AN ESTIMATE OF THE IMPACT OF DEDUCTIBLES ON THE DEMAND FOR MEDICAL CARE SERVICES

PREPARED UNDER A GRANT FROM THE U. S. DEPARTMENT
OF HEALTH, EDUCATION, AND WELFARE

JOSEPH P. NEWHOUSE, JOHN E. ROLPH,
BRYANT MORI, MAUREEN MURPHY

R-1661-HEW
OCTOBER 1978

Rand
SANTA MONICA, CA. 90406

The research reported herein was performed pursuant to Grant No. 016B-7901-P2021 from the U.S. Department of Health, Education, and Welfare, Washington, D.C. The opinions and conclusions expressed herein are solely those of the authors, and should not be construed as representing the opinions or policies of any agency of the United States Government.

ISBN 0-8330-0065-9
LC Catalog Card Number 78-21763

Published by The Rand Corporation

AN ESTIMATE OF THE IMPACT OF DEDUCTIBLES ON THE DEMAND FOR MEDICAL CARE SERVICES

PREPARED UNDER A GRANT FROM THE U. S. DEPARTMENT
OF HEALTH, EDUCATION, AND WELFARE

JOSEPH P. NEWHOUSE, JOHN E. ROLPH,
BRYANT MORI, MAUREEN MURPHY

R-1661-HEW
OCTOBER 1978

Rand
SANTA MONICA, CA. 90406

PREFACE

This report was written as part of the nonexperimental work under the Rand Health Insurance Study, which is being conducted under a grant from the U.S. Department of Health, Education, and Welfare. Whereas earlier nonexperimental work concentrated on measuring the effects of variation in the coinsurance rate on the demand for medical services, this report presents estimates of variation in a deductible on the demand for medical care. Because there are no known estimates of demand behavior with respect to a deductible, the results presented in this report are of significance. Appreciably more reliable results will be possible when data from the experimental portion of the Health Insurance Study become available in several years. In the interim, these results, however unsatisfactory, are not likely to be significantly improved upon.

The empirical results complement a theoretical model of demand under a deductible presented in another Rand report, *Deductibles and the Demand for Medical Services: The Theory of the Consumer Facing a Variable Price Schedule under Uncertainty*, by Emmett Keeler, Joseph Newhouse, and Charles Phelps, R-1514-OEO/NC, December 1974 (also in *Econometrica*, April 1977). The present report should be of interest to analysts concerned with estimating demand under various health insurance plans or under alternative national health insurance legislation.

SUMMARY

Using data on insurance premiums for policies with varying deductibles, together with a distribution of medical expense, we estimate the relationship between deductibles and the demand for medical care. The estimates are limited to deductibles in the range of $50 to $1000 (per person per year, 1975 dollars).

The results indicate that demand is quite sensitive to variation in a deductible in the region of $50 and becomes steadily less sensitive as the deductible rises above $75. (The estimated demand curve is graphed in Fig. 4.3, p. 33). Our results are consistent with a theoretical model of demand for medical care given a deductible, as well as with what is known about the responsiveness of demand to variation in coinsurance. The results show that the size of the deductible will have an important effect on the amount of public funds used in a national health insurance program and thereby on the distribution of payments among the population (in a tax-financed program). Finally, no evidence was found that an increase in the deductible causes increased expenditures by deterring efficacious preventive care, but the data are not well suited to test this hypothesis.

Our method included the estimation of the parameters of a distribution that describes medical care expense (out-of-pocket plus insurance payments), given an insurance plan with a $50 deductible. In estimating these parameters, we assume that as the deductible changes, the distribution will shrink toward zero in a multiplicative fashion. Insurance premium data can be used to estimate the shrinkage factor if we assume that the premium is proportional to the insurer's payment. (The insurance companies assert that this is the case.) The actual proportionality factor is estimated from the distribution of expense at a $50 deductible and the premium for that policy. Once the distribution of expense, conditional upon various deductibles, has been estimated, it is straightforward to estimate both the mean expenditure at those deductibles (leading to the aggregate demand curve) and the cost to the insurer. Under a public plan, this latter figure would appear, of course, as a budgetary cost.

ACKNOWLEDGMENTS

The authors received extremely valuable comments from Alain Enthoven, Emmett Keeler, Charles Phelps, and Finis Welch on various drafts of this report. They are also grateful to George Harris of Arthur D. Little, Inc., for making data available on expenditures by federal employees.

The research reported herein was performed as part of the Rand Health Insurance Study, conducted under a grant from the U.S. Department of Health, Education, and Welfare, Washington, D.C. The opinions and conclusions expressed in this report are solely those of the authors, and should not be construed as representing the opinions or policy of any agency of the United States Government.

