0.02)	0.076 (0.02)
Chiropractors, podiatrists, Christian Science practitioners visits per person	0.72 (0.12)	0.12 (0.06)
Speech therapists visits per person[d]	0.0002 (0.0002)	0.007 (0.006)
Expenditures per person (1983 dollars)[d]	63 (13)	15 (5)

NOTE: The sample includes all participants present at enrollment, while they remained in the Seattle area.

[a]Comparison significant at $p < 0.05$.

[b]One case accounts for two-thirds of this mean. Inpatient psychiatric cases account for one-sixth of this mean.

[c]A face-to-face visit is one for which a separate charge would have been made in fee-for-service. Excludes radiology and pathology, and pregnancy, speech therapy, psychotherapy, chiropractic, podiatric, and Christian Science practitioner visits.

[d]Comparison significant at $p < 0.01$.

IV. DISCUSSION

THE LITERATURE ON PREPAID GROUP PRACTICE

Our results show minor and generally insignificant differences between the two GHC groups, implying that results from noncontrolled studies may not be seriously contaminated by selection effects. In particular, imputed expenditures in the Experimental group were 6 percent less than in the Control group. Based on the standard errors for these expenditure figures, which are between 5 and 10 percent of the mean, it is unlikely that there is a large difference between these groups.

The validity of our results is strengthened by the general consistency between them and the results in the literature regarding prepaid group practices. Luft's review of several noncontrolled studies found that prepaid group practice hospitalized from 15 to 40 percent less than fee-for-service (Luft, 1978, 1981). In the analogous comparisons from our study, GHC Controls were 40 percent less likely to go into the hospital than those in the Free Fee-for-Service group and about 5 to 20 percent less likely to be admitted than those in the cost-sharing groups.

It is plausible that the difference in admission rates between the GHC Control group and the Free Fee-for-Service plan should be as high as those observed in the literature. The Free Fee-for-Service plan had better ambulatory benefits than virtually all fee-for-service plans studied in the literature, and more extensive coverage of ambulatory services leads to more hospitalization among those using fee-for-service physicians (Newhouse, Manning, Morris, et al., 1981, 1982). For the same reason the difference between the admission rates in the GHC Control group and in the cost-sharing plans should be near the low end of Luft's range.

Outpatient visit rates among GHC Controls were insignificantly higher than in the Free Fee-for-Service plan but significantly higher than with cost sharing ($p < 0.01$). Luft finds roughly similar results among a variety of studies he reviews (Luft, 1981). Thus, the comparison of use among the GHC Control group with use among the fee-for-service groups resembles analogous comparisons in the literature and thus lends support to the validity of the GHC Experimental results.

THE LOWER HOSPITALIZATION RATE AT GHC

Services delivered in the hospital account for around half of all expenditure on personal health services in the United States (Waldo and Gibson, 1982). That GHC could lower them so markedly relative to the Free Fee-for-Service group invites closer scrutiny. To achieve so great a reduction in hospitalization rates, GHC could be providing more preventive care, or it could be treating cases on an outpatient basis and avoiding hospitalization. The data examined to date provide no clear explanation, however. Although preventive visits at GHC are more numerous than in the fee-for-service plans, the hospitalization rates are not significantly different from those in the fee-for-service cost-sharing plans, which had only about half as many preventive visits as the GHC plans. Moreover, about two-thirds of the preventive visits are for well-child care and gynecologic examinations. Because gynecologic and pediatric admissions account for a minority of hospitalizations, it seems unlikely that preventive care could account for much of the large difference in hospitalization rates. Indeed, despite the concentration of preventive care among children, the percentage reduction in admission rates among children is roughly similar to (and is insignificantly different from) the percentage reduction among adults ($\chi^2(5) = 3.83$, $p > 0.50$) (Table 8).

