RISK SELECTION IN THE MASSACHUSETTS STATE EMPLOYEE HEALTH INSURANCE PROGRAM

Running Title: Risk Selection in Health Insurance

Wei Yu, Ph.D.[1]

Randall P. Ellis, Ph.D.[2]

Arlene Ash, Ph.D.[3]

[1]Health Economics Resource Center and Cooperative Studies Program, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA

[2]Department of Economics, Boston University, Boston, MA.

[3]Health Care Research Unit, Section of General Internal Medicine, Evans Department of Medicine, Boston Medical Center, Boston, MA.

Address Correspondence and Reprints to:

Wei Yu, Ph.D.

Health Economics Resource Center

VA Palo Alto Health Care System

795 Willow Road (152 MPD)

Menlo Park, CA 94025

Telephone: (650) 493-5000 x23157

Fax: (650) 617-2639, E-Mail: wei.yu2@med.va.gov

RISK SELECTION IN THE MASSACHUSETTS STATE EMPLOYEE HEALTH INSURANCE PROGRAM

Running Title: Risk Selection in Health Insurance

ABSTRACT

Using the Diagnostic Cost Group (DCG) model developed from a national sample, we examine biased selection among one fee-for-service (FFS) plan, one preferred provider organization, and several health maintenance organizations (HMOs) in Massachusetts. The proportions of enrollees in low-risk groups are higher in the HMO plans and lower in the FFS plan. The average age in the FFS plan is 9 years greater than that in the HMO plans. Actual premiums are not consistent with risk levels among HMO plans, resulting in gains in some HMO plans and losses in others as high as 20% compared to expected expenses as computed by the DCG model.

INTRODUCTION

Capitated payment has been widely used in managed care plans to contain healthcare costs. Risk adjustment of the capitation payment rates is a frequently recommended strategy for dealing with variations in health status of capitated enrollees. If rates are not risk adjusted, health plans have incentives to try to attract the healthy and avoid the sick. These incentives can lead to biased selection, in which certain plans enroll disproportionately healthy people and other health plans, in particular indemnity plans, are left with the disproportionately sick [1-3]. Although risk adjusted payments have received prominent attention as a strategy for reducing biased selection both domestically and internationally [4-7], it is not clear, in practice, whether available rate-setting methods offset the problem.

In this study we used the Diagnostic Cost Group (DCG) model [8-12] to examine biased selection and explore the relationship between expected costs and premiums among plans in a state employee insurance program. The model was developed from a national private insurance sample with nearly 1.4 million people. We measured risk selection and predicted cost differences using data from the Massachusetts State employee health insurance program during fiscal years 1994 and 1995. This program offered a traditional indemnity plan (FFS), one Preferred Provider Organization (PPO) and several Health Maintenance Organization (HMO) plans. Payments to plans were negotiated in the absence of measures of plan differences in their enrollees' health status. We first compared plans with respect to demographics, medical payments, and DCG-measured risk. Then we compared plans' relative risks with relative premiums, and calculated risk-adjusted premiums to reveal potential gains and losses.

Our results document some of the strongest evidence of biased selection yet found, a difference of 56 percent in expected costs between the traditional indemnity and HMO plans, even though benefit coverage features are similar. Though premiums set by the Massachusetts Group Insurance Commission (GIC) reflect the relatively higher risk of the indemnity plan, they

are not consistent with relative risks among HMO plans. The consistency between the premium and the relative risk in the FFS plan reflect a common consensus about selection effects between indemnity and managed care plans. The inconsistency between the premiums and the relative risks in the managed care plans needs further examination. It may be partially due to plan differences in the ability to capture the diagnostic codes used to measure risk. It may also suggest an unstable market for managed care, where health plan exit or entry seems likely. This study highlights the usefulness of risk adjustment for identifying, and potentially compensating, plans for differences in their enrollees' health status.

