Volume Title: Inquiries in the Economics of Aging

Volume Author/Editor: David A. Wise, editor

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-90303-6

Volume URL: http://www.nber.org/books/wise98-2

Publication Date: January 1998

Chapter Title: Insurance or Self-Insurance? Variation, Persistence, and Individual Health Accounts

Chapter Author: Matthew Eichner, Mark B. McClellan, David A. Wise

Chapter URL: http://www.nber.org/chapters/c7079

Chapter pages in book: (p. 19 - 49)

1 Insurance or Self-Insurance? Variation, Persistence, and Individual Health Accounts

Matthew J. Eichner, Mark B. McClellan, and David A. Wise

Economists have for some time emphasized the desirable incentive properties of catastrophic health insurance. Under such a system individuals would pay for their own health care unless the expenses were very large. Thus the temptation to spend too much, the "moral hazard," that is created by typical insurance provisions would be reduced or eliminated. Indeed, Arrow (1963) demonstrated that in the presence of moral hazard and risk aversion a catastrophic plan is optimal. A practical argument against catastrophic insurance, however, is that a very large fraction of families have almost no liquid savings and would find it hard to make even small out-of-pocket payments, especially if they were not anticipated. Perhaps in recognition of this fact, employees have shown a willingness to pay very high premiums to avoid uncertainty about health expenditures, and many firms have experienced very low participation rates in "major risk" plans with substantial deductibles even though these plans are typically offered at generously low premiums.

The attention of American health reformers in recent years has turned to managed care plans, which place little reliance on "demand" incentives, instead using "supply side" and contractual restrictions to limit spending under insurance plans with low deductibles and copayments. Apparently, such plans have had some success in limiting health expenditures, although at the cost of increased regulation of doctor-patient transactions. But, ex post, insured pa-

Matthew J. Eichner is assistant professor of finance and economics at Columbia University Graduate School of Business and a faculty research fellow of the National Bureau of Economic Research. Mark B. McClellan is professor of economics and medicine at Stanford University and a faculty research fellow of the National Bureau of Economic Research. David A. Wise is the John F. Stambaugh Professor of Political Economy at the John F. Kennedy School of Government, Harvard University, and director of the Health and Retirement Research Programs at the National Bureau of Economic Research.

Research support from the National Institutes of Health through a grant to the National Bureau of Economic Research is gratefully acknowledged.

tients bear only a small fraction of the cost of their care, suggesting that incentives for cost control are far from optimal. We begin with this paper a research agenda that emphasizes—and maybe reconsiders—price incentives in health insurance reform.

We explore the feasibility of catastrophic health insurance established in conjunction with individual health accounts (IHAs). Such an arrangement holds the potential for both reducing health care cost and encouraging saving. Under this plan, the employer establishes both a high-deductible health insurance plan and an IHA. Annual contributions to the IHA are equal to a substantial fraction of the deductible. Employee health care costs below the deductible are then paid out of the IHA; costs above the deductible are paid by the insurance plan. Assets remaining in the account when the employee retires, or becomes Medicare eligible, are then available for other purposes. The motivation for the parallel saving and insurance plans, of course, is that each employee is spending his or her "own" money for medical care, except in the event of serious illness. The plan thus combines the desirable features of catastrophic coverage for reducing medical expenditures, as advocated by Feldstein (1971) and by Feldstein and Gruber (1994), with a mechanism that creates a reserve from which individual expenses can be paid.

But even if the IHA component provides the necessary liquidity, it may still be thought to be inequitable. To the extent that individuals experience different health shocks over many years, the plan could lead to large differences in IHA accumulations. A person who is never sick will accumulate large IHA balances, while someone who is always sick will accumulate nothing. If individual health expenditures over a working lifetime vary little, all persons will have the same IHA balance at retirement. On the other hand, if average individual expenditures vary widely over the working life, the plan may look like a savings plan to the healthy and self-insurance to the chronically ill. Though mitigating measures could be appended to a basic IHA plan to limit such inequality, unequal accumulation is an unavoidable consequence of individual financial risk. Equal accumulations under an IHA plan can only be guaranteed if all individuals have the same health shock experience over their lifetimes. Of course, equal premiums under a comprehensive health plan with first-dollar coverage also assure equal individual cost.

Because individual health shocks clearly vary, the feasibility of an IHA plan depends on whether the gains (improved incentives for efficient health care spending and increased savings) outweigh the costs (more variation in individual health care costs than under more generous insurance plans). Indeed, we believe that in practice feasibility may depend largely on what the variation in IHA balances would "look like." Thus, as a crucial first step, we address that issue in this paper. Within the context of an illustrative IHA plan, we develop preliminary empirical evidence on the distribution of medical expenditures—and hence savings—under an IHA plan. Our analysis is based on longitudinal health insurance claims data from a large firm. In this analysis, we assume no behavioral response whatever to the increased cost sharing under an IHA plan. Feldstein and Gruber (1994) suggest that the response could be substantial. To the extent that individuals respond to price incentives, expenditure will be more equal—perhaps substantially more equal—than our results suggest. We hope to provide an upper bound on expenditure variation, as measured by variation in IHA accumulation.

We begin with examples that illustrate the enormous variation in health care costs in a single year and the persistence of individual expenditures from one year to the next. The illustrations are similar to those in McClellan and Wise (1994). That paper emphasized persistence in individual health care costs and demonstrated that persons with large expenditures in one year are likely to have large average expenditures, whether calculated over one or several years. In addition, we emphasize here that, although not insignificant, the relationship between expenditures two years apart is substantially weaker than the relationship between expenditures one year apart. And thus, while persistence is important, the descriptive data also suggest that high expenditure levels typically do not last for many years.

Next we explain the statistical model that underlies our conclusions on the distribution of health care expenditures. The goal is to approximate the distribution of medical expenditures over a working lifetime in a large firm. We have data on employee expenditures over a three-year period. We estimate a model that captures the pattern of expenditures among employees and then use the model to simulate the lifetime distribution of expenditures. We give particular attention to two issues: One is the extent of persistence, the expected expenditure in one year conditional on expenditure in prior years. The second is the "unexplained" residual variance, or "shock," in expenditure conditional on expenditure in prior years. An important aspect of the data is that this unexplained variance is very large and is not approximated well by any analytic distribution. Thus our simulation procedure depends heavily on nonparametric analysis based on the empirical distribution of conditional expenditures.

We next explain the results of simulations based on the model. We find that many employees will have no large medical expenditures over an entire working life. Others will have one or more episodes that generate large expenditures in one year and possibly in at least a few subsequent years as well. The concentration of expenditures that is observed in one year, and even when three years are combined, declines consistently as expenditures are cumulated over more years. Nonetheless, even over a working lifetime there is a noticeable concentration of expenditures. We illustrate the implications of the concentration by considering the distribution of balances at retirement in an IHA account with a $2,000 annual contribution. Although the vast majority of participants retire with substantial IHA balances, some incur substantial out-of-pocket costs and thus retire with only small balances. About 80 percent of employees are left with at least 50 percent of total IHA contributions, but about 5 percent have less than 20 percent.

We conclude with a discussion of the implications and limitations of this preliminary analysis. We comment on issues that are not addressed and on future research plans.

1.1 The Data and Summary Description

1.1.1 Medical Claims Data

The data are medical claims of employees in a large Fortune 500 manufacturing firm. The analysis is based on all fee-for-service insurance claims over the three-year period 1989–91. Over this period approximately 300,000 employees and their dependents were covered through these insurance plans.

The firm has two fee-for-service plans, one for hourly and another for salaried employees. The hourly plan, with benefits negotiated in union contracts, provides "first-dollar" coverage for virtually all health care. Because of this virtually unlimited coverage, hourly employees have no financial incentives to join managed care or HMO plans, though specific provider relationships and location considerations may provide some nonfinancial incentives. The salaried plan has an annual deductible of $200 per individual and $250 per family, a 20 percent coinsurance rate for all expenses, and an out-of-pocket annual limit (including the deductible) of $500 per family. Routine physical examinations are not covered. Both plans incorporate limited case management for certain high-cost medical conditions and concurrent review of hospital stays. The hourly plan includes preadmission certification requirements for certain elective admissions; patients who elect admission despite precertification denial are responsible for 20 percent copayments up to $750 per individual and $1,500 per family. Both plans also require second opinions for 16 elective surgical procedures, though the procedures are covered regardless of the second opinion finding. Both plans have very generous hospital stay limits: 365 days per stay, renewable after 60 days out of the hospital. Mental health and substance abuse inpatient care has a stricter day limit of 45 days, also renewable after 60 days out of the hospital. (During the time period covered by the data, a managed care program was implemented for mental health and substance abuse services.)

1.1.2 Summary Data

Many studies have documented that medical expenditures in a particular year are concentrated among a small proportion of the insured. Less evidence, and hence less attention, has been directed to the persistence of individual expenditures over longer time periods and to the relationship between persistence and concentration of expenditures. Together, both have important implications for insurance in general and in particular for the feasibility, incidence, and other consequences of insurance market reforms. Before presenting more formal analysis, we present descriptive evidence on these issues using longitudinal individual claims data for the period 1989–91. We consider first a tabular description of the relationship between expenditures in three successive years. We then consider the concentration of expenditures and present more detailed descriptions of persistence, considering expenditures in consecutive years conditional on the decile (or quintile) ranking of expenditures in the first year.

Medical Claims in Successive Years

The distributions of expenditures in 1990 and in 1991 conditional on 1989 expenditures are shown in table 1.1. Consider, for example, persons aged 18–35 in the first panel of the table. Conditional on no expenditure in 1989, 4 percent of persons have expenditures above $5,000 in 1990. In contrast, over 20 percent of persons with expenditures above $5,000 in 1989 have expenditures above $5,000 in 1990. Persistence appears to increase with age. Almost 30 percent of persons aged 46–55 who had expenditures greater than $5,000 in 1989 also had expenditures greater than $5,000 in 1990. In contrast, only 5 percent of the persons in this age group with no expenditures in 1989 had expenditures above $5,000 in 1990. If there were no persistence across years, only 8 percent of the individuals with expenditures greater than $5,000 would have such high expenditures again in 1990.