CONTENTS

Section 1

INTRODUCTION

Estimating the effect of a deductible on the demand for medical care services has both theoretical and policy significance. Although the theory of an optimal deductible is still incomplete, it is apparent that any attempt to calculate an optimal deductible will require quantification of how demand responds to variation in the deductible.*

The demand response to variation in a deductible is also of significance to the national health insurance debate, because no deductible, or a very small deductible, will lead to a marked increase in the demand for outpatient physician visits, perhaps 75 percent or more (Newhouse, Phelps, and Schwartz, 1974). Such an increase would almost certainly cause nonprice rationing of physician services, something many would consider undesirable. Unfortunately, one can, at present, only conjecture how large a deductible would suffice to keep demand from rising sharply (Newhouse, Phelps, and Schwartz, 1974; Keeler, Newhouse, and Phelps, 1977). In this report, we present the first evidence of how demand might respond to variation in deductibles.

By contrast with deductibles, numerous estimates of how coinsurance affects the demand for medical services have been made.† Much

---

*Arrow (1963) analyzed the case of perfectly inelastic demand and showed that deductibles--as opposed to coinsurance--were a desirable way to structure insurance for a given premium. Arrow (1976) subsequently analyzed the case of elastic demand with a constant unit price (coinsurance). He showed that to estimate the welfare effects of coinsurance, one must have a knowledge of the demand curve for medical care, but he did not explicitly treat insurance plans with nonconstant unit prices (e.g., deductibles). Keeler, Newhouse, and Phelps (1977), using both conventional welfare analysis (in particular, assuming that the market price of medical care represented its marginal social cost) and assumed values for demand response, showed that deductibles of several hundred dollars were likely to be optimal. Their conclusion would be reinforced to the degree that no deductibles and reimbursement at "usual, customary, and reasonable" levels lead to inefficiencies in production (Newhouse, 1977).

†See Feldstein, 1971; Davis and Russell, 1972; Phelps and Newhouse, 1974a; Rosett and Huang, 1973; Newhouse and Phelps, 1974; Newhouse, Phelps, and Schwartz, 1974; Newhouse and Phelps, 1976.

more work has been undertaken on coinsurance than on deductibles for both theoretical and empirical reasons. In the remainder of this section, we take up the theoretical reasons for the comparatively greater understanding of coinsurance than of deductibles; the empirical reasons will be discussed in Section 2.

Standard economic theory can accommodate coinsurance straightforwardly, because coinsurance simply multiplies the unit price by a parameter between zero and one.* A deductible, on the other hand, presents the consumer with a price schedule; he must pay 100 percent of the cost up to a certain level, but after that, the price is subsidized. Such a price schedule violates the conventional assumption that the consumer can buy as many units as he wants at a given price. In the usual (one-period) economic model of choice, a deductible leads to a kinked budget line (hyperplane) and two local maxima, one of which, in general, will be the global maximum.

Figure 1.1 illustrates this situation. The budget line is ABC; if the consumer purchases D units of medical care (the deductible), the relative price of medical care is reduced. The I_0, I_1, and I_2 curves are indifference curves; 0 and 0' represent local optima along budget line ABC. In this example, 0 is the global optimum; the consumer does not exceed the deductible.

The one-period model illustrated in Fig. 1.1 is a special case. The deductibles in most insurance policies are usually high enough so that the first consumption decision on medical care (e.g., an office visit) will not satisfy the deductible, although subsequent consumption decisions may cause the consumer to exceed the deductible. In this case, the maximizing decisionmaker (the physician, the consumer, or some combination) will assess, at the time of each consumption decision, the probability that the consumer will satisfy the deductible and make decisions accordingly. Thus, decisions will be sequential

---

*Indeed, as long as unit price is unchanged, the comparative statics of quantity demanded with respect to coinsurance are identical to those with respect to price, save for the effect of a change in coinsurance on the insurance premium, which leads to an empirically negligible income effect. (See Phelps and Newhouse, 1974b, Appendix A.)

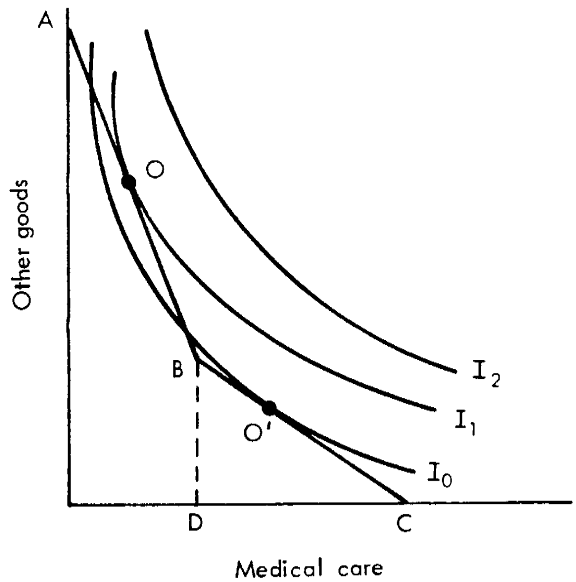

Fig. 1.1--Effect of a deductible in the usual one-period model

and will be made under uncertainty, but the theory illustrated in Fig. 1.1 does not contemplate such circumstances.