METHODS AND DATA

Data

We obtained data from the GIC, including claims and eligibility status of state employees, state-employee retirees, and dependents in Massachusetts for two fiscal years: year 1 spanned July 1, 1993 through June 30, 1994 and year 2 was from July 1, 1994 through June 30, 1995. At that time, Massachusetts state employees under age 65 could choose among three types of plans: fee-for-service (FFS) traditional indemnity coverage, a Preferred Provider Organization (PPO), and a choice of several Health Maintenance Organizations (HMOs). The FFS indemnity plan offered in this program was very generous, with only a $75 individual deductible ($150 for a family) and a flat $5 fee per office visit after the deductible, up until a $750 stop-loss (both individual and family). The PPO plan features were identical to the FFS plan, although deductibles and stop-losses were higher for out-of-plan use ($150 deductible, $3000 stop-losses, individual and family). All HMOs offered by the GIC were required to charge a flat fee of $10 per visit at this time. Because cost sharing was so low and covered services were standardized across plans, covered expenses provided a picture of spending that was probably only very slightly distorted by cost sharing differences across plans.

Our method required both an "enrollment" and a "claims" file. The enrollment file

contained the reason, nature and type of coverage for each person who was entitled to health care benefits. Each record also included the enrollee's age, sex, and months of eligibility. The claims (or encounter) file identified the recipient of each service and contained diagnostic, treatment and payment information for each medical encounter.

Approximately 270,000 people were eligible at some time over the two-year study period. The enrollees were distributed across 12 health plans: one indemnity plan, one PPO, and 10 HMOs. One HMO plan did not report any diagnostic information in year 1 and three other HMOs had very incomplete cost information (the apparent average cost per person was only half of that for other HMOs). Hence, we excluded data from these four HMOs from our analysis. Since most people above 65 had alternative insurance coverage (such as Medicare), we excluded people who were 65 or older on June 30, 1995 and people covered by other senior insurance programs. We also excluded people with missing age or sex, and dependents who were students over age 24. The analytic file consisted of one FFS plan, one PPO plan and six HMO plans. It captured the experience of 159,936 people who were eligible at least one month in each of the two study years.

During the study period, premium rates were set by the GIC based on bids from each plan. The GIC evaluated these bids in light of the demographic composition of each plan's enrollees, establishing separate premiums for families and individuals. No measure of plan risk due to illness burden was used in rate setting. We obtained premium rates for individuals and families for each plan in FY 1994 from the GIC.

The Commission had been closely monitoring these data for some years, most importantly, enforcing uniform reporting of International Classification of Diseases codes, 9th revision and clinical modification (ICD-9-CM) on dummy claims. By 1994, most plans had achieved similar levels of diagnostic coding, as measured by the percent of physician records with at least one diagnosis recorded.

Enrollment Information

Many existing payment models, including HCFA's Average Adjusted Per Capita Cost (AAPCC) methodology, used beneficiary age and sex (age-sex cells) as predictors. For this study, we calculated beneficiary age as of the first day of year 2. We also defined two additional enrollment-based variables, known to affect costs in the GIC data: "employee status" and "early retirement." The marker for employee status distinguished the primary contract holder (a state employee) from his or her dependents; the early retirement marker identified under-age-65 employees entitled to retirement benefits. We compared plans' enrollees by age, sex, mean numbers of eligible months, percent with no cost in year 2, reported and predicted cost, and distribution of levels of risk.

Constructing a Cost Variable

The total annual cost of medical services for a person was calculated by summing up all covered expenses during the year. The methods used to impute costs to individuals differed across the plans. For the FFS and PPO plans, the covered expense was the amount of submitted expenses that were eligible for payment by the plan, including any deductible, coinsurance, and coordination of benefits expenses. For the HMOs, costs were assigned to each encounter reported to the Commission following detailed instructions from the GIC, which reflected fee schedules and average cost pricing. Since these so-called "dummy claims" were not connected to monetary transfers, we used the language of "reported" rather than "actual" costs throughout this paper. We did not truncate our spending measure, since GIC plan payments did not reflect any stop-loss or truncation provisions, and our risk adjustment model was also for untruncated spending. Pharmacy bills were not included when calculating expenditures for any of the plans, since not all plans provided pharmacy coverage.