Comparison of the first panel of table 1.1 and the second—which shows data for 1991 conditional on spending in 1989—reveals that persistence diminishes with time. For example, in the 18–35 age group, about 15 percent of those with expenditures above $5,000 in 1989 had expenditures above $5,000 in 1991, compared to almost 21 percent in 1990. For the persons aged 46–55, about 22 percent of those with expenditures above $5,000 in 1989 had expenditures exceeding $5,000 in 1991, compared to almost 29 percent in 1990.

Individual Concentration of Expenditures

The relationship between persistence and concentration is shown in figure 1.1. The figure shows that in 1989 about 80 percent of cost was incurred by 10 percent of enrollees, roughly comparable to concentration results from other studies. Fifty percent of employees incurred virtually no cost. Figure 1.1 also shows the concentration of annual expenditures *averaged* over the two-year period 1989–90 and over the three-year period 1989–91. If there were no persistence in cost from one year to the next, costs averaged over several years would be much less concentrated among a few enrollees than costs in a single year. The curves show that, although concentration declines as the time period increases, even over three years a small proportion of employees incur an enormous fraction of health care costs. Averaged over three years, 10 percent of enrollees account for a full 65 percent of expenditures. If there were no persistence from year to year, then averaged over many years the cost accounted for by the highest cost decile would approach 10 percent. The more formal analy-

Table 1.1 **Percentage Distribution of 1990 and 1991 Expenditures, by 1989 Expenditure Interval and by Age**

Age	1989 Expenditure ($)	1990 Expenditure ($)				
		0	0–300	300–1,000	1,000–5,000	Above 5,000
0–17	0	43.9	37.25	11.42	5.57	1.86
	0–300	45.4	34.84	12.56	5.75	1.44
	300–1,000	29	32.32	22.97	12.19	3.52
	1,000–5,000	28.1	25.69	20.92	19.05	6.25
	Above 5,000	23.28	19.32	17.43	20.76	19.21
18–35	0	40.94	33.97	12.92	8.25	3.92
	0–300	51.62	26.57	11.36	7.27	3.19
	300–1,000	35.55	25.5	18.76	13.69	6.51
	1,000–5,000	32.9	20.93	17.12	19	10.05
	Above 5,000	29.17	17.36	14.76	17.92	20.79
36–45	0	40.27	33.06	13.6	8.65	4.41
	0–300	36.57	33.03	16.11	10.12	4.18
	300–1,000	23.93	26.83	24.07	18.25	6.92
	1,000–5,000	18.88	19.97	21.19	27.93	12.04
	Above 5,000	18.63	16.82	16.67	22.81	25.07
46–55	0	37.65	32.14	14.52	10.15	5.53
	0–300	30.99	34.29	17.9	11.61	5.21
	300–1,000	17.84	26.35	28.2	19.47	8.13
	1,000–5,000	15.59	19.55	22.41	28.76	13.7
	Above 5,000	14.41	15.73	16.88	24.38	28.6
56–65	0	31.14	34.73	15.64	10.89	7.6
	0–300	35.87	31.89	16.52	10.14	5.58
	300–1,000	22.62	24.54	26.68	17.36	8.8
	1,000–5,000	19.17	18.87	21.66	26.07	14.23
	Above 5,000	21.5	13.82	15.67	22.42	26.6
0–17	0	28.29	46.98	14.87	7.49	2.37
	0–300	51.94	29.34	11.47	5.67	1.58
	300–1,000	38.92	28.29	19.03	10.31	3.45
	1,000–5,000	39.44	23.35	17.52	14.47	5.22
	Above 5,000	38.65	18.18	17.43	14.11	11.64
18–35	0	28.57	39.13	15.95	10.81	5.53
	0–300	63.58	19.21	8.98	5.54	2.69
	300–1,000	52.13	18.86	14	10.26	4.75
	1,000–5,000	48.7	16.2	13.49	14.87	6.73
	Above 5,000	46.32	13.9	11.17	13.69	14.92
36–45	0	24.43	40.28	17.38	11.98	5.91
	0–300	40.43	29.4	15.88	9.83	4.46
	300–1,000	31.02	24.29	21.79	15.94	6.96
	1,000–5,000	26.95	18.88	20.37	23.77	10.03
	Above 5,000	27.19	16.51	15.15	20.3	20.86
46–55	0	23.23	37.97	17.87	13.13	7.79
	0–300	32.73	31.41	17.85	11.89	6.13

Table 1.1 (continued)

Age	1989 Expenditure ($)	1990 Expenditure ($) 0	0–300	300–1,000	1,000–5,000	Above 5,000
	300–1,000	21.86	24.75	26.06	18.54	8.79
	1,000–5,000	20.33	19.27	20.69	26.75	12.95
	Above 5,000	22.66	15.85	16.05	23.6	21.83
56–65	0	29.58	33.78	16.54	11.92	8.18
	0–300	44.32	25.3	14.75	9.7	5.93
	300–1,000	34.87	19.25	21.41	15.85	8.62
	1,000–5,000	31.96	15.95	18.28	21.31	12.5
	Above 5,000	38.09	13	12.46	18.33	18.13

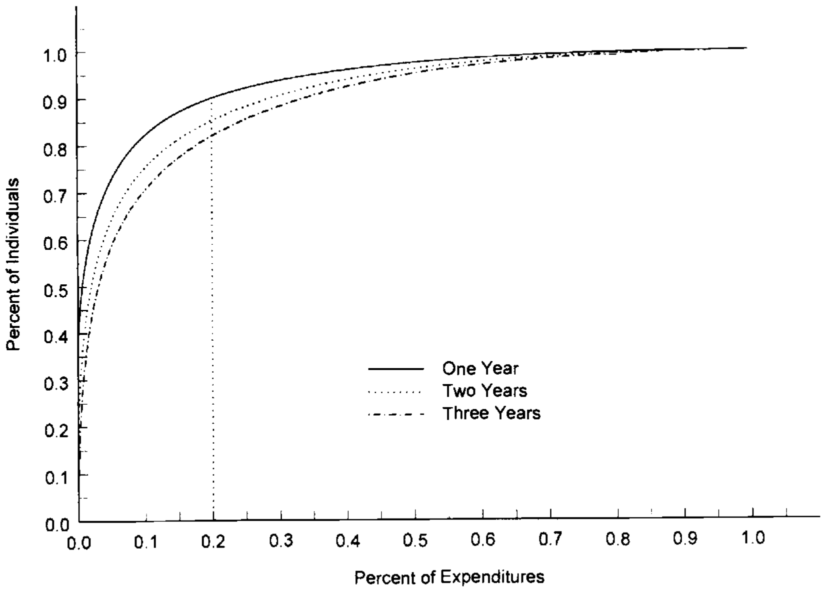

Fig. 1.1 Concentration of expenditures over one, two, and three years

sis presented below shows that concentration declines continuously as more and more years of expenditures are cumulated but is still substantial even averaged over a working lifetime.

Expenditure Decile and Subsequent Expenditures

Table 1.1 shows persistence in expenditures for selected age groups. Figure 1.2 presents a more detailed picture of persistence for all ages combined. Enrollees are divided into deciles based on 1989 claims. The figure shows expen-

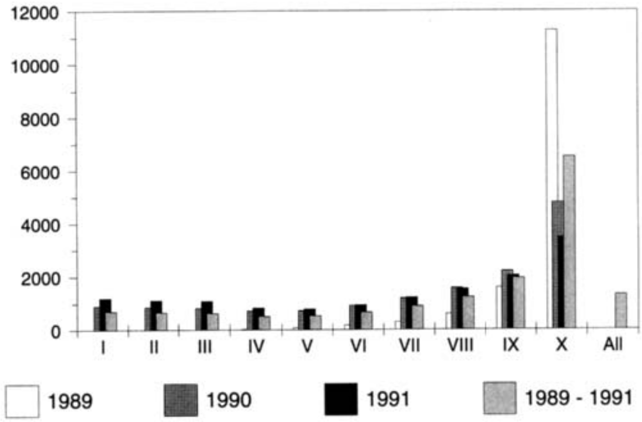

Fig. 1.2 Mean annual cost by 1989 decile

ditures in 1989 by decile and then, also by 1989 decile, average annual expenditures one year later in 1990 and two years later in 1991. For comparison, the figure also shows average expenditures for all enrollees over these three years, which was $1,314. Persons in the tenth decile in 1989 in that year spent over eight times as much as the average. They spent close to five times the average in 1990 and almost three times the average in 1991. Averaged over all three years, those in the highest decile in 1989 spent about five times the average.

An alternative description is shown in figures 1.3A, 1.3B, and 1.3C. Figure 1.3A shows the distribution of costs by quintile over the 1989–91 period for each 1989 quintile (determined by 1989 expenditures). The figure shows, for example, that almost 60 percent of persons who were in the highest quintile in 1989 were also in the highest quintile averaged over three years. Another 35 percent were in the second highest quintile averaged over three years. By contrast fewer than 10 percent of persons in the lowest 1989 quintile were in the highest quintile over three years, and only about 10 percent were in the second highest quintile. Figure 1.3B shows the distribution of costs in 1990, and figure 1.3C shows the distribution in 1991, conditional on the 1989 quintile. Over 40 percent of persons in the highest quintile in 1989 are in the highest quintile in 1990, and about 35 percent are in the highest quintile in 1991.

These descriptive data show that, on average, persons with high expenses in one year also tend to have much higher than average expenses in the next year and also higher than average, but lower, expenses in the following year. These data are used in more formal estimation to allow extrapolation of this persistence pattern over subsequent years. The formal analysis also shows that the importance of the persistence revealed in mean expenditures appears to be dominated by the enormous variation in expenditure shocks, conditional on any past expenditure pattern.