Outpatient visit rates in the two GHC groups are similar to those in the Free Fee-for-Service group; if some problems for which fee-for-service participants are hospitalized were being managed on an outpatient basis at GHC, one might expect such rates to be higher. Of course, the similar outpatient visit rates at GHC could be a combination of more intensive outpatient treatment of those whom fee-for-service physicians would hospitalize, combined with less intensive treatment of those who would not be admitted in any event. Further investigation will be required to address this possibility.

HOW MANY DOLLARS WOULD BE SAVED IF HMO ENROLLMENT INCREASED?

The 28-percent difference in imputed expenditure between the GHC Experimental plan and the Free Fee-for-Service plan is striking (Table 2). One can, however, ask how much error the imputation process might have introduced. Although more detailed analysis could yield a more precise figure, it seems unlikely that the true difference could be much less than 25 percent. Both admissions and total hospital days at GHC were 40 percent below the Free Fee-for-Service plan (Table 2), while ambulatory visit rates were similar. How many dollars might

Table 8

PERCENT OF SAMPLE WITH ONE OR MORE ADMISSIONS PER YEAR FOR CHILDREN AND ADULTS
(Standard errors in parentheses)

Plan	Children (Age Less Than 18)	Adults (Age 18 or Over)
GHC Experimental	3.5	9.2
	(0.56)	(0.68)
GHC Control	3.6	7.8
	(0.70)	(0.73)
Fee-for-Service		
Free	6.2	13.7
	(1.13)	(1.71)
25 percent	5.8	10.6
	(1.92)	(1.62)
95 percent	3.2	11.6
	(1.08)	(1.62)
Individual deductible	6.0	8.7
	(1.64)	(1.26)

NOTE: The sample consists of all participants present at enrollment while they remained in the Seattle area. Except for decedents, observations on partial years of participation are deleted. A chi-squared value for comparability of response is 3.83 with 5 d.f., $p > 0.50$.

such a reduction in use save? First, suppose the reduction in admissions was random. In that case inpatient expenditure, which accounts for somewhat over half of total expenditure, would fall 40 percent.[1] If ambulatory expenditure in the two systems were similar, total expenditure would fall by about a quarter, as our imputed figures indicate.

But suppose the 40-percent reduction in admissions was not random but rather was disproportionately made up of short-stay admissions. If this in fact occurred, it would imply that GHC also reduced the length of stay among those it did admit. Otherwise the percentage reduction in days would have been less than 40 percent. A combination of reduced admissions and reduced length of stay that together yielded a 40 percent reduction in hospital days could certainly cause a true reduction in expenditure on the order of one-quarter.

[1]See Newhouse, Manning, Morris, et al. (1981, 1982).

Moreover, our estimates of the savings do not account for any efficiencies that GHC may have enjoyed in the delivery of physician services—such as greater substitution of paramedical personnel. To estimate the magnitude of such efficiencies, if any, would require a study of costs within each system, something we did not attempt. Nonetheless, it is difficult to escape the conclusion that true expenditure was substantially less in the two GHC groups than in Free Fee-for-Service.

One might question whether the Free Fee-for-Service plan is a relevant reference group. Few individuals in the population who use fee-for-service physicians face no cost sharing, and one might, therefore, adopt the cost-sharing plans as a standard of comparison. In the extreme case of the 95-percent coinsurance plans, the differences between GHC and fee-for-service narrow sharply.

Does this mean that if the share of the population enrolled in prepaid group practices were to increase markedly from its current value of 5 percent, Americans' use of the hospital would not change much? In fact, one might expect considerably less hospitalization as HMOs' market share grows, because the national hospitalization rate is not far from the free plan value. Using fee-for-service data from four sites (including Seattle), we previously found that the annual likelihood of one or more hospitalizations among average Americans under 65 in 1977, 9.5 percent, was close to the free plan value of 10.2 percent, whereas the values in the cost-sharing plans ranged from 7.2 to 9.0 percent.[2] Because the national average appears close to the free plan rate, our data suggest scope for a substantial drop in inpatient use.