Partial-year eligibility presented problems for risk adjustment. When a person was only eligible for part of year 1, we might miss some diagnostic information that could be used for

prospective risk adjustment. For people who were only eligible for part of year 2, total observed costs were not directly comparable to costs for people who were present for the whole year. In the GIC study over 90% of its participants remained enrolled for the entire year; the mean number of eligible months was 11.8 for each plan. Analysis using a different sample found that correcting for this eligibility variation in the regression model resulted in very minor differences in regression results and predictions when the mean number of eligible months for a plan was more than 11 [13]. Hence we did not make any adjustment for partial year eligibility in the base year for this analysis, and used the basic DCG model as was. Individuals were included in the analytic file as long as they were eligible for at least one month in each of year 1 and year 2.

We accommodated the problem of partial eligibility in year 2 by using a weighting algorithm for people who were present in year 1 but only eligible for part of year 2. To get the correct average payments for all beneficiaries, including those who left the plan during year 2, we annualized payments and weighted observations, as follows. Expenditures during year 2 were divided by the fraction of year 2 (in increments of one month) each beneficiary was present, which annualized total payments. For example, if a man remained alive and enrolled in a plan only into the sixth month of year 2 and generated $6,000 of reimbursements, then his annualized payment would be $12,000. If annualized amounts were simply entered into regressions and calculations of means, this overstated the contribution of such partial-year eligibility. Therefore, we weighted each person's annualized cost by the fraction of the year that he or she was eligible for coverage. The weight for the man in the above example was 0.5. This process of annualizing and weighting observations resulted in unbiased estimates of the average and total payments for a group in which individuals were eligible for different fractions of the year.

Predicting Cost with the DCG Risk-Adjustment Method

We used the DCG Version 3 model for commercially insured populations, also known as the DCG Hierarchical Condition Categories (DCG/HCC) model, as described in Ash et al. [8]. This model used age, sex and diagnostic information (ICD-9-CM codes) from one year to predict annual health care spending for the subsequent year. It identified the range of medical problems present and used this information to predict higher costs for plans that enrolled more people with serious, chronic conditions. The model's cost prediction for a person was the same whether a particular illness was coded once or many times during year 1 and did not increase simply because more resources were expended. Hierarchies (the "H" in HCC) were used to retain only the most serious among related medical conditions. For example, for a man seen only for a cough and congestive heart failure (CHF), the model retained both; if the same man was also seen for chronic obstruction pulmonary disease (COPD) and hypertension, only the "dominant" conditions (COPD and CHF) were retained. If a distinct coexisting condition, such as diabetes, was also present, the model also took into account the seriousness of this additional problem in calculating predicted year 2 costs.

Modeling

The prospective DCG/HCC model was developed using a nationally representative, private insurance data set with 1.4 million people, and estimated using weighted least squares. We used the nationally estimated model to predict cost for our Massachusetts State population in two steps. First, we entered age, sex, and diagnosis information from our sample into the national model to obtain a DCG/HCC expected cost for year 2 for each person. Second, we recalibrated these predictions to the GIC sample by regressing year 2 costs on 3 variables: dummy markers for "early retirement" and "employee status" and the DCG/HCC prediction (or risk score). This 3 degree-of-freedom (df) model captured substantial complexity because the risk score was itself a function of age, sex and the diagnostic profile. To illustrate the explanatory power of the risk score in predicting cost, we also estimated a 17 df demographic model with 16

age-sex cells plus "employee status" and "early retirement." The key difference between the demographic and DCG models was the risk score generated by the national DCG/HCC model. We reported the R^2, or percent of total variation explained, for each model.

Risk Determination

We classified enrollees into 11 risk groups by level of DCG-predicted cost. Because about half of the study population had medical expenses below $1,000 (38% for the FFS plan, 52% for the PPO, and 60% for HMO plans), at the low end we used risk interval widths of $250. The wider intervals used for those with higher predicted risk were needed to ensure reasonable numbers in each category. At the top, we grouped all people with predicted cost of $5,000 or more. For some comparisons, we combined the eight plans into three categories: FFS, PPO, and HMOs.