Keeler, Newhouse, and Phelps (1977) formally develop, as a dynamic programming problem, a model that incorporates sequential behavior under uncertainty. We briefly sketch this model here. If the consumer and/or his physician expects to exceed the deductible, the model predicts that the consumer's behavior will approximate that of an individual with no deductible.* If the consumer and/or his physician is uncertain about exceeding the deductible, the model predicts that the insured consumer will consume medical services at a somewhat greater rate than the uninsured, because the expected price of future consumption is reduced as total expenditure rises toward the deductible. If the consumer is quite certain of *not* exceeding the deductible, the model predicts that his behavior will approximate that of the uninsured.

---

*If it is anticipated that the consumer will exceed his deductible with certainty, the only difference caused by a deductible is the income effect that the individual with a deductible policy must bear (i.e., the deductible acts as a loss in income, but not as a change in price); this income effect is partially offset by the lower premium that one pays for insurance with a higher deductible.

It follows that variation in the deductible around zero, and variation in a deductible that is high enough so that it is unlikely to be exceeded, should cause little variation in demand. Because existing data on demand response to coinsurance show a substantial change in demand between uninsured and fully insured states (Newhouse, Phelps, and Schwartz, 1974), one may infer that the response of a given consumer to variation in a deductible will be sigmoid, as shown in Fig. 1.2.

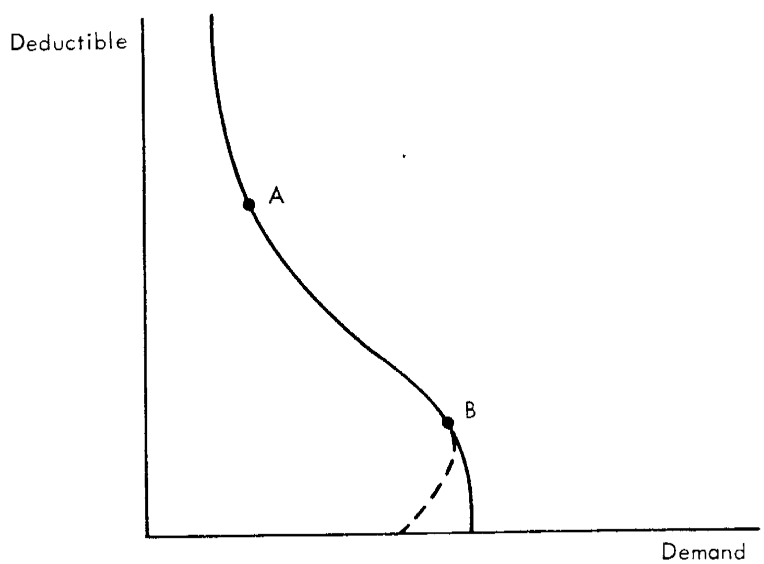

Fig. 1.2--Demand as a function of a deductible

The response of aggregate demand to variation in the deductible should also be sigmoid. Most consumers would expect to exceed a very low deductible, so their behavior would not differ much from the zero-deductible case. As the deductible rises, an increasing fraction of consumers expect not to satisfy the deductible. Thus, for an intermediate range of deductibles, one would anticipate that demand would respond to variation in a deductible, as these consumers change from behavior approximating full insurance to behavior approximating no insurance. At high deductibles, most individuals do not expect to satisfy the deductible, so variation in the deductible again causes

little effect on demand. This situation is depicted in Fig. 1.2 by the solid line. Between A and B, variation in the deductible exerts a substantial effect, but above A and below B there is little effect. Such "demand curves" are predicted by the model in Keeler, Newhouse, and Phelps (1977). One objective of the present report is to discover whether this behavior is observed in reality, and if it is, to say as much as possible about where A and B might lie, i.e., to estimate the solid curve shown in Fig. 1.2.

An objection to the shape of the solid curve in Fig. 1.2 can be made if deductibles deter preventive services and therefore cause higher total consumption of medical care (in the steady state) (Roemer et al., 1975). The assumptions necessary to establish this argument are as follows: (1) Certain preventive services are efficacious. (2) Individuals do not know the value of these preventive services and are unwilling to undertake them if they have to pay for them. (3) If, however, these preventive services were subsidized, individuals would consume appropriate amounts of them. If these assumptions are correct, the market demand curve shown in Fig. 1.2 should bend in toward the origin at low levels, as shown by the dashed line.* Thus, as deductibles rose, usage over a certain range would actually increase (in the steady state), because effective preventive care would not be demanded. We also wish to test for this possibility.

---

*Because preventive care may take time to exert an effect, any individual's demand curve would depend on prior consumption of preventive services (and, implicitly, on the number of years he had had a policy with a particular deductible). Figure 1.2 is intended to represent the steady-state demand curve if all had policies with a given deductible level for a considerable period of time.