On the other hand, the reduced cost due to lower PGP inpatient use will be partially offset by the increased use of ambulatory services relative to current fee-for-service cost-sharing plans. National ambulatory use rates fall between the rates observed in the 25-percent and 95-percent fee-for-service plan estimates.[3] On balance, the net effect from increased enrollment in PGPs will still be a saving unless cost sharing in the fee-for-service system were to increase above present levels.

POLICY IMPLICATIONS

Plainly, there is much less hospitalization among those receiving care at GHC than among those with similar benefits (i.e., no cost sharing) who use fee-for-service physicians. Because of our experimental design, we can virtually rule out population characteristics as an

[2]Ibid.

[3]Ibid.

explanation of the lower hospitalization rate in GHC. We conclude that GHC physicians are simply practicing a different style of medicine from that of fee-for-service physicians. Although our study is limited to a single, not necessarily typical, prepaid group practice, the general consistency between its results and those in the literature indicates that a less hospital-intensive style of medicine than that practiced by the average physician is possible.[4]

But is such a style desirable? The results presented above shed little light on that question. We have taken extensive measures of the health status of our participants[5] and future analysis of these data should detect any pronounced effects of the different style of medical treatment on health status. Nonetheless, prepaid group practices have existed for decades, and it seems unlikely that there can be large deleterious health effects from their style of medicine.

By contrast, the different style could well affect patient satisfaction. Indeed, because many people choose not to enroll in a prepaid group practice, it seems almost certain that their expected level of satisfaction, were they to join, would be less than in fee-for-service medicine. To be sure, some may decline the option to join because they receive no cost advantage (e.g., the employer pays the entire health insurance premium). But others do pay more to receive their care from the fee-for-service system, and one can surmise that they believe they receive something in return.

Whatever the motivation of persons choosing the fee-for-service style, many argue that such persons ought to pay all the additional costs.[6] For this to occur, employers (or the government) would have to pay an equal sum for each available health plan option, rather than, as many do, more for a fee-for-service than an HMO plan. If employers did pay an equal sum, price competition between HMOs and fee-for-service insurance plans could well increase.

Some fear that increased price competition without regulation of benefit packages will bring risk selection, that is, better risks in one plan than another (Enthoven, 1980). We found no appreciable difference in use between the GHC Experimentals and Controls, implying no risk selection in this case. Nonetheless, this result may not generalize. Economic theory suggests that risk selection could occur if individuals know more about their future health demands than do insuring

---

[4]See Luft (1981) and Nobrega, Krishan, Smoldt, et al. (1982).

[5]See Brook, Ware, Rogers, et al. (1983) for an overview. More detail can be found in Rand monographs belonging to series with the governing numbers of R-2262-HHS, and R-2898-HHS; in Eisen, Donald, Ware, and Brook (1980); and Brook, Ware, Davies-Avery, et al. (1979).

[6]Enthoven (1980) and McClure (1982).

organizations or HMOs,[7] and nonexperimental studies have found evidence of risk selection.[8]

In sum, GHC delivered a different, less expensive style of medicine than did fee-for-service practitioners in the Seattle area when both treated comparable groups who faced no cost sharing. Adding cost sharing to the fee-for-service insurance plans brought expenditures more closely into line with those of GHC, but appeared to result in yet another style of care, one with markedly fewer ambulatory contacts. How the GHC style fares on dimensions other than expense remains an open question, but the lower expense at GHC cannot be explained by differences in the population that it treats.

---

[7]See Arnott and Stiglitz (1982), Cave (1984), and Rothschild and Stiglitz (1976).

[8]See Luft (1981) and Eggers and Prihoda (1982).

Appendix

STATISTICAL METHODS

We have used three alternatives to analysis of variance (ANOVA) methods to obtain the estimates reported in Tables 5 to 8. There were two major reasons for going beyond simple ANOVA. For expenditures (Table 5) there was a notable gain in precision from including age and sex as explanatory variables, and from using a statistical technique that models the distribution of expenditures. Second, for all outcomes, the usual ANOVA estimates of standard errors (and other inference statistics) are biased because they do not account for positive correlation across family members and for the same person over time. The usual ANOVA estimates of the sample means are unbiased but inefficient.