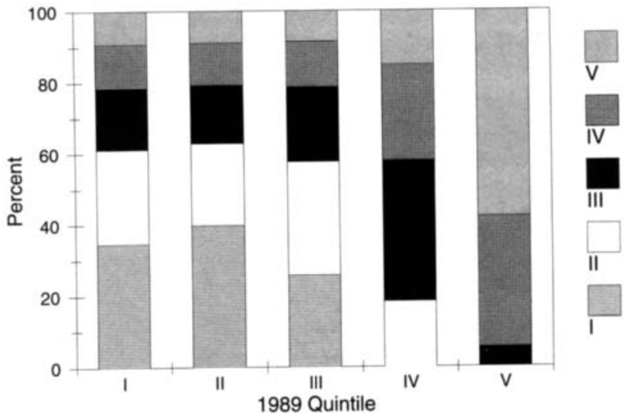

Fig. 1.3A 1989–91 Quintile by 1989 quintile

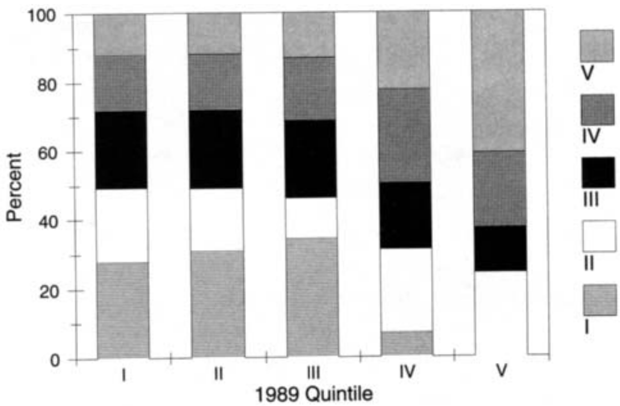

Fig. 1.3B 1990 Quintile by 1989 quintile

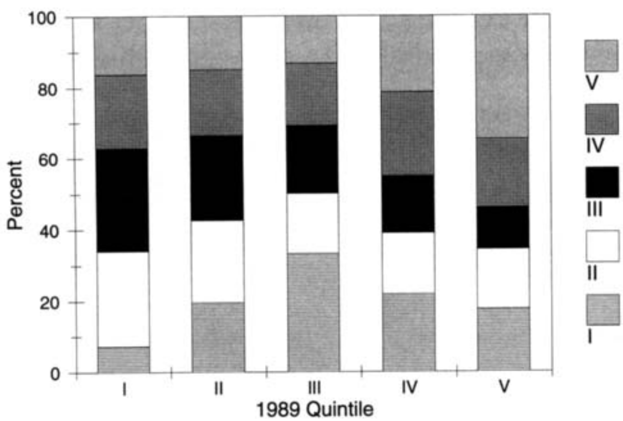

Fig. 1.3C 1991 Quintile by 1989 quintile

1.2 A "Model" of Persistence and Expenditure Shocks

Our goal is a formal description of medical expenditures that will allow us to simulate the pattern of expenditures over the working life. We begin with a description of the model specification and then explain the simulation procedure. A critical feature of the model is the extent to which it captures actual expenditure patterns, and thus we give considerable attention to the model fit.

1.2.1 Specification

The descriptive specification must capture two critical features of health care expenditures: (1) the enormous variation across individuals in the same year and (2) the persistence of expenditures from one year to the next. In this version of the analysis we describe annual expenditures. We assume that medical expenditures in year t, M_t, can be predicted by three factors: (1) demographic characteristics, denoted by D, which include age, sex, and employment status (hourly or salaried); (2) past expenditures M_{lag}, which in this version include expenditures in years $t - 1$ and $t - 2$; and (3) random shocks, ε:

(1) $$M_t = \alpha + \beta D + \gamma M_{\text{lag}} + \varepsilon .$$

The critical components are the random shocks and persistence (measured by γ). Because a large fraction of employees have no expenditures in a given year, it is useful to consider explicitly the expected value of M_t, given by

(2) $$\text{E}(M_t) = \Pr[M_t \leq 0] \times 0 + \Pr[M_t > 0] \times \text{E}(M_t \mid M_t > 0) .$$

We estimate the two components—$\Pr[M_t > 0]$ and $\text{E}(M_t \mid M_t > 0)$—separately. The first is estimated using a linear probability specification, and the second using a log linear regression.

The model estimates, together with exact specifications, are shown in appendix table 1A.1. The specification used in the analysis is presented in the bottom panel of the table. The basic structure of the specification, however, is more easily seen in the simpler specification that is shown in the top panel. This specification shows three variables (D1, D2, D12) that identify persons with no expenditures in $t - 1$, no expenditures in $t - 2$, and no expenditures in either prior period. Variables containing expenditure amounts for persons who had claims in prior years are defined using similar notation. For example, M1 gives expenditure amounts for persons who had claims in period $t - 1$ (for whom D1 = 0) and is zero for persons who did not file a claim in period $t - 1$. For persons with claims in both $t - 1$ and $t - 2$ (the "base" group), there are two expenditure variables, M*1 and M*2, for $t - 1$ and $t - 2$, respectively. The estimated coefficients are difficult to interpret individually. (E.g., the coefficient associated with D1 in the simplified expenditure equation [−0.7485] indicates that the expenditure in period t for persons with zero expenditures in both $t - 1$ and $t - 2$ is about 75 percent lower than the expenditures—*evaluated at* $M^*1 = M^*2 = 0$, the "intercept"—of persons with positive expenditures in both prior periods.) Thus we give scant attention to individual parameter estimates; instead we emphasize below the degree to which the specification reproduces actual expenditure patterns. The more flexible specification relaxes the simplified version in two ways: the lagged expenditure variables are piecewise linear and the lagged expenditure variables are interacted with age—distinguishing persons who are younger from those who are older than 45.

Possibly the most important component of the estimates is the large residual variance. Consider a given set of right-hand variables and the associated mean expenditure. The estimated standard error of the estimate (1.644) suggests that to capture, say, 95 percent of expenditures one would have to cover the range from 0.04 to 27 times the mean.

1.2.2 Prediction and Simulation Method

The key to prediction is the distribution of random shocks. We want the distribution that is used in prediction to "match" as closely as possible the actual distribution, which is extremely skewed. Here, we use the distribution conditional on the demographic variables D and lagged expenditures. In particular, given D and 25 cells in the five-by-five matrix of $t - 1$ and $t - 2$ expenditure intervals (used in table 1.1), we randomly choose from the distribution of residuals from the two components of equation (2), using a six-year window centered at the age of the individual whose expenditures are being predicted. Given D and the expenditure history captured by M_{lag}, we follow this exact procedure: First, choose a residual from the first component of equation (2) conditional on the demographic characteristics and expenditure history (as captured by the individual's position in the five-by-five matrix discussed above). If the first component of equation (2) evaluated at the independent variables and the chosen residual is greater than 0.5, the individual is assumed to have positive expenditures. Second, and only if the individual is assumed to have positive expenditures, choose a residual from the second component of equation (2) again conditional on demographics and expenditure history. Using this selected shock, we predict the magnitude of expenditure for those assumed to have positive expenditure.

The goal here is not to obtain "behavioral" estimates of marginal effects of predictor variables. Instead, we seek a prediction procedure that captures both the dynamic and cross-sectional features of health care expenditure over a lifetime. Below, we evaluate how well we have succeeded in attaining this goal.

This initial analysis has at least one potentially important limitation. It assumes that, given expenditures in $t - 1$ and $t - 2$, expenditures in prior years add no additional information about expenditures in t. Prior expenditures may, however, contribute additional information and could yield better predictions of persistence.

1.2.3 The Fit

There are at least four ways to check the extent to which the model captures the actual distribution of medical expenditures. All are based on a comparison between simulation results and the actual data. Some of the comparisons emphasize the dynamic properties of our simulation process—the degree to which the simulated expenditures capture the distribution of expenses over time. Others emphasize the extent to which the short-run model predictions capture the persistence observed in the three-year data panel. We consider the mean of expenditures by age, actual versus predicted expenditure distributions in year t given expenditures in $t - 1$ and in $t - 2$, actual versus predicted persistence over a three-year period, and the distribution of lifetime predictions at particular ages versus actual distributions at those ages.

Actual versus Predicted Means by Age

Figure 1.4 shows actual and predicted average expenditures by age for hourly and salaried males and females. The actual averages are based on the full sample of 230,497. The predicted averages are determined as follows: Begin with a sample of 1,000 employees aged 25. Then apply equations (1) and (2) repeatedly, producing a stream of expenditures for each person through age 60. The predicted averages for a given age are the averages of the predicted

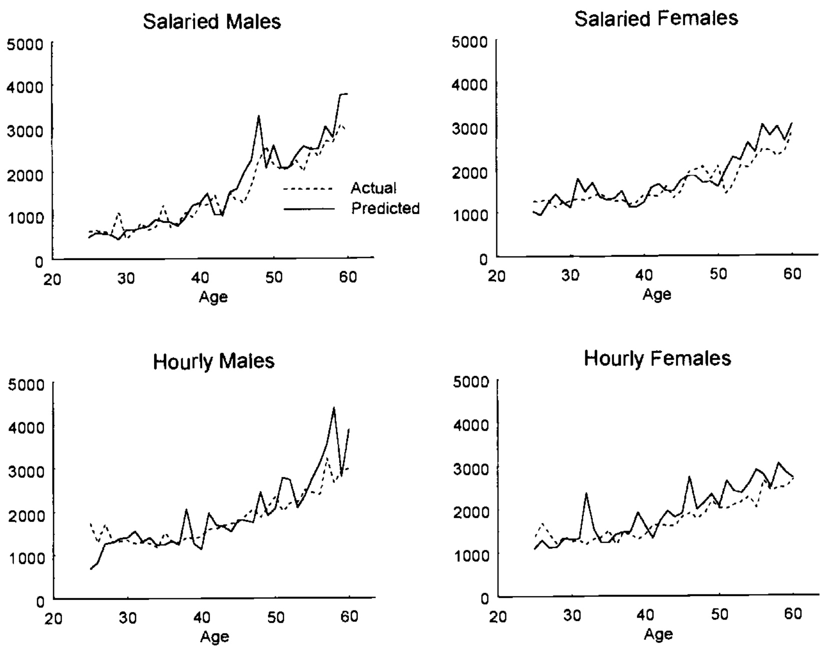

Fig. 1.4 Actual vs. predicted annual expenditures (dollars)

values at that age. Because the predicted values are based on a rather small number of persons there is more variation in the predicted than the actual averages, but the overall match seems quite close. (One might assume that any simple model—like a Tobit—would yield such a match. But because of the very skewed distributions of health expenditures this is not the case. Indeed, simple Tobit estimates yield means that are at least twice as large as the actual means.)