Section 2

GENERAL APPROACH TO ESTIMATING DEMAND FOR MEDICAL CARE SERVICES

Empirical analysis of demand response to variation in a deductible differs substantially from analysis of demand response to variation in coinsurance. Perhaps most significantly, a claim with the insurer will generally be filed in the case of coinsurance, but claims are typically not filed by individuals who do not exceed their deductibles. (Indeed, lower filing rates are the principal reason why administrative costs of insurance policies with deductibles should be lower than those without deductibles.) Therefore, when one uses data from insurance claims to compare policies with differing deductibles, it is straightforward to compute the differential cost to the insurer; but it is impossible to compute the effect of the deductible upon total demand, because no information is available on the portion of the distribution lying below the deductible.

An alternative to the use of claims data for estimating demand as a function of deductibles is to use data from household surveys. While such data may permit an estimate of total demand (if total expense data as opposed to out-of-pocket expense data are gathered), they tend to have little information on the insurance policy. Moreover, health insurance policies vary on so many dimensions that existing surveys do not afford samples that are large enough for one to estimate a pure effect on a deductible. For example, some policies pay up to $A per day for hospital room and board, whereas others pay for a semiprivate room in its entirety for B days; still others may pay C percent of the cost after a $D deductible is satisfied. Additionally, some policies may pay up to $E for laboratory services, whereas others will pay for laboratory services in excess of $F; etc. Many surveys do not gather information on insurance policies at this level of detail, and those few that do, do not have a large enough sample for our purposes.

Our approach in estimating demand has been governed by our inability to find data describing expense distributions from policies differing only in the deductible amount. We have only been able to

obtain data on the *premiums* that two major health insurance companies charge to insure a specified group of individuals with policies that differ only in the size of the deductible. We have also obtained the *distribution of claims* for an actual insurance plan with a $50 deductible.

Given these data, we approached the problem as follows. We adjusted the data on premiums and claims to make them comparable. We treated the premium as proportional to the expected value of the insurer's payout (the insurers tell us that this assumption is correct). Thus, the premium data provided us with information on the size of the expense distribution lying above the deductible. We then combined this information with an assumption about the shape of the entire expense distribution (based in part on the claims data) so that we could infer a mean for the entire expense distribution as the deductible changed.

Our assumption about the entire expense distribution was that it comes from a family of distributions that is normally distributed after a Box-Cox power transformation of the raw data, except for a mass at zero expenditure. The Box-Cox transformation is $(y^{\lambda} - 1)/\lambda$ for $\lambda \neq 0$ and $\log(y)$ for $\lambda = 0$. The normal and the lognormal are special cases of this transformation in which λ equals 1 and 0, respectively. As λ decreases, the distribution of the raw data becomes heavier-tailed. The mean μ and the variance ν of the transformed distribution, as well as λ, are estimated from the claims distribution with a $50 deductible. Special account is taken of underreporting in the vicinity of the deductible, and the probability of zero expenditure is estimated by using extraneous data. As the deductible increases, this distribution is assumed to shrink multiplicatively (the probability density shifts toward the origin in a proportional fashion). The multiplicative factor is estimated from the premiums charged for policies with higher deductibles. Given the distribution of total medical expenses as a function of the deductible, means can be computed straightforwardly. Before turning to the details of the estimation procedure, however, we will describe the premium and claims data.

THE PREMIUM DATA

Two large national insurance companies gave us quotations in 1973 prices on the monthly premium they require to insure an employee group with the following characteristics: the median age is 40; half the employees are females; half the employees earn less than \$5000 annually and half earn between \$5000 and \$10,000 (1973 dollars); all live in the Los Angeles area. There were no special group discounts. Premiums were quoted at six levels of a deductible: \$50, \$75, \$100, \$200, \$500, and \$1000 per person per year. Premiums were quoted for employees only, as well as for the employee and all dependents. We were unable to make use of the information on dependents because our claims data cannot be aggregated into families.* Thus, our comparisons are for employees only. Data from a study of variation in coinsurance suggest that this restriction could cause our estimates of demand responsiveness to a deductible to be biased toward zero, because female dependents showed the greatest responsiveness of any subgroup to changes in coinsurance (Phelps and Newhouse, 1972). Data from that study suggest that responsiveness of employees to price may be some 20 percent less than that of the entire population (including dependents). The amount of understatement in our estimates is therefore probably nontrivial; however, we would not expect inclusion of dependents to change our qualitative conclusions.

The premiums were quoted for policies with both a 20-percent and 25-percent coinsurance rate above the deductible. There was a lifetime maximum of \$25,000, with a \$1000 automatic annual reinstatement of benefits. Pregnancy-related expenditures were not covered, and only 50 percent of neuropsychiatric expenditures were covered. The quotations given for this policy by the two companies are shown in Table 2.1.

The premiums are assumed equal to

$$(1 + \ell)(1 - c)\int_D^U (x - D)\, dF(x) \ , \qquad (2.1)$$

---

*In other words, the additional premium for dependents is not on a per-dependent basis, although the claims data are.