In the presence of correlated errors, the ANOVA and ordinary least squares estimates of the standard errors overstate the precision of the estimate. Our data exhibit positive correlation for some outcomes across family members and across time for each person. For the probabilities of any use in Tables 5 and 8, we used a variant of Huber's (1967) approach to correct the standard errors for correlation. The correction makes no assumption about the form of the correlation; see Rogers (1983) for details.

To apply Huber's formula for intrafamily correlation, we consider the family as the unit of observation and linear regression on individuals as an M-estimator. Linear regression is not the maximum likelihood estimator because individuals within a family have correlated responses. The result is that if we define:

$$R_1 = \underset{\text{families}}{\text{sum}} \quad \underset{\substack{\text{within}\\ \text{family}}}{(\text{sum } X_i r_i)'} \quad \underset{\substack{\text{within}\\ \text{family}}}{(\text{sum } X_i r_i)}$$

$$R_2 = (X'X)^{-1}$$

where X_1 stands for the matrix of observed data and r_i is the vector of residuals for family member i, then $R_2 R_1 R_2$ is an asymptotically consistent estimate of the covariance matrix of the regression parameters, regardless of the form of intrafamily correlation or heteroscedasticity.

This formula generalizes to allow for intertemporal correlation for individuals nested within families.

GENERALIZED LEAST SQUARES

For the estimates of annual visit and admission rates (Tables 6 and 7), we used a more parametric approach to intertemporal and intrafamily correlation in order to include information on full and partial years of participation. We assumed that the correlation in the use rates could be approximated by a nested variance components model,

$$y_{ift} = \mu + \nu_f + \epsilon_i + \eta_{ift} \tag{1}$$

where y = outcome (e.g., visits);
μ = constant (mean) to be estimated;
ν_f = unobserved family effect for family f;
ϵ_i = unobserved individual effect for individual i net of ν; and
η_{ift} = unobserved error terms for individual i in family f in period t net of ν and ϵ;

and

$$E(\nu) = E(\epsilon) = E(\eta) = 0$$

where $\nu \sim$ i.i.d across families,

$\epsilon \sim$ i.i.d across persons, and

$\eta \sim$ i.i.d across observations.

With T_i years of data, each person's average annual visit rate is

$$\overline{y_{if.}} = \mu + \nu_f + \epsilon_i + \overline{\eta_{if.}} = \mu + \theta_{if} \tag{2}$$

where $\overline{y_{if.}}$ is the individual mean for y and $\overline{\eta_{if.}}$ is the person's mean for the error component η_{ift}. The variance of the error term (θ_{if}) is

$$\begin{aligned} E(\theta_{if}\,\theta_{jk}) &= \sigma_\nu^2 + \sigma_\epsilon^2 + \sigma_\eta^2 / T_i && i = j \text{ and } f = k \\ &= \sigma_\nu^2 && i \neq j \text{ and } f = k \\ &= 0 && i \neq j \text{ and } f \neq k \end{aligned}$$

We estimated Eq. (2) by generalized least-squares (GLS) to obtain more efficient estimates than possible with ANOVA in the presence of intertemporal and intrafamily correlation and the heteroscedasticity of θ.

FOUR-EQUATION MODEL

We have used a four-equation model to estimate imputed expenditures on medical services, rather than the more common analysis of variance (ANOVA) or analysis of covariance (ANOCOVA) techniques. This model is the same one developed and described in Duan, Manning, Morris, and Newhouse (1982, 1983).[1] Our choice is dictated by three characteristics of the distribution of medical expenses. First, a large proportion of the participants will use no medical services during the year. Second, the distribution of expenses among users is very skewed. Third, the means and variances of medical expenses are quite different for users of any inpatient services and for users of only outpatient services.