Actual versus Predicted Third-Year Expenditure Distributions

Figures 1.5A through 1.5D show the actual versus predicted distributions of third-year (1991) expenditures at selected ages—30, 40, 50, and 60—for hourly and salaried men and women. The predicted distribution is based on the demographic variables and expenditures for the first two years (1989 and 1990). Overall, the predicted and actual distributions are very similar.

Actual versus Predicted Persistence

Table 1.2 shows mean actual and predicted expenditures, conditional on the expenditure interval in each of the preceding two years. To illustrate, consider persons aged 36–45: for employees who had no claims in periods $t - 1$ and $t - 2$, the actual mean expenditure in period t was $1,295, compared to a predicted mean of $1,337. Overall, the predicted values capture quite well the pattern in the actual data with one exception. For individuals with spending above $5,000 in both $t - 1$ and $t - 2$, the prediction of $11,949 substantially exceeds the actual value of $9,934. Because of the very skewed distribution of shocks, random draws of very high shocks in the simulation can have a substantial effect on predicted means.

Distributions of Lifetime Predictions at Selected Ages versus Actual Distributions

Our simulations are designed to predict the expenditures of employees over a working lifetime. The comparison above suggested that the distribution of the model predictions of third-year expenditures were very close to the actual distribution of third-year expenditures. Now we consider a comparison that is intended to test the long-run implications of the model. We start with the expenditures of a sample of persons aged 25 in 1991. Then we simulate their expenditures through age 60. We want to know in particular that the distribution of predicted expenditures approximates the actual distribution at older ages. Figures 1.6A and 1.6B show predicted versus actual distributions at ages 45 and 55. Two comparisons are made: The first compares the predicted distribution at age 45 with the distribution of the actual expenditures of persons 45 years old in our sample. The second comparison shows the distribution of predicted and actual cumulated expenditures over three years, ages 45 through 47. Overall, the distributions of predicted expenditures are very close to actual distributions. Given the small (1,000) sample used for the simulations, simula-

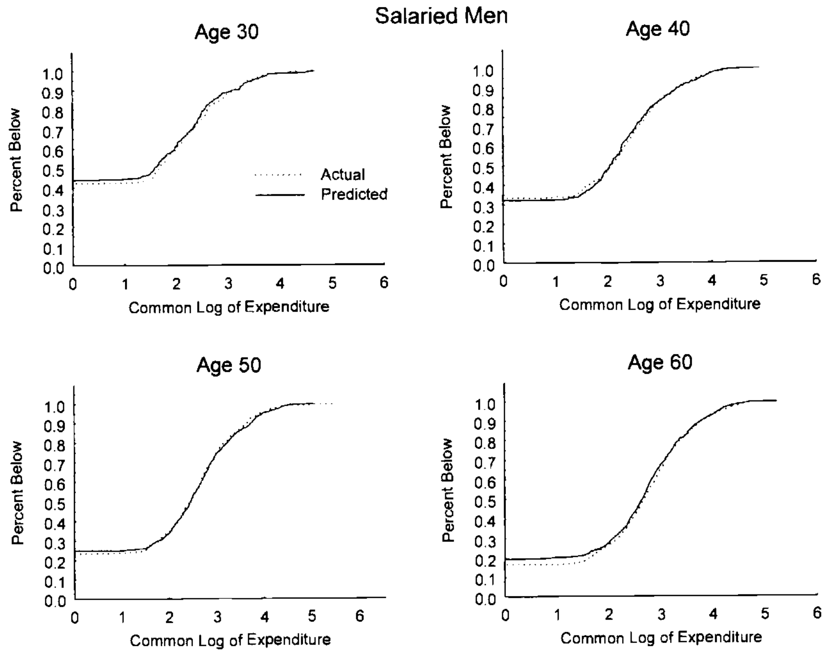

Fig. 1.5A Actual vs. predicted third-year expenditures by age: salaried men

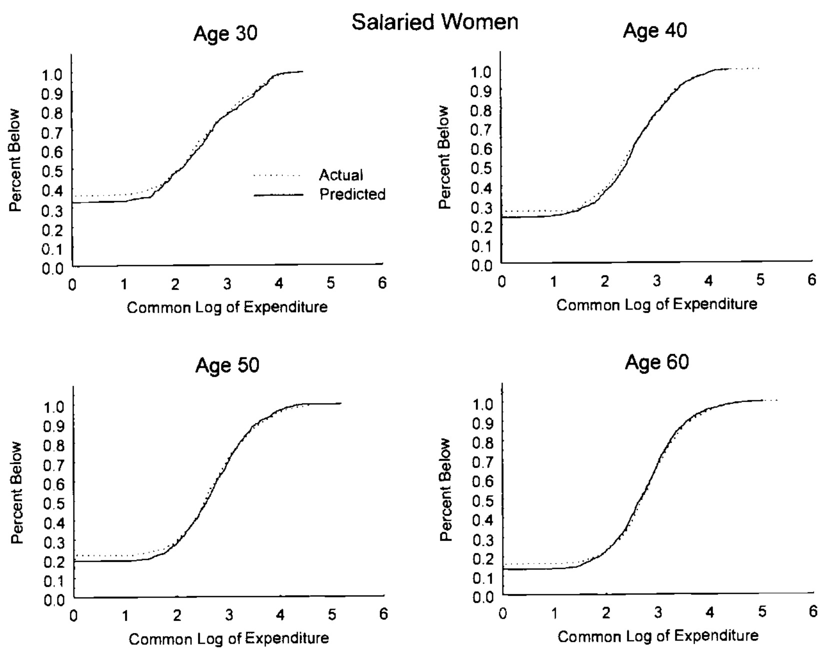

Fig. 1.5B Actual vs. predicted third-year expenditures by age: salaried women

Fig. 1.5C Actual vs. predicted third-year expenditures by age: hourly men

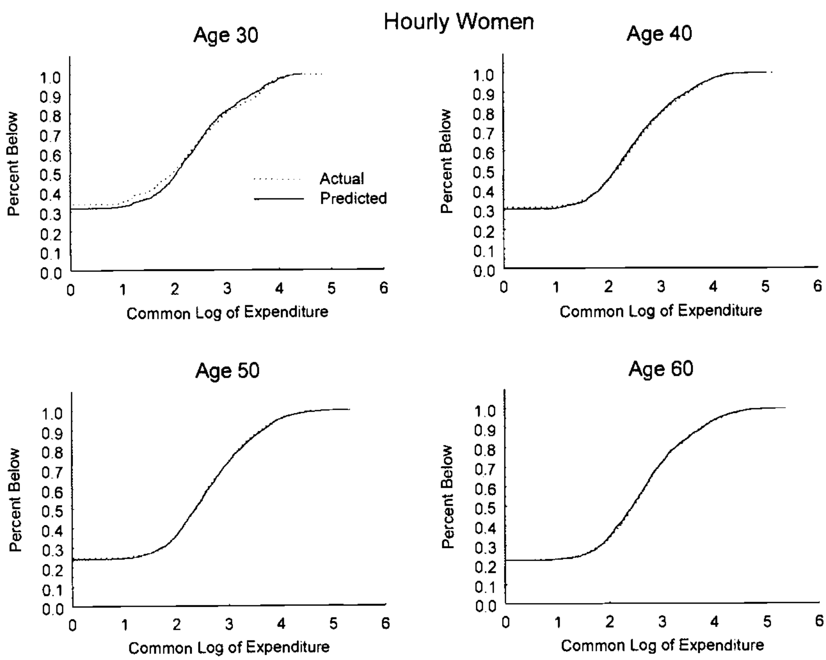

Fig. 1.5D Actual vs. predicted third-year expenditures by age: hourly women

Table 1.2 **Mean 1991 Expenditures Conditional on 1989 and 1990 Expenditures (dollars)**

1989 Expenditure ($)	1990 Expenditure ($): 0	0–300	300–1,000	1,000–5,000	Above 5,000
			Enrollees Aged 36–45		
			Actual Mean for 1991		
0	1,295	742	1,162	1,677	4,383
0–300	435	761	1,130	1,732	3,584
300–1,000	607	825	1,452	2,334	4,390
1,000–5,000	697	1,154	1,719	2,449	4,178
Above 5,000	844	1,811	2,175	4,003	9,934
			Predicted Mean for 1991		
0	1,337	727	1,206	1,828	4,153
0–300	459	780	1,199	1,797	3,751
300–1,000	639	776	1,419	2,507	5,903
1,000–5,000	1,206	1,386	1,847	2,543	4,659
Above 5,000	1,014	1,727	2,281	3,991	11,949
			Enrollees Aged 46–55		
			Actual Mean for 1991		
0	1,718	1,042	1,471	2,132	4,854
0–300	725	1,038	1,490	2,066	4,238
300–1,000	819	1,589	1,882	2,398	5,267
1,000–5,000	950	1,564	2,223	3,413	6,377
Above 5,000	1,593	2,069	2,965	4,327	10,380
			Predicted Mean for 1991		
0	1,739	991	1,477	1,998	5,824
0–300	691	1,094	1,554	2,364	4,163
300–1,000	695	1,674	1,729	2,440	4,809
1,000–5,000	1,086	1,524	2,076	3,459	7,031
Above 5,000	3,253	1,642	3,343	4,545	13,012

tions based on different samples yield somewhat different comparisons. But our experience has been that there is no appreciable difference in the overall results.