Relative Risk

We compared plans on their relative risks and relative premiums. The relative risk for a plan was the ratio of its DCG/HCC-predicted mean cost to the mean for the entire study population. The relative premium for a plan was the ratio of its actual premium to the average premium for the study population, which reflected the adjustment of resource allocation by the GIC through the negotiation process. In an economically efficient system, premiums would reflect differences in health status (relative risk) among plans. We compared relative premiums with relative risks to examine the variation consistency between the actual resource allocation and the DCG predicted risk. We calculated relative premiums separately for individuals and families.

Gain and Loss after DCG Risk Adjustment

To examine potential gains and losses among the eight insurance plans when premium rates were risk-adjusted by the DCG/HCC model, we calculated percentage changes in payments. A risk-adjusted premium for a plan was calculated by multiplying the average

premium of the entire study population by the plan's relative risk. We then calculated percentage changes between the risk-adjusted and the actual premiums.

Statistical Significance

All statistical analyses used SAS statistical software, version 6.11 (SAS Institute, Cary, North Carolina). Using ANOVA and ANOVA with option DUNCAN, we tested differences in mean predicted payments among the three types of health insurance plans. Because of the large sample size, nearly all reported differences were statistically significant at the $p<0.01$ level.

RESULTS

Data Completeness

Percentages of diagnostic and procedure codes reported by each plan were summarized in Table 1. For inpatient claims, all plans had relatively complete diagnostic information. The FFS, the PPO plan, and 4 of the 6 HMO plans reported principal diagnostic codes for every inpatient stay. The other two HMO plans reported diagnostic information on 99% and 95% of inpatient records. Most plans also reported some secondary diagnostic codes, although these rates varied by plan and in part reflected the structure of claims and encounter files, such as whether multiple diagnoses could appear on the same or separate "claims" and "line items". Half of all plans, including the FFS and the PPO, recorded diagnoses on 95% of outpatient claims. Four HMO plans, however, reported diagnoses less frequently on outpatient records, ranging from 57% to 75%. Overall, all plans in the analytic file had good diagnostic information (ICD-9-CM codes) for inpatient services and reasonably good information for outpatient services.

Insert Table 1 here

Selection Effects

Selection effects in the three types of health insurance plans were clearly indicated by differences in average age and cost (Table 2). People enrolled in the FFS plan were generally

older and had much higher medical costs than those in other plans. The average age of people enrolled in the FFS plan was about 9 years older than in HMO plans and the average cost in the FFS plan was more than twice that in the HMOs. When dividing enrollees into risk groups, the HMO and PPO plans enrolled more people in the low-risk groups (< $1,500) and fewer in the high-risk groups (Figure 1). For example, the group with the highest predicted cost contained 6% of the total enrollees in the FFS plan, 5% of PPO plan enrollees, and just 2% of HMO plan enrollees.

Insert Table 2 and Figure 1 here

DCG Risk Adjustment

The demographic model explained only 1.9% of the variation in cost in the GIC data, while the DCG/HCC model explained 9.3%. Relative risks were compared with relative premiums to examine differences between the actual resource allocation and the risk of each plan estimated by the DCG model (see Figure 2). Though the FFS plan enrolled relatively sicker people (risk is 22% higher than average), the premiums closely reflected the risk difference (premium is 20% higher than average). Relative risk for the PPO plan was 4% higher than average, while its individual premium was 1% lower than average and its family premium was 1% higher. Among HMO plans, the data showed surprisingly large variations in consistency between relative premiums and relative risks. Four HMO plans were not paid as much as their relative risk and two HMO plans were paid considerably more. At least part of the variation inconsistencies between relative premiums and relative risks among HMO plans were probably due to incomplete diagnostic codes.

Insert Figure 2 here

DISCUSSION

This study has documented enormously biased selection among Massachusetts State health insurance plans. On average, the PPO and HMO enrollees were 8 and 9 years younger

than those enrolled in the indemnity plan. The reported medical cost of the managed care plans averaged about half of that of the indemnity plan. These differences remained despite the fact that we excluded enrollees over age 65, who were disproportionately enrolled in the FFS plan from this analysis. Their inclusion would reveal even greater differences.