Table 2.1

MONTHLY INSURANCE PREMIUMS FOR POLICIES WITH VARYING DEDUCTIBLES AND COINSURANCE RATES, LOS ANGELES AREA, 1973

Deductible ($)	Company A: Employee Only		Company B: Employee Only	
	20-Percent Coinsurance	25-Percent Coinsurance	20-Percent Coinsurance	25-Percent Coinsurance
50	19.96	18.19	16.21	14.75
75	17.67	16.11	15.16	13.80
100	16.15	14.72	13.64	12.41
200	13.09	11.93	10.57	9.62
500	7.44	6.78	6.05	5.51
1000	3.93	3.58	3.42	3.11

where ℓ is a loading charge to cover administrative costs and profit, c is the coinsurance rate, F(x) is the cumulative distribution function of claims filed with the company for a given deductible and coinsurance rate, D is the deductible, and U is the upper limit.* In order to infer the parameters of F(x) from our data on premiums, we must make an assumption about the loading factor ℓ: we assume that ℓ is independent of the deductible level. This assumption is supported by verbal communication with the companies, who stated that a constant ℓ was used in giving us quotations.

THE CLAIMS DATA

The distribution of claims is taken from the largest group health insurance plan in the country, the Federal Employees Health Benefits Plan. A substantial portion of the federal employees are insured through a consortium of commercial insurance companies in a plan

*This formulation assumes that the accounting period for the upper limit is the same as the accounting period for the deductible (e.g., annual). In practice, upper limits are typically lifetime limits with automatic restoration of a certain amount (often $1000) each year. Thus, U should be the amount remaining in the individual's lifetime maximum. This is not observed, and we have used the lifetime maximums as though they were annual maximums. Because few individuals should be in the vicinity of their lifetime maximums (less than .02 percent in 1 year), and because the entire maximum will be restored upon evidence of "insurability," this approximation should introduce little error.

administered by the Aetna Life and Casualty Company. A 20-percent sample was taken of all claimants in this plan in 1970; the sample included 75,321 claimants. We have worked with a subsample of these claimants in order to enhance the comparability between the group upon which the premiums are based and the federal employee population. Our total sample is 11,737 claimants, which come from a population of 31,990 insureds.

We have made a number of adjustments to the distribution of claims to make them comparable to our premium data because of the differences in the population covered by the differences in the insurance policy. These adjustments are described in Appendix A.

Regardless of how successful our adjustments are, they will not be perfect. Moreover, the premiums contain a loading factor that must be taken into account before comparisons with the claims distribution can be made. Because we do not know the loading factor (and it clearly differs for the two companies), and because our adjustments leave an unknown amount of error, it seemed best to compute a loading factor for each company, such that $P_{i,50} = (1 + \ell_i)E_{50}$, where $P_{i,50}$ is the premium for a policy with a \$50 deductible that company i charges (i = 1,2) and E_{50} is the expected payout for a policy with a \$50 deductible given the adjusted claims distribution. This relationship provides an estimate of ℓ_i. The premiums charged for policies with higher deductibles were then deflated by the estimated value of $1 + \ell_i$ so that we could compare them with the adjusted claims distribution. This procedure implies that any multiplicative errors we have made in adjusting the claims distribution are embodied in the estimated loading factor and will not affect our subsequent estimates.

Our dollar figures are in 1973 Los Angeles area prices. Because medical care prices in Los Angeles are well above average, 1973 Los Angeles dollars are approximately equal to 1975 national dollars.* Thus, our estimated demand curves should be treated as being in approximately 1975 dollars.

---

*Data on city-worker family budgets show that the medical care budget in Los Angeles was some 22 percent above the national urban average in 1971 (Bureau of Labor Statistics, 1973). From 1973 to 1975, the national Consumer Price Index for medical care services rose to 24 percent.

Section 3

ESTIMATING THE MEDICAL EXPENSE DISTRIBUTION WHEN THE DEDUCTIBLE IS $50

To compare the distribution of expenses for $50 deductible insurance policies with policies containing other deductibles, we needed to characterize the expense distributions in a mathematically convenient form. Our approach was to estimate the expense distribution for a $50 deductible policy as a member of a rich parametric family of probability distributions. The expense distribution for insurance policies with other deductibles was then derived by estimating a parameter in this family of distributions. In this section, we describe our estimation of the probability distribution of expenses for a $50 deductible policy from the federal employees' data.

In choosing a distributional form for medical expenses, other workers have found that the lognormal or gamma distributions fit their data well.* Jackson (1968) has developed a methodology for testing whether observations come from one of these two distributions. However, there are no theoretical reasons to justify limiting our search to the gamma or lognormal forms. It would be desirable to estimate the expense distributions from a larger class that includes the lognormal and gamma distributions. This idea motivates our choice of the three-parameter family of Box-Cox power transformations (Box and Cox, 1964).