Other Validation Comparisons

Finally, we considered two additional comparisons to confirm that the persistence implications of the model were consistent with the data. Recall that the model assumes that, given expenditures in $t - 1$ and $t - 2$, expenditures in prior years add no information about expenditures in year t. But although it is likely that there is also some information contained in the pattern of expenditures in previous years, the model lag structure does imply that expenditure shocks will "last" for several years.

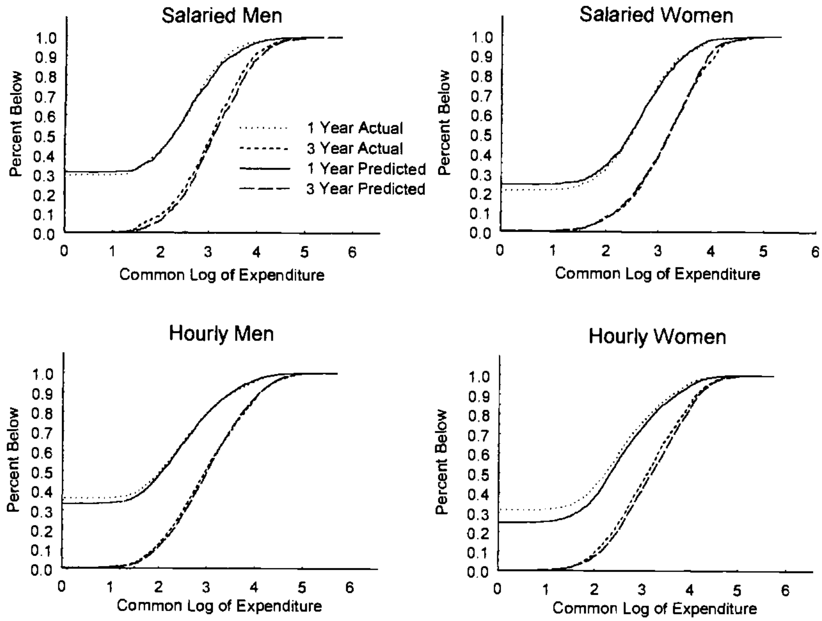

Fig. 1.6A Actual vs. predicted life cycle distribution at age 45 by gender and employment status

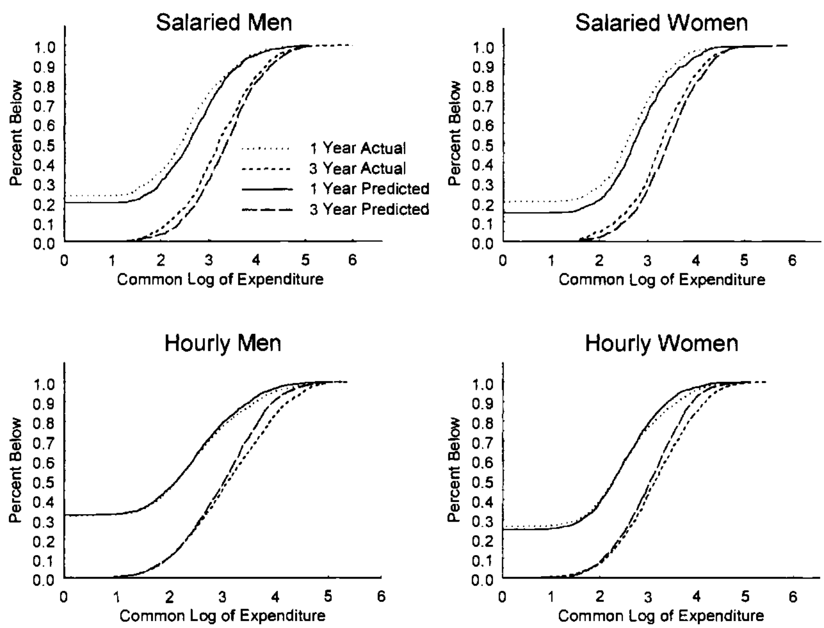

Fig. 1.6B Actual vs. predicted life cycle distribution at age 55 by gender and employment status

First, we considered all persons with high expenditures—over $10,000 and over $15,000—at ages 35 and 45 in 1989. We then used the model to simulate their expenditures in the 10 subsequent years. We compared the predictions in the subsequent two years with the actual data. For the first two years, for which we have matching actual data, the actual and simulated means are very close. The lag structure in the model is of course a way to extrapolate the decline in expenditures to future years, and the simulations imply that after four or five years the expenditures of persons with large shocks approach the overall sample mean. The simulated and actual "decay" patterns are shown in appendix figure 1A.1.

Second, we considered the future expenditures of persons with specific 1989 diagnoses that are typically associated with high expenditures. In particular, we were concerned that the expenditure decay in these cases be consistent with the implications of the model. We considered these 1989 diagnosis: acute myocardial infarction (AMI), cancer, mental health disorder (with inpatient care), and pregnancy. Only 45 percent of 1989 AMI patients had expenditures greater than $1,000 one year later in 1990; only 34 percent had expenditures greater than $1,000 two years later in 1991. (Over 14 percent had zero expenditures in 1990, and 25 percent in 1991.) Less than 25 percent of cancer patients had expenditures over $1,000 in 1990, and only 20 percent in 1991. There was more persistence in the expenditures of inpatient mental health patients: 54 percent had expenditures over $1,000 in 1990, and 42 percent in 1991. Pregnancy is one of the most important contributors to firm health care costs, but with minimal persistence. Only 17 percent of women with pregnancy-related diagnosis in 1989 had expenditures over $1,000 in 1990, and only 13 percent in 1991. We take these results as evidence that our simulated decay rates are not unreasonable. In particular, we find no reason to suspect that they are too rapid.

We conclude that simulated expenditure patterns match closely actual expenditure patterns revealed in the three years of our data.

1.3 Simulation Results

We have simulated the lifetime expenditures of 1,000 employees who begin work at age 25 and retire at age 60. We realize that few, if any, persons would work for the same firm for that length of time, but it is the expenditure pattern that we want to capture, assuming that employees continued to face an insurance scheme like the one at this firm.

1.3.1 Distribution of Lifetime Expenditures

The distribution of cumulative expenditures at selected ages is shown in figure 1.7A. Figure 1.7B shows the concentration version of the data. Over a working lifetime, expenditures of salaried males vary from less than $10,000 (about 10 percent of employees) to over $100,000 (about 10 percent of employ-

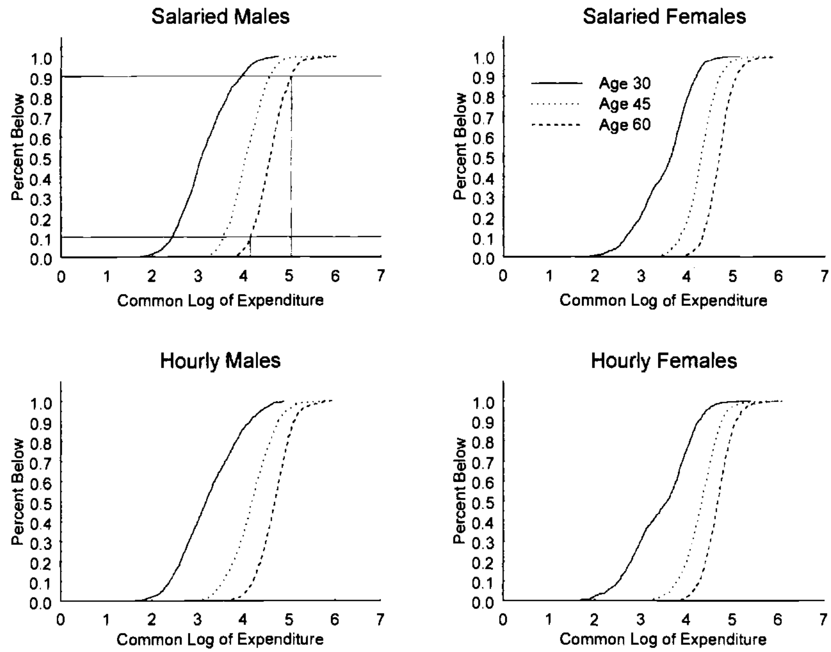

Fig. 1.7A Simulated distribution of expenditures at various ages

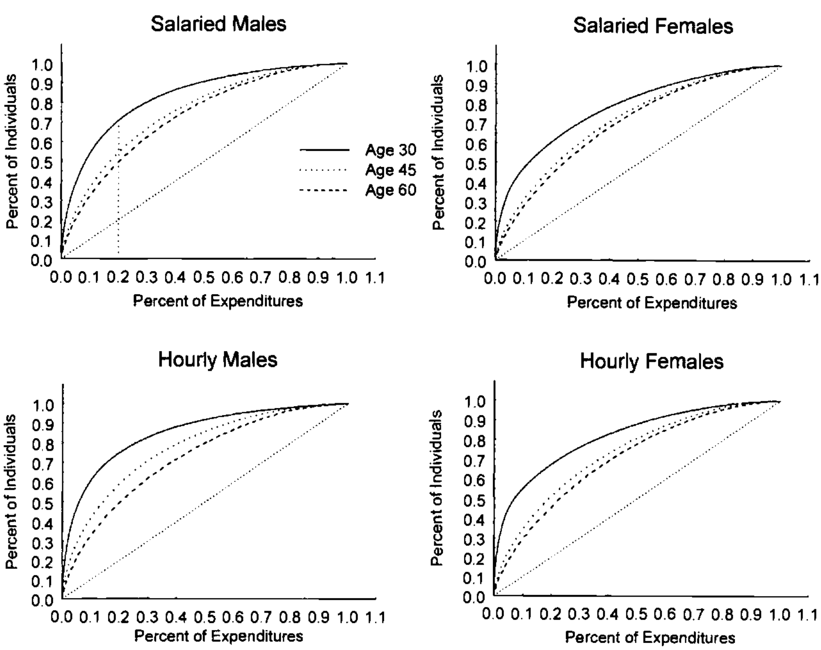

Fig. 1.7B Cumulative expenditure concentration at various ages

ees). The median is about $32,000, as shown in figure 1.7A. The distributions for the other gender and employee status groups are similar to those for salaried males.