Since operating efficiency might contribute to the substantial cost difference between the managed care plans (HMOs and PPO) and the indemnity plan, we examined enrollment distributions among 11 risk groups. The distribution showed that HMO plans and the PPO enrolled higher proportions of people in the low risk groups (≤ $1,500) than the indemnity plan (Figure 1). This distribution was also consistent with the difference in management intensity of the three types of plans. The pattern suggested that the most strictly managed plans (HMOs) enrolled the highest proportion of people in low risk groups (≤ $1,500) and the least-controlled plan (FFS) enrolled the lowest proportion. The PPO plan was less strictly managed than the HMO plans. Patients in the PPO did not need to get a referral to see a specialist or another physician. The risk distribution in the PPO was between that of the HMO plans and the indemnity plan (Figure 1).

Our data could not distinguish whether the observed selection was due to consumer choice or management strategy. Although employees chose plans, plans might follow strategies that differentially attracted healthy people. None of the HMOs was "for-profit" during the study period. Our study suggested that selection effects could be serious even among non-profit HMOs.

Using the DCG method, we found that a substantial portion of the cost differences among plans could be explained by differences in demographics and health status. The remainder might be due to differences in pricing or utilization control efforts. Based on its relative risk, the average predicted cost for the indemnity plan was $2,499, which was $550 lower than its average reported cost (Table 2). At least part of this $550 was likely due to consumer behavior when

health care services were not strictly managed. In contrast, risk-adjusted predicted cost for all people in the managed care plans was $1,761 while the average reported cost was $1,403 (Table 2). The $358 saving might be due to utilization management or to unobservable factors. For example, people who chose an HMO might prefer more conservative treatment than those choosing an indemnity plan. It was also possible that our model had not fully measured how much healthier the HMO enrollees truly were.

The Group Insurance Commission (GIC) managed the Massachusetts State health insurance program. At the time of this study, payments to plans (premiums) were set by the GIC, based upon bids from each plan. If health plan payments did not adjust for risk, plans had powerful incentives to divert money that could be used to improve health care to marketing to attract healthy people. Non-risk-adjustment payment method created competition among plans, but could reward favorable risk selection as much or more than efficient health care delivered.

The FFS plan enrolled relatively sicker people and its additional risk was reflected in its higher premium. Large risk variation, however, existed among HMO plans, with the premiums not reflecting differences in risk. If premium rates were risk-adjusted using the DCG/HCC model, two HMO plans would gain and four others would lose (Figure 3). Thus, although the HMOs overall were overpaid, some HMOs accepted lower payments than appeared to be justified by their risk. Competitive bidding might contribute to the variation since the reported expenses of "riskier" HMOs 2 and 3 (averaging about $1,300 each) were, in fact, higher than for the lower risk enrollees in HMOs 5 and 6 (averaging $1,200 and $1,100, respectively), although the premiums "did not track." (HMO3 had the lowest individual premium and HMO6 the highest family rate.). This study could not tell whether HMO3 was more efficient, because important factors such as quality of service were not measured.

Insert Figure 3 here

Incomplete diagnostic reporting (Table 1) or missing encounter records might also contribute to observed variations in either risk or cost. However, most of the missing diagnosis codes were from outpatient care and plans differ widely in the completeness of such data. To compare plans absent the effects from different levels of missing outpatient data, we fit a DCG/HCC model using only inpatient diagnosis codes. Although less than 5% of the population was hospitalized during year 1, this inpatient model still explained 8% (versus 9.3% for the all-diagnosis model) of variation in reported expenses.

The DCG/HCC model used for this study was developed for HCFA to adjust its capitated payment rates to Medicare managed care risk plans [8]. It was calibrated to a national private insurance data set in which most (85%) of the enrollees were covered by indemnity plans. This raised concerns about using this model for HMO or PPO plans whose treatment strategies might differ. However, the DCG/HCC model explained cost differences in the current file at least as well as in the original benchmark sample (R^2 = .094 here versus .084 in the benchmark validation sample). This suggested that the relationship between diagnostic profiles and expenditures in these Massachusetts data, which included many managed care enrollees, was similar to that of national FFS plans.