These characteristics imply that ANOVA and ANOCOVA techniques will yield very imprecise (though unbiased) estimates of the effects of health insurance, age, and sex on the expenses for medical services, even for a fairly large sample size such as we have. To increase the precision of our expenditures estimates, we have included covariates for age and sex. As Duan, Manning, Morris, and Newhouse (1982, 1983) have shown, a model that exploits the characteristics of the medical expenditure distribution can provide more precise estimates.[2]

---

[1]Estimates based on Duan, Manning, Morris, and Newhouse (1982, 1983) can be found in Newhouse, Manning, Morris, et al. (1981, 1982).

[2]In Duan, Manning, Morris, and Newhouse (1983), we carried out a split-sample analysis as an empirical comparison of the predictive performance of models similar to the models used in this study. We compared these models in the analysis of medical expenditures for fee-for-service data, which have very similar distribution characteristics (e.g., some nonusers and highly skewed positive expenses) to the present data. That analysis was an application of the classical cross-validation technique. We randomly split the sample into two subsamples, estimated our models on the first subsample, then predicted the observations in the second subsample using the models fitted on the first subsample. To evaluate the predictive performance of the models, we compared the predictions and actual observations in the second subsample. If a technique (e.g., ANOCOVA) is inefficient or if it overfits the data, it will not perform well on the second (prediction) subsample. The criteria that we used in the split-sample analysis were mean forecast squared error and mean forecast bias.

Both ANOVA and ANOCOVA performed relatively poorly in the split-sample analysis. The four-part model used in this study performed statistically significantly better than either ANOVA or ANOCOVA in terms of mean forecast squared error; the four-part model was insignificantly different from a two-part model (a probit for any use,

We have modeled the distribution of medical expenses by partitioning the enrolled population into three groups: nonusers, users of only outpatient services, and users of any inpatient services; the model examines the expenses of the last two groups of users separately.[3] The first equation of the model is a probit equation for the probability that a person will receive any medical service during a year—from either outpatient or inpatient sources. Thus, this equation separates users from nonusers and therefore addresses the first characteristic described above (a large proportion of the population makes no use of 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 any medical use. This equation separates these two user groups, and thus addresses the third characteristic noted above (different means and variances 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 the total annual medical expenses for users of any inpatient services. Note that this last equation includes *both* outpatient and inpatient expenses for users of any inpatient services.

The log transformation of annual expenses practically eliminates the undesirable skewness in the distribution of expenses among users, earlier described as the second characteristic. The log transformation yields nearly symmetric and roughly normal error distributions, for which the least squares estimate is efficient. We therefore expect the estimates from this model to be more precise than those obtained from ANOVA and ANOCOVA. Because the residuals from the log level of use equations are only roughly normal, we use a consistent retransformation technique to obtain consistent estimates on the raw dollar scale (Duan, 1983). Because the errors are mildly heteroscedastic by plan, we use separate retransformation factors for each plan.

All inference statistics have been corrected for intrafamily and intertemporal correlation.

---

and a log of medical expenses if any positive use). However, the two-part model was rejected because it yields inconsistent plan comparisons; the two-part model underweights the inpatient response.

[3]In this analysis we did not use either the Heckman model or the Adjusted Tobit Model. Both models make assumptions about specifications which cannot be validated but which, nevertheless, have major implications for predictions. Second, unlike Heckman's application, our nonusers are not missing data cases but truly zero users. Finally, both the Heckman and Adjusted Tobit Models have multiple maxima and other statistical and numerical problems that the four-part model is free from. For a fuller discussion, see App. A in Duan, Manning, Morris, and Newhouse (1982), and Duan, Manning, Morris, and Newhouse (1984).

In addition to the techniques described above, we also used negative binomial regression models to estimate the effect of insurance plan on the admission and visit rates. Such an approach incorporates the information contained in the distribution of counts, namely the large proportion of zeros and skewness in the outcomes. Although the results were somewhat more precise than those reported in Table 5, our qualitative conclusions were not changed. Nor did incorporating age and sex covariates alter the conclusions. For the negative binomial regression models, we used Rogers' (1983) adaptation of Huber's (1967) approach to correct for intrafamily correlation. The negative binomial approach corrects directly for intertemporal correlation.