Figure 1.7A shows that the distribution of expenditures is much less concentrated over a lifetime than over five years (at age 30). Figure 1.1 (above) shows that in a single year about 80 percent of expenditures are accounted for by about 10 percent of employees. For salaried males, figure 1.7B shows that after five years (at age 30) about 29 percent of employees account for 80 percent of expenditures. Over a lifetime about 48 percent of employees account for 80 percent of expenditures.

1.3.2 IHA Balances

Given the distribution of expenditures described above, how might an IHA plan work? We consider this plan:

- The employer puts $2,000 in each employee's IHA at the beginning of each year.
- The health insurance plan has a $4,000 annual deductible, with expenses below the deductible paid by the employee (out of the IHA) and 100 percent of expenditures above the deductible covered by the health insurance plan. If the IHA balance goes to zero, all expenses are paid by the insurance plan.

The distribution of IHA balances at selected ages is shown in figure 1.8. Consider salaried males: After five years (at age 30), about 50 percent of men have balances close to $10,000. Only about 10 percent have balances less than about $6,000. After a lifetime, there is more variation in the IHA balances, but most employees are left with a substantial accumulation. About 90 percent of the employees have a balance at age 60 that exceeds $25,000, while 75 percent have more than $40,000 and 50 percent have more than $50,000. The distributions are similar for salaried females and for hourly employees.

Another way to understand the implications of the plan is to consider the proportion of IHA contributions that remain at selected ages. The distribution of these proportions is shown in figure 1.9. Two features of the distribution stand out: the fraction declines with age, but even at retirement the fraction remaining is large for almost all employees. At retirement, only about 20 percent of employees have less than 50 percent of their contributions, about 10 percent have less than 35 percent, and about 5 percent have less than 20 percent. And 50 percent still hold more than 70 percent of their IHA contributions.

The average balance remaining in the IHA is shown by age in figure 1.10. The fraction can be compared to the 45 degree line that represents the accumulation path if there are no withdrawals to cover health care costs. The fraction remaining is higher for salaried than for hourly employees and highest for salaried males. The average balance of salaried males is about $46,000 at age 60.

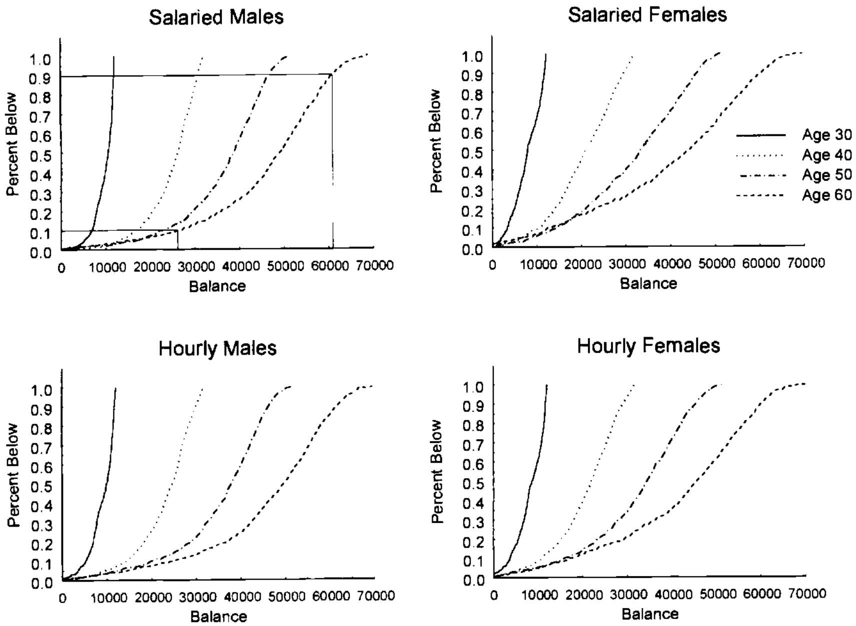

Fig. 1.8 IHA balances (dollars) at selected ages

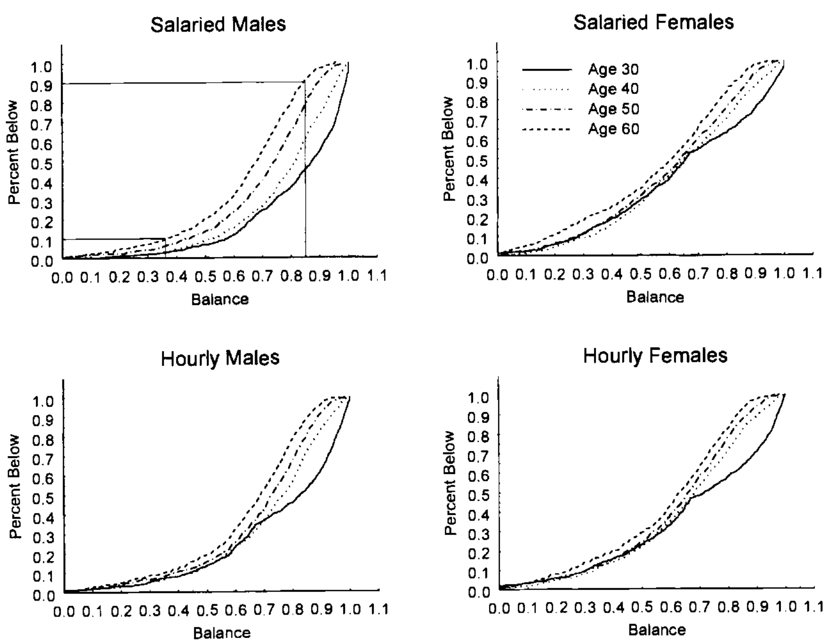

Fig. 1.9 IHA balances (percent of contributions) at selected ages

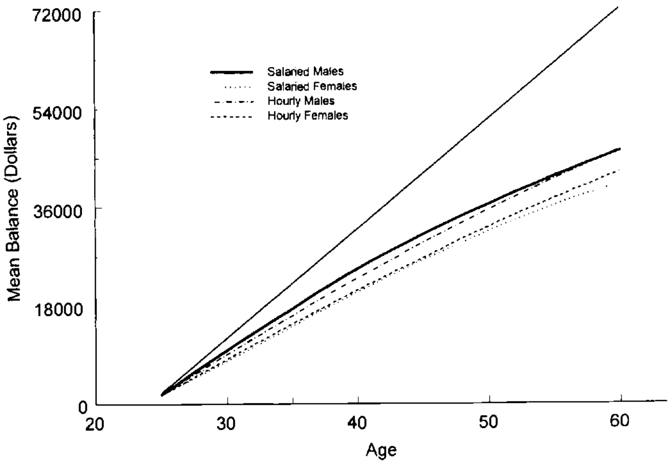

Fig. 1.10 Mean IHA balances by age

The higher average for salaried than for hourly employees may be attributed in large part to differences in plan provisions. The hourly plan provides first-dollar coverage, while the salaried plan includes copayments and a deductible. We have not accounted here for any behavioral effects. If the difference in expenditures of hourly and salaried employees is due to plan provisions, it may be an indication that the behavioral response to a catastrophic plan could be substantial as well.

1.4 Discussion

A health insurance system featuring IHAs combined with a catastrophic insurance plan would insure employees against high health care costs but would also subject a large fraction of expenditures to the discipline imparted by "spending your own money." A potential drawback of this scheme is increased risk, measured by variation in IHA balance accumulation. The variation depends critically on the lifetime distribution of health care costs. With emphasis on this issue, our goal was to present preliminary evidence on the feasibility of an IHA plan.

We developed a nonparametric method to describe the longitudinal distribution of health care expenditures in a large firm. Comparisons with actual firm data suggest that the model captures well the important features of the expenditure distribution. Using this model to simulate the lifetime distribution of individual health care costs, we evaluated the implications of variation in health care expenditures for variation in IHA balances at retirement. Although the plan would produce a range of balances across employees, approximately 80 percent of employees would retain more than 50 percent of their IHA contributions. Only about 5 percent of employees would retain less than 20 percent of their contributions. These outcomes do not appear to us to be so extreme as to make the plan a nonstarter.

To say more we need to incorporate additional components into the analysis. While we believe that such a plan would reduce medical expenses substantially, we must quantify the behavioral response that would occur as employees spend more of their own money. Rough calculations suggest that the proportion of payments subject to this restraint would increase from virtually zero to as much as 45 percent (depending on whether the payment that pushes the total over the $4,000 deductible is counted). And we believe that such a plan could be structured to increase retirement saving. The benefits of reduced health care costs and increased saving must be considered against the risk associated with increased variance in lifetime medical expenditures. Traditional insurance plans that offer near total insurance, and erect a large moral hazard along the way, dampen this variability in lifetime expenditures. By considering risk aversion and time preferences, we can address these trade-offs more formally in the context of utility analysis.

An IHA plan could have important implications for the composition and, possibly, the level of employee compensation. The overall annual cost of the firm's current health plan is around $1,400 per participant. With no behavioral effects, the IHA plan we consider would reduce this premium to around $700 per participant. Employee health care costs would increase by about $700. Assuming no reduction in other nonwage benefits, the IHA contribution of $2,000 would result in a net increase in total employee benefits of approximately $1,300. Evidence on 401(k) plans suggests that offsetting effects on other employee benefits are unclear. Many 401(k) plans were established with no apparent reductions in other components of employee compensation, at least in the short run.