REFERENCES

1. R.S. Brown, D.G. Clement, J.W. Hill, S.M. Retchin, and J.W. Bergeron, Do health maintenance organizations work for Medicare, Health Care Financing Review 15 (1993) 7-23.

2. J.P. Newhouse, Reimbursing health plans and health providers: selection versus efficiency in production, Journal of Economic Literature 34 (1996) 1236 - 1263.

3. D. Altman, D. M. Cutler, and R. J. Zeckhauser, Adverse selection and adverse retention, American Economic Review, Papers and Proceedings 88 (May 1998) 122-126.

4. J.P. Newhouse, M. Beeuwkes Buntin, J.D. Chapman, Risk adjustment and Medicare: taking a closer look, Health Affairs 16 (1997) 26 - 43.

5. N. Rice and P.C. Smith, Capitation and risk adjustment in health care, Health Care Management Science 3 (2000) 73 - 75.

6. S. Peocock and L. Segal, Capitation funding in Australia: imperatives and impediments, Health Care Management Science 3 (2000) 77 - 88.

7. K. Beck, Growing importance of capitation in Switzerland, Health Care Management Science 3 (2000) 111 - 119.

8. A.S. Ash, R.P. Ellis, G.C. Pope, J.Z. Ayanian, D.W. Bates, H. Burstin, L.I. Iezzoni, E. MacKay, and W. Yu, Using diagnoses to describe populations and predict costs, Health Care Financing Review 21(3) (Spring, 2000) 7 - 28.

9. R.P. Ellis, G.C. Pope, L.I. Iezzoni, et al., Diagnosis-based risk adjustment for Medicare capitation payments, Health Care Financing Review 17 (1996) 101 - 128.

10. A. Ash, F. Porell, L. Gruenberg, et al., Adjusting Medicare capitation payments using prior hospitalization data, Health Care Financing Review 10 (1989) 17 - 29.

11. R.P. Ellis, Employee choice of health insurance, Review of Economics and Statistics 71 (1989) 215 - 223.

12. R.P. Ellis and A. Ash, Refinements to the diagnostic cost group (DCG) model, Inquiry 32 (1995) 418 - 429.

13. A. Ash, R. P. Ellis, W. Yu, et al. Risk Adjustment for the Non-Elderly. Research Report prepared for Health Care Financing Administration, June 1998.

Table 1. Diagnostic Coding Frequencies by Plan in 1994

Plan	*Inpatient* Percent of claims with: First Diagnosis	Second Diagnosis	Procedure Code	*Outpatient* Percent of claims with: First Diagnosis	Second Diagnosis	Procedure Code
FFS	100 %	0 %	70 %	95 %	0 %	68 %
PPO	100	45	81	97	5	68
HMOs:						
1	100	30	80	57	5	52
2	100	54	34	69	11	54
3	100	62	37	100	23	87
4	100	64	80	100	20	97
5	99	0	85	70	0	64
6	95	7	22	75	9	46