REFERENCES

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.

Brook, R. H., J. E. Ware, A. Davies-Avery, A. L. Stewart, et al., "Overview of Adult Health Status Measures Fielded in Rand's Health Insurance Study," *Medical Care,* 1979; 17(7), Supplement: 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,* 1983; 309:1426–1434.

California Medical Association, Committee on Relative Value Studies, *1974 Revision of the 1969 California Relative Value Studies,* Sutter Publications, Inc., San Francisco, 1975.

Cave, J.A.K., *Equilibrium in Insurance Markets with Incomplete Information: Adverse Selection under Asymmetric Information,* The Rand Corporation, R-3015-HHS, 1984.

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

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, 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 Business and Economic Statistics,* 1983; 1(2):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,* 1984; 2(3):283–289.

Eggers, P., and R. Prihoda, "Pre-enrollment Reimbursement Patterns of Medicare Beneficiaries Enrolled in 'At-Risk' HMOs," *Health Care Financing Review,* 1982; 4(1):55–73.

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

Enthoven, A., *Health Plan,* Addison-Wesley, Reading, Mass., 1980, Chap. 5.

Gaus, C. R., B. S. Cooper, and C. G. Hirschman, "Contrasts in HMO and Fee-for-Service Performance," *Social Security Bulletin*, 1976; 39(5):3–14.

Goldberg, G. A., *The Health Insurance Experiment's Guidelines for Abstracting Health Services Rendered by Group Health Cooperative of Puget Sound*, The Rand Corporation, N-1948-HHS, 1983.

Huber, P. J., "The Behavior of Maximum Likelihood Estimates under Nonstandard Condition," Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967; I:221–233.

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.

Luft, H. S., "Assessing the Evidence on HMO Performance," *Milbank Memorial Fund Quarterly*, 1980; 58:501–536.

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

Luft, H. S., "How do Health-Maintenance Organizations Achieve their 'Savings'?: Rhetoric and Evidence," *New England Journal of Medicine*, 1978; 298:1336–1343.

MacColl, W. A., *Group Practice and Prepayment of Medical Care*, Public Affairs Press, Washington, D.C., 1966.

Manning, W. G., J. P. Newhouse, and J. E. Ware, "The Status of Health in Demand Estimation: Beyond Excellent, Good, Fair, or Poor," in V. R. Fuchs (ed.), *Economic Aspects of Health*, University of Chicago, 1982:143–184.

Manning, W. G., A. Leibowitz, G. A. Goldberg, and J. P. Newhouse, "HMOs Versus Fee-for-Service: Incentives or Selection," The Rand Corporation, unpublished working paper, 1984.

McClure, W., "Implementing a Competitive Medical Care System Through Public Policy," *Journal of Health Policy and Law*, 1982; 7:2–44.

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

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*, 1981; 305:1501–1507; also published by The Rand Corporation, R-2847-HHS, 1982.

Nobrega, F. T., I. Krishan, R. K. Smoldt, et al., "Hospital Use in a Fee-for-Service System," *Journal of the American Medical Association*, 1982; 247:806–810.

Perkoff, G. T., L. Kahn, and P. J. Haas, "The Effects of an Experimental Prepaid Group Practice on Medical Care Utilization and Cost," *Medical Care,* 1976; 14:432–449.

Rogers, W. H., "Analyzing Complex Survey Data," The Rand Corporation, unpublished working paper, 1983.

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

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

Various authors, *Measurement of Physiologic Health for Children* (series), The Rand Corporation, R-2898-HHS, various dates.

Waldo, D. R., and R. M. Gibson, "National Health Expenditures, 1981," *Health Care Financing Review*, Rev. 1982; 4(1):1–35.

RAND/R-3029-HHS