The actual additional cost to be divided between the firm and its employees may be substantially lessened not only by any behavioral effect but also by the favorable tax treatment accorded to 401(k) and other retirement programs, as well as health insurance premiums. For example, if contributions to an IHA were also treated in this manner and perhaps combined with a retirement savings program, copayments and deductibles now made by employees with after-tax earnings might be replaced by payments made with pretax earnings from an IHA account.

We also need to explore variations in the structure of IHA plans to understand the implications of alternative IHA contributions, savings accumulations, and out-of-pocket expenditures. Our analysis reveals that small changes in the structure of the plan can substantially alter the "way the numbers look." For example, if the IHA contribution is made at the end of the year—and thus is not available to fund expenditures in the current year—virtually no one would be left with IHA accumulations less than 35–40 percent of their contributions (although this would generate a somewhat higher insurance premium). The results would also look quite different if the IHA contribution were $1,000 instead of $2,000.

We intend to consider all these issues more thoroughly in future work. We also plan to reestimate the parameters of our model with longer panels of expenditure data when they become available, and to consider other kinds of health plan reforms from the perspective of lifetime expenditures.

Appendix

Table 1A.1 Model Parameter Estimates

	Linear Probability		Log Linear if Amount > 0	
Variable	Estimate	*t*-Statistic	Estimate	*t*-Statistic
	Simplified Specification			
Constant	0.831471	130.294065	5.968484	177.958314
Age	−0.001120	−9.105254	0.007943	12.126717
Sex	0.044624	5.543992	0.207343	4.951444
(Age)(Sex)	−0.000237	−1.441283	−0.003874	−4.481654
Hourly	−0.023953	−12.092770	−0.027556	−2.862969
D1	−0.449010	−184.611500	−0.542924	−33.122375
D2	−0.183226	−73.009823	−0.302466	−23.604705
D12	0.203681	69.794990	−0.748530	−60.388999
M1	0.000004	14.447127	0.000043	30.102726
M2	−0.000003	−13.460566	0.000041	9.894052
M*1	0.000001	10.385432	0.000040	60.611868
M*2	−0.000000	−2.210547	0.000028	37.152903
	Piecewise Linear with Age Interaction			
Constant	0.767076	76.081033	5.218912	102.262185
Age	−0.005531	−31.751904	0.001529	1.688721
Sex	0.038068	4.781396	0.199626	4.886164
(Age)(Sex)	−0.000180	−1.105580	−0.004235	−5.019273
Hourly	−0.017851	−9.112001	−0.001713	−0.182322
D1	−0.267524	−28.154423	0.078253	1.357532
D2	−0.072295	−7.645898	0.138964	2.828726
D12	0.421501	54.552813	0.200218	5.465471
M1	0.000407	11.373989	0.001937	10.145843
$M1_{300}$	−0.000308	−6.716501	−0.001345	−5.583627
$M1_{1000}$	−0.000097	−7.018825	−0.000560	−8.153725
M2	0.000223	5.931243	0.001604	6.105962
$M2_{300}$	−0.000256	−5.280451	−0.001434	−4.221094
$M2_{1000}$	0.000030	1.989208	−0.000157	−1.449693
M*1	0.000529	19.042785	0.001386	10.339624
$M{*}1_{300}$	−0.000441	−12.932473	−0.000490	−3.015251
$M{*}1_{1000}$	−0.000088	−9.944736	−0.000868	−21.430002
M*2	0.000183	6.669677	0.000811	6.275513
$M{*}2_{300}$	−0.000171	−5.061111	−0.000228	−1.444769

Table 1A.1 (continued)

Variable	Linear Probability Estimate	Linear Probability *t*-Statistic	Log Linear if Amount > 0 Estimate	Log Linear if Amount > 0 *t*-Statistic
	Piecewise Linear with Age Interaction			
M^*2_{1000}	−0.000013	−1.448336	−0.000562	−13.744969
Age45	0.131443	14.109861	0.071001	1.571068
(D1)(Age45)	−0.016230	−1.306935	0.132904	1.749148
(D2)(Age45)	−0.035120	−2.738827	0.091063	1.360564
(D12)(Age45)	−0.043627	−4.236622	0.099796	2.078911
(M1)(Age45)	0.000055	1.075926	−0.000170	−0.622411
($M1_{300}$)(Age45)	−0.000063	−0.963313	0.000211	0.616798
($M1_{1000}$)(Age45)	0.000006	0.310614	−0.000046	−0.479565
(M2)(Age45)	−0.000079	−1.587965	−0.000168	−0.477427
($M2_{300}$)(Age45)	0.000066	1.027559	0.000211	0.463816
($M2_{1000}$)(Age45)	0.000014	0.709447	−0.000034	−0.239341
(M*1)(Age45)	0.000047	1.335356	0.000065	0.381546
(M^*1_{300})(Age45)	−0.000070	−1.623723	−0.000028	−0.136822
(M^*1_{1000})(Age45)	0.000022	2.014475	−0.000041	−0.809818
(M*2)(Age45)	−0.000006	−0.168130	0.000305	1.868320
(M^*2_{300})(Age45)	−0.000014	−0.317268	−0.000340	−1.710830
(M^*2_{1000})(Age45)	0.000019	1.673649	0.000025	0.494918

Variable Definitions:

Age	Age in 1989
Sex	1 if female; 0 otherwise
(Age)(Sex)	Age interacted with Sex
Hourly	1 if hourly worker; 0 otherwise
D1	1 if no expenditures in period $t-1$; 0 otherwise
D2	1 if no expenditures in period $t-2$; 0 otherwise
D12	1 if no expenditures in periods $t-1$ and $t-2$; 0 otherwise
M1	Expenditure in $t-1$ if no expenditure in $t-2$
$M1_{300}$	Expenditure in $t-1$ minus 300 if no expenditure in $t-2$ and expenditure in $t-1$ above 300
$M1_{1000}$	Expenditure in $t-1$ minus 1,000 if no expenditure in $t-2$ and expenditure in $t-1$ above 1,000
M2	Expenditure in $t-2$ if no expenditure in $t-1$
$M2_{300}$	Expenditure in $t-2$ minus 300 if no expenditure in $t-1$ and expenditure in $t-2$ above 300
$M2_{1000}$	Expenditure in $t-2$ minus 1,000 if no expenditure in $t-1$ and expenditure in $t-2$ above 1000
M*1	Expenditure in $t-1$ if expenditure in both $t-1$ and $t-2$
M^*1_{300}	Expenditure in $t-1$ minus 300 if expenditure in both $t-1$ and $t-2$ and expenditure in $t-1$ above 300
M^*1_{1000}	Expenditure in $t-1$ minus 1,000 if expenditure in both $t-1$ and $t-2$ and expenditure in $t-1$ above 1,000
M*2	Expenditure in $t-2$ if expenditure in both $t-1$ and $t-2$
M^*2_{300}	Expenditure in $t-2$ minus 300 if expenditure in both $t-1$ and $t-2$ and expenditure in $t-2$ above 300
M^*2_{1000}	Expenditure in $t-2$ minus 1,000 if expenditure in both $t-1$ and $t-2$ and expenditure in $t-2$ above 1,000

(*continued*)

Table 1A.1 (continued)

Age45	1 if age in 1989 above 45; 0 otherwise
(D1)(Age45)	D1 interacted with Age45
(D2)(Age45)	D2 interacted with Age45
(D12)(Age45)	D12 interacted with Age45
(M1)(Age45)	M1 interacted with Age45
($M1_{300}$)(Age45)	$M1_{300}$ interacted with Age45
($M1_{1000}$)(Age45)	$M1_{1000}$ interacted with Age45
(M2)(Age45)	M2 interacted with Age45
($M2_{300}$)(Age45)	$M2_{300}$ interacted with Age45
($M2_{1000}$)(Age45)	$M2_{1000}$ interacted with Age45
(M*1)(Age45)	M*1 interacted with Age45
($M*1_{300}$)(Age45)	$M*1_{300}$ interacted with Age45
($M*1_{1000}$)(Age45)	$M*1_{1000}$ interacted with Age45
(M*2)(Age45)	M*2 interacted with Age45
($M*2_{300}$)(Age45)	$M*2_{300}$ interacted with Age45
($M*2_{1000}$)(Age45)	$M*2_{1000}$ interacted with Age45

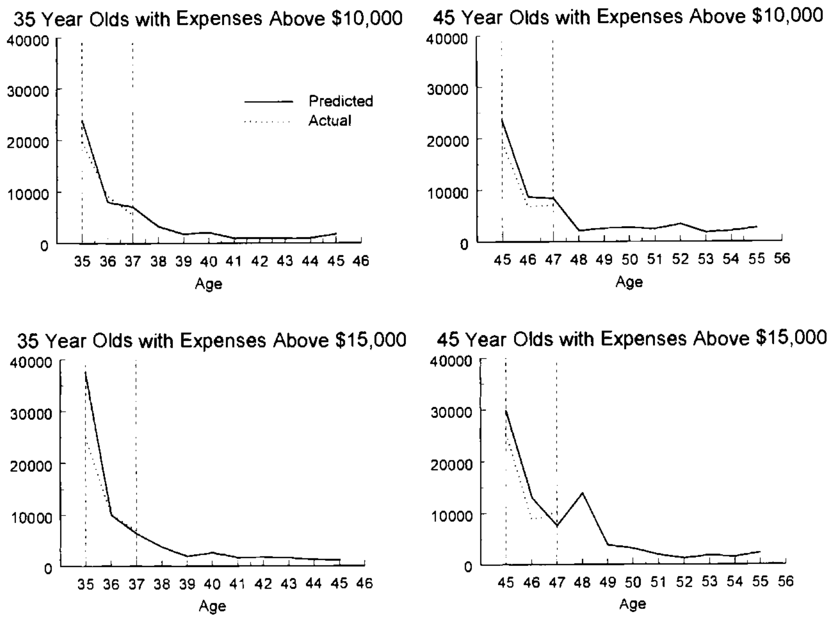

Fig. 1A.1 Decay pattern after large shocks: salaried men

References

Arrow, Kenneth J. 1963. Uncertainty and the welfare economics of medical care. *American Economic Review* 53:941–73.