Let U be the random variable denoting nonzero expenditures. Define the power transformation (as a function of λ) by

$$
\begin{aligned}
Z_\lambda &= \frac{U^\lambda - 1}{\lambda} \qquad \text{if } \lambda \neq 0 \,, \\
&= \log U \qquad \text{if } \lambda = 0 \,.
\end{aligned}
$$

We assume that Z_λ has a normal distribution with mean μ and variance ν (written $Z_\lambda \sim N(\mu, \nu)$). This family of distributions of U includes the lognormal distribution when $\lambda = 0$, whereas the value $\lambda = \frac{1}{3}$ closely

---

*Friedman (1974); Keeler, Morrow, and Newhouse (1977).

approximates the gamma distribution.* Our model will allow us to estimate μ, ν, and λ simultaneously from the data and, as a by-product, permits us to test the hypotheses $\lambda = 0$ and $\lambda = \frac{1}{3}$. If p is the probability that an individual will have zero expenses for the year, the cumulative distribution function of yearly medical expenses Y is

$$P(Y \le y) = p \qquad \text{if } y = 0 ,$$

$$= p + (1 - p)\Phi\left(\frac{z - \mu}{\sqrt{\nu}}\right) \qquad \text{if } y > 0 , \qquad (3.1)$$

where $z = (y^{\lambda} - 1)/\lambda$ (or log y if $\lambda = 0$) and Φ is the standard normal distribution function:

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \exp\,(-x^2/2)\, dx .$$

Thus the density function† $f_Y(y)$ is‡

$$f_Y(0) = p ,$$

$$f_Y(y) = (1 - p)\,\frac{1}{\sqrt{2\pi\nu}} \exp\,[-(z - \mu)^2/2\nu]\,y^{\lambda-1} \qquad \text{if } y > 0 . \qquad (3.2)$$

The density function is depicted in Fig. 3.1.

---

*More precisely, the Wilson-Hilferty approximation implies that if U is gamma, then $Z_{\frac{1}{3}}$ is approximately normal. See Hinkley (1975) for a tabulation of the exact appropriate values of λ for varying shape parameters of the gamma distribution.

†Strictly speaking, this is a frequency function for Y = 0 and a density for Y > 0, but we use the term "density" to avoid awkwardness.

‡Initially, a one-third random sample of the claims was used for the exploratory data analysis leading to our final choice of a model, as given in Eq. (3.2). We fitted a four-parameter Box-Cox power transformation to the expense distribution above \$50. That is, for nonzero expenses U, we estimated λ_1 and λ_2 in

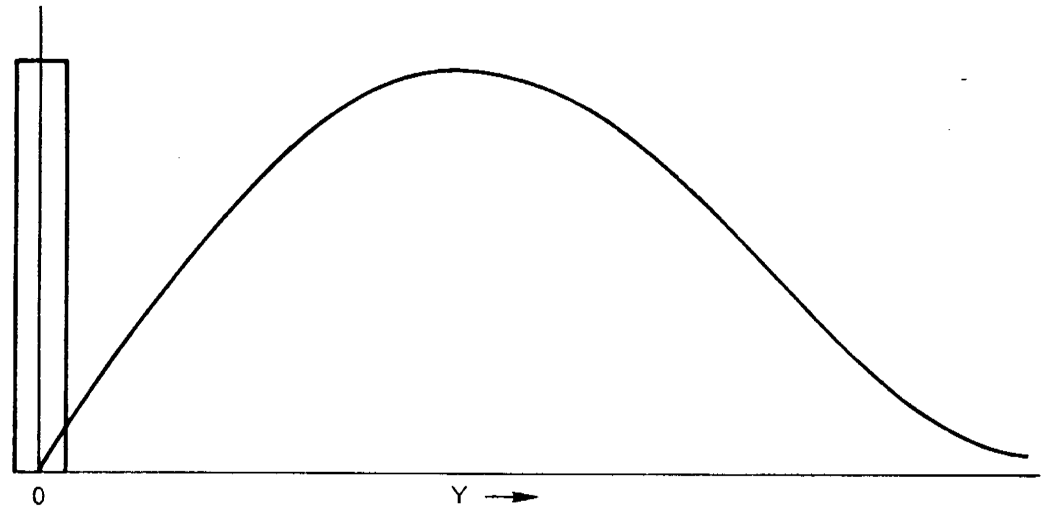

Fig. 3.1--Simple schematic of expense distribution

MAXIMUM LIKELIHOOD ESTIMATION OF THE EXPENSE DISTRIBUTION

As described below, we estimate μ, ν, and λ in the power transformation model by the method of maximum likelihood. Before writing down the likelihood function of the claims observations, two features of our claims data need special attention: (1) The lower end of the distribution is truncated because people with yearly expenses below the deductible do not file claims. (2) Because the high and low option of federal employees' plans treat expenses above \$20,000 differently, we have censored the observations at \$18,000; we have used the actual values of the claims below \$18,000 and the number of claims above \$18,000 to make our estimates.