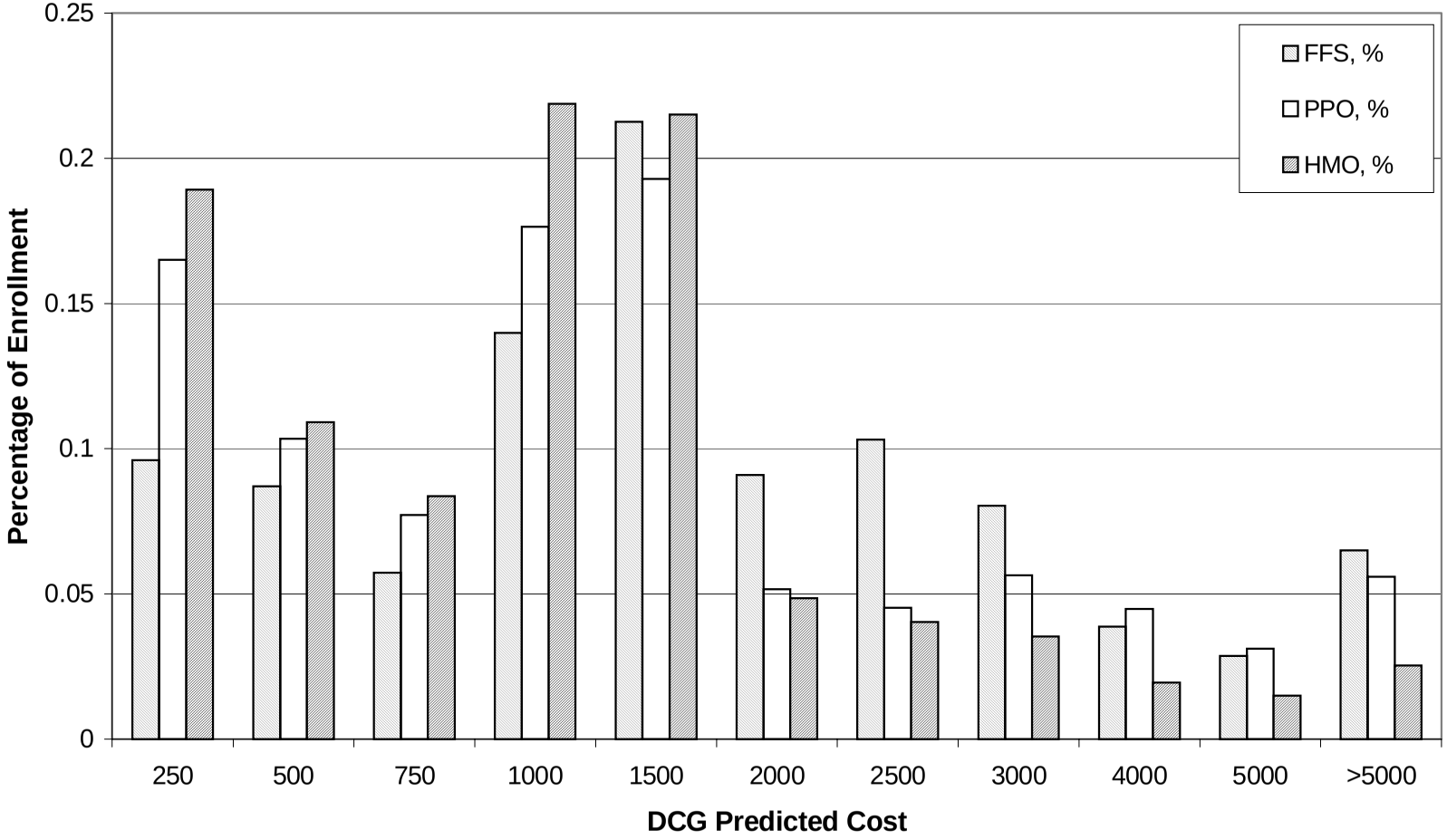
Note: Numbers shown are the percentages of hospital and physician claims with the indicated information.

Table 2. Year 2 Demographic and Cost Characteristics by Plan Type

Plan Type	N	Mean age	Mean Months Eligible	Percent with Zero Expenses	Actual Expenses		DCG-Predicted Expenses	
					Mean	CV	Mean	CV
Total	159,936	34	11.78	17	$ 2,052	412	$ 2,052	126
FFS	62,976	40	11.79	16	$ 3,049	361	$ 2,499	118
PPO	21,991	32	11.80	18	$ 1,572	415	$ 1,989	132
HMO*	74,969	31	11.78	17	$ 1,352	450	$ 1,523	134

*Include 6 HMO plans.

Figure 1. Distribution of Enrollment by DCG Predicted Cost and Plan Type



Note: Numbers on the X-axis are upper bounds to intervals, E.g., "250" contains all persons whose DCG-predicted costs lie in the interval from $0 to $250.