Feldstein, Martin. 1971. A new approach to national health insurance. *Public Interest* 23:93–105.

Feldstein, Martin, and Jonathan Gruber. 1994. A major risk approach to health insurance reform. NBER Working Paper no. 4852. Cambridge, Mass.: National Bureau of Economic Research.
McClellan, Mark B., and David A. Wise. 1994. Where the money goes: Medical expenditures in a large corporation. Paper prepared for the JCER-NBER Conference on Health Care, December.

Comment Jonathan Gruber

I really enjoyed this paper, and I think that it helps us to focus on an important debate within the health economics community. Many economists and health policy analysts are in agreement that increasing price incentives and consumer shopping will increase efficiency in health care markets. But there is considerable divergence over the point at which those price incentives should be introduced.

Traditionally, economists such as Arrow and Feldstein emphasized incentives at the *point of service*, such as through high copayment rates and deductibles. The effectiveness of this mechanism was demonstrated by the RAND Health Insurance Experiment, which found that individuals did use significantly less medical care if they faced a significant price on the margin.

More recently, however, the emphasis has shifted to price incentives at the *point of health insurance plan choice*. Rather than demand-side copayments, the Clinton plan emphasized having employees pay a larger share of their insurance premiums, to get them to shift to managed care plans, which use supply-side controls. The effectiveness of this mechanism is more questionable, however. While strictly defined HMOs do appear to have lower costs, even controlling for selection, there is little evidence that other forms of managed care have been particularly successful in lowering costs. And no form of managed care has managed to restrain cost growth, either at the level of a particular managed care institution or if one compares medical cost growth across states with more or less managed care in place.

Thus, it seems sensible to shift the focus back to demand-side cost sharing, at least to some extent. The natural way to do this is with large copayments and deductibles, but these are often viewed as politically unpalatable. An alternative, which has the same incentive properties but simply redistributes resources, is to have "individual health accounts" (IHAs). Under this plan, individuals get an account with some funds in it. They then face 100 percent of their medical costs up to some level, say $2,000. Any spending up to this level comes out of their accounts, and they get to keep the rest. Under the particular variant described by the authors, they keep it in a pensionlike savings account

Jonathan Gruber is professor of economics at the Massachusetts Institute of Technology and a faculty research fellow of the National Bureau of Economic Research.

that they get at retirement, which has the virtue of subsidizing retirement savings.

An important criticism of this plan, however, is that serial dependence in health expenditures could lead to substantial inequities in plan accumulation. That is, if individuals who are high spenders are lifetime high spenders, then they will use up their money every year and have little in their accounts at retirement. Although it is only mentioned briefly by the authors, this has potentially nasty distributional implications. Those individuals who are lifetime high spenders may be exactly the persons about whose retirement resources we are most worried, for example, because they were unable to advance in their jobs due to health problems, or because their spouses had to stay home to help with the health problems. Furthermore, these are the persons who will continue to face high expenditures when retired, given the high deductibles and copayments under the Medicare program.

So the critical question to ask in evaluating the magnitude of this redistribution is: is there such a thing as lifetime sick or healthy? Strikingly, however, we know virtually nothing about the intertemporal correlation of medical spending for an individual. I know of only one article on this subject, by Dan Feenberg and Jon Skinner, using data on catastrophic spending from tax returns, but this is very specialized data on only a limited set of spending events.

Into this critical gap jump the Eichner et al. team. The goal of their careful empirical work is exactly to measure this illusive parameter: the intertemporal correlation in medical spending. The authors do so using a unique database on medical spending at a large firm. They have information on spending for 300,000 workers over a three-year period. And they use these data to demonstrate two facts about the distribution of medical spending. First, medical spending is very concentrated at a point in time. Second, the *intertemporal* concentration is much lower than the point-in-time concentration. The raw data show that, of those who spend more than $5,000 on medical care in one year, only 20 percent spend that much two years later. And only 35 percent of those in the highest quintile of spending in one year are in that quintile two years later.

The authors then present a formal statistical model to project the persistence of expenditures. This is a two-part model, with one equation for predicting whether there is spending and one equation modeling the level of spending conditional on having expenditure. The dependent variable is the level of spending in the third year. The explanatory variables include a limited set of demographic characteristics and detailed controls for spending over the previous two years.

Their model predicts only moderate persistence over a period as long as five years, and little persistence over a lifetime. This is illustrated most graphically in appendix figure 1A.1, which shows that spending is predicted to revert rather quickly to the mean for high spenders. Overall, the authors find that, over a lifetime, 48 percent of workers account for 80 percent of expenditures, rather than the 10 percent in a single cross section. So this suggests that there will not be a problem with substantial inequities in the ultimate distribution of IHA balances.

I find this to be quite an important and provocative conclusion, and what I would like to do is offer a couple of observations or criticisms on the methods. First, the obvious limitation of the empirical work is that the authors have only three years of data to work with in projecting lifetime concentrations. The authors are aware of this limitation and ultimately hope to remedy it by adding more data. However, this may not be as important a limitation as one would initially think. Examining their appendix figures and results, one notes that most of the decay in spending happens after only one year, so that having more years of data may not be very important.

In fact, this suggests an additional specification check of their model. The authors could fit the model based on the second year of data, rather than the last year, and model spending as a function of a one-year lag only. They could then do an out-of-sample projection on the third year to see how well the model fits. This would be more convincing than the tests that they do now, since it is both truly out of sample and can be compared to actual outcomes. Another way to say this is: if a second year of lag does not add much to the fit of the model, then further years may not matter either.

On the other hand, more years could make an important difference if medical spending follows a sort of "S-s" pattern, rather than a smooth intertemporal decline. That is, individuals may have chronic illnesses that flare up, requiring expensive care every fourth year but not much in between. I do not have much insight into whether spending is more likely to be continuously high or to flare up, but presumably the team could usefully draw on Dr. McClellan's insights here as to how much of a problem this is likely to be in reality.

Second, there are a couple of specification issues with their model. One picky point is that, even though they only have a limited set of demographic variables, they do not use them as fully as possible. Given their sample size, there would be no problem including much more detailed age-by-sex categories in the model, rather than just a linear age * sex interaction. A more important point, which is recognized by the authors, is that they build in no correlation between the error terms of the 1/0 spend/don't spend equation and the level-of-spending equation. These errors are clearly correlated in an important way, and accounting for this could improve the fit of the model even further. I appreciate the difficulty of modeling this error correlation with such bizarrely distributed data, but the new innovations in nonparametric modeling of error distributions could perhaps be usefully applied here.

Third, the analysis does not incorporate any behavioral response to the IHA plan. If demand for medical care is elastic at the point of service, overall spending will fall under an IHA plan. But, at the same time, the lifetime concentration of medical expenditures may increase, if the chronically high spenders are less elastic in their responses than are low spenders. So introducing behavioral responses could exacerbate the ultimate inequities in IHA balances at retirement. This is clearly an important priority for future work with these data.

Finally, in terms of where this work goes next, I think that the authors should spend some time thinking about three critical design questions involved in an IHA-type plan. First, should this be a savings account or should individuals just get the money back at the end of the year? There are obvious advantages to structuring IHAs as savings vehicles. But one cost of doing so is that individuals may not perceive the funds left in the IHA to be as valuable as if they got the money back. This is important because the elasticity of response to this incentive may be a function of the value that individuals place on the money that they get to keep in their accounts.

This is a major concern with current cafeteria plans such as that at MIT. These plans allow individuals to put away a certain amount of money that can be used on a pretax basis for out-of-pocket medical spending during the year. But whatever is left over at the end of the year is lost to the individual. As a result, individuals with money left at the end of the year have a "wheel of fortune" mentality, either purchasing unnecessary care or substituting forward care that would have been delivered in some future year. Presumably, such a severe response would not arise with an IHA where you got the money back. But it might arise with an IHA where the money is saved, if individuals do not value those savings, which may be true for younger employees. More generally, the behavioral response to the incentives embodied in an IHA might be larger if the cash is received back directly rather than saved. So there is a trade-off here between the cost of limiting the incentive effects on medical spending and the gains from inducing increased saving by employees.

Second, should IHAs use deductibles or copayments? An alternative to individuals' paying all of their costs up to $2,000 would be to have them pay 50 percent of their costs up to $4,000. This would make individuals somewhat less sensitive to medical spending but do so over a much larger range. In my work with Martin Feldstein on behavioral responses to catastrophic health care plans, we found that the reduction in deadweight loss is larger for a copayment plan than for a deductible plan with the same maximum out-of-pocket exposure. This is because the initial increases in cost sharing have the greatest marginal reduction in deadweight loss, so spreading a moderate cost-sharing increase over a greater sensitive range is more efficient. Furthermore, we also found that the increase in risk bearing was much greater for the deductible plan. This follows obviously from the point that the total out-of-pocket exposure is the same under both plans, but individuals will spend more out of pocket on average with the deductible plan. Thus, our calculations suggest that a copayment plan dominates a deductible plan if the goal is to reduce deadweight loss while minimizing the increase in risk bearing. In future work, the authors should consider the distributional implications of copayment plans instead of deductible plans, along with building in an elasticity-of-spending response.

Finally, what other government interventions might be necessary in an IHA world? It is worth noting that there is little use of very high deductible plans in practice. There are two obvious reasons for this. One is the tax subsidy to employer-provided insurance, which distorts individual preferences toward more generous insurance plans. The other is the possibility of supplemental insurance purchase, which could undo the spending incentives put into place by a catastrophic-type plan. If an IHA is to be an effective means of controlling spending, some type of restriction on the use of such supplemental plans may have to be put in place.

To conclude, let me highlight my belief that this line of research has the potential to help guide the policy agenda on IHAs. And I urge the authors to continue to update and improve their vast array of data in order to confirm and extend their interesting results.