The observed claims distribution has only values of Y above the \$50 deductible. Underreporting may exist in the region just above

---

$$[(U + \lambda_2)^{\lambda_1} - 1]/\lambda_1 \qquad (\text{or } \log(U + \lambda_2)) .$$

The lognormal distribution was also fitted to these data. Likelihood-ratio tests, chi-square goodness-of-fit tests, and various data plots were used to evaluate the fit and to compare it with the three-parameter (μ, ν, λ) family actually used. The fit of the four-parameter family to the data was not sufficiently better than that of the three-parameter family to justify the extra parameter; the lognormal family clearly was not rich enough.

this deductible. Consider the individual with \$60 of yearly medical expenses; with a \$50 deductible and a 20-percent coinsurance rate above the deductible, the insurer will repay the insured \$8 = (= 0.8 × (60 - 50)). This amount may not be a sufficient incentive for the individual to keep a record of his expenses, obtain the appropriate doctor's certification, and submit this claim. We have found that underreporting is likely to be confined to the region \$50 above the deductible (i.e., below \$100 total expenditure with a \$50 deductible). In particular, estimates of the parameters μ, ν, and λ in Eq. (3.2) are not sensitive to choosing a figure larger than \$100 and estimating the parameters only from claims exceeding the larger figure, whereas they are sensitive to choosing a lower figure (see Appendix B). The estimation of the portion of the distribution below \$100 is described later in this section (and in Appendix B).

ESTIMATION OF THE EXPENSE DISTRIBUTION ABOVE \$100

At the upper end of the claims distribution, the high option plan covers claims up to a \$50,000 lifetime maximum, and the low option plan covers claims up to a \$20,000 lifetime maximum. The high and low option plans reinstate \$2000 and \$1000 per year, respectively, toward this maximum when it has been depleted. Because we combine adjusted high and low option claims into one file, individuals with claims near the upper limit should respond differently, depending on their option. For this reason, claims above \$18,000 (\$2000 short of the upper limit for the low option) are counted only as being above \$18,000 rather than at their exact values. In effect, we assume that the variation in the upper limit has a negligible effect on behavior below annual expenses of \$18,000. (Only 0.02 percent of the individuals exceeded \$18,000 in annual expenditures.)

We can now write down the likelihood function for claims, taking into account their truncation* from below at \$100 and censoring from

*Because individuals who do not file claims could have either zero or nonzero expenses, we truncate and do not use such individuals in estimating the parameters μ, ν, and λ. This information is used separately for estimating the portion of the expense distribution below \$100.

above at \$18,000. Let n be the number (10,242) of claimants with expenditures above \$100 and less than \$18,000; let m(= 6) be the (random) number of claimants with expenses aboove \$18,000; and let $x_1, \ldots, x_n$ be the values of the n claims between \$100 and \$18,000. Then, from Eq. (3.2), the likelihood function is given by

$$L(x_1, \ldots, x_n, m \mid n + m, \mu, \nu, \lambda)$$

$$= [1 - \Phi(\xi_\ell)]^{-(n+m)} [1 - \Phi(\xi_u)]^m (2\pi\nu)^{-n/2} \exp\left[- \sum_{i=1}^{n} \frac{(z_i - \mu)^2}{2\nu}\right] \prod_{i=1}^{n} x_i^{\lambda-1} , \tag{3.3}$$

where Φ is the standard normal distribution function, $z_{x_i} = (x_i^\lambda - 1)/\lambda$ or $\log x_i$ if $\lambda = 0$ (abbreviated to z_i), $\xi_x = (z_x - \mu)/\sqrt{\nu}$, and $\ell = 100$ and $u = 18{,}000$. Note that z and ξ depend on (μ, ν, λ). For a given set of $x_1, \ldots, x_n$ and m, the likelihood function L in Eq. (3.3) can be maximized with respect to the parameters (μ, ν, λ). A brief description of our computational algorithm for doing this is given in Appendix C.

The first row of Table 3.1 gives the maximum likelihood estimates of the parameters (μ, ν, λ). These are the estimates we shall use in the remainder of the report. Also given in Table 3.1 are the results of testing whether $\lambda = 0$ and of testing whether $\lambda = \frac{1}{3}$ corresponding to a lognormal and gamma distribution, respectively. The fact that the

Table 3.1

PARAMETER ESTIMATES AND LIKELIHOOD RATIO TESTS ON λ

λ	Log Likelihood Function	μ	ν
0.11123	-64525.227	4.0807	15.8360
0	-64544.689	4.5161	3.0245
$\frac{1}{3}$	-64527.380	-433.2500	4204.2000