Figure 2. Percentage Deviations from Average of Relative Risks and Premiums by Plan

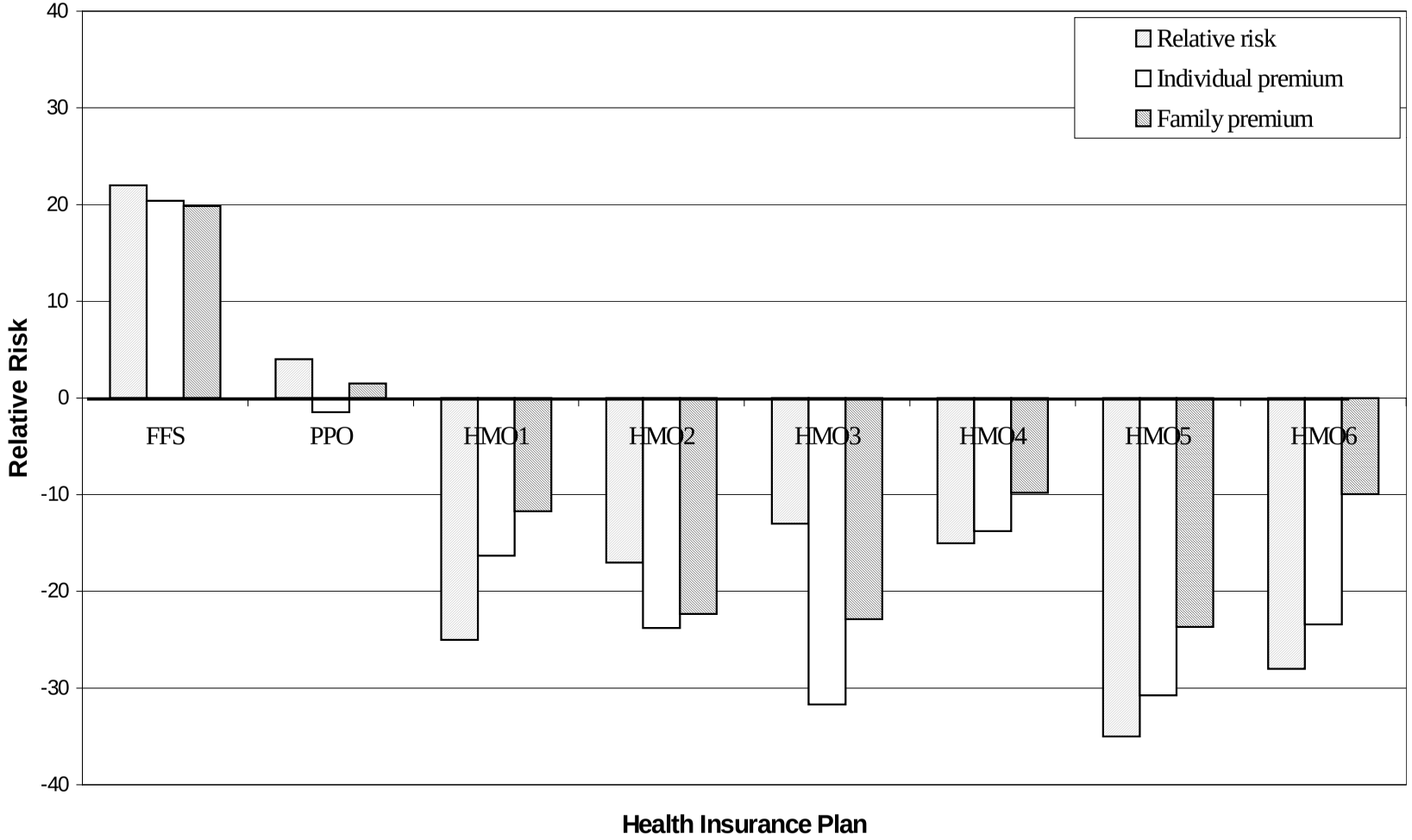


Figure 3. Percentage Change in Premiums with Risk Adjustment by Plan