estimated value of λ, $\hat{\lambda} = 0.11123$, is between these values is consistent with the conclusion of some researchers that the lognormal distribution ($\lambda = 0$) is appropriate for medical expenses and the conclusions of others that the gamma distribution ($\lambda = \frac{1}{3}$) is appropriate for medical expense distributions. Minus twice the difference between the log likelihood at the null hypotheses ($\lambda = 0$ or $\lambda = \frac{1}{3}$) and at the maximum likelihood estimate ($\lambda = 0.11123$) has approximately a chi-square distribution with one degree of freedom under the null hypothesis. The hypothesis that $\lambda = 0$ is rejected at the 0.0001 level ($\chi^2 = 38.9$), and the hypothesis that $\lambda = \frac{1}{3}$ is rejected at the 0.05 level ($\chi^2 = 4.32$). Although the hypothesis $\lambda = \frac{1}{3}$ is not rejected as convincingly as $\lambda = 0$, the associated parameter values $\mu = -433$ and $\nu = 4204$ imply that all but the area that is eight standard deviations to the right of the mean is truncated at \$100. The probability mass beyond eight standard deviations is less than 10^{-9}, so that this model for $\lambda = \frac{1}{3}$ is not consistent with the proportion of expenses above \$100 under any reasonable model.

ESTIMATION OF THE EXPENSE DISTRIBUTION BELOW \$100

Inference about the expense distribution below \$100 is complicated by two factors: (1) underreporting of expenses between \$50 and \$100 and (2) no reporting of expenses below \$50. The situation may be described by referring to Fig. 3.2. The solid line depicts an assumed true density of *total* medical expenses, and the dashed line AB gives an observed density between \$50 and \$100 of *reported* expenses (claims that are filed).

Before describing our procedure for estimating the density $f_Y(y)$ of medical expenses below \$100, we need some additional notation. Let $n(a, b)$ be the number of individuals with *actual* expenses greater than or equal to a but less than b, and let $m(a, b)$ be the corresponding number of individuals with *reported* expenses (claims) in this interval. We know $m(50, 100)$ and $m(0, 50)$. Now, since $m(50, 100) < n(50, 100)$ (we assume only underreporting, not overreporting), the only useful information available about the probability distribution of actual expenses below \$100 is the fact that $m(0, 100) = n(0, 100)$. This one

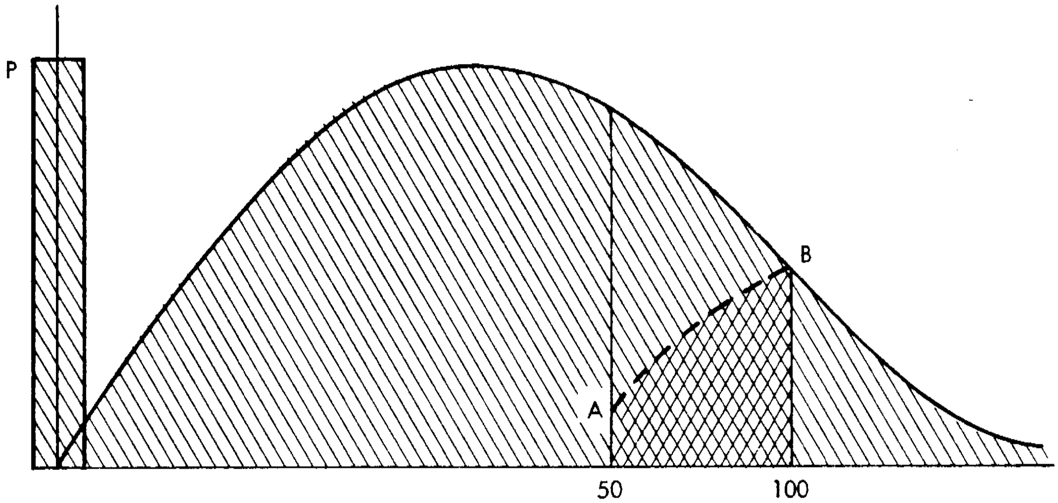

Fig. 3.2--Schematic of reported and total expense distributions

equation will, at best, allow us to estimate p, the probability of zero expenses.*

A straightforward way to estimate both p and the density between 0 and 100 that was ultimately *unsuccessful* is as follows: Assume that positive expenses between 0 and 100 have a power-transformed normal distribution with the parameters (μ, ν, λ) estimated by $(\hat{\mu}, \hat{\nu}, \hat{\lambda})$ from the data above 100, as described earlier. Then p, the probability of zero expenses, can be estimated by equating the ratio of the sample proportion of individuals above and below \$100 yearly expenses to the ratio of the probabilities

$$\frac{m(0,\ 100)}{m(100,\ \infty)} = \frac{p + (1 - p)\displaystyle\int_0^{100} \frac{1}{\sqrt{2\pi\nu}} \exp\ [-(z - \mu)^2/2\nu]x^{\lambda-1}\ dx}{(1 - p)\displaystyle\int_{100}^{\infty} \frac{1}{\sqrt{2\pi\nu}}\ \exp\ [-(z - \mu)^2/2\nu]x^{\lambda-1}\ dx} . \qquad (3.4)$$

---

*Observed values are m(0, 50) = 20,253; m(50, 100) = 1489; m(100, 18,000) = 10,242; m(18,000, ∞) = 6; thus m(0, ∞) = 31,990.

This results in an estimate of p as

$$1 - \frac{m(100, \infty)}{F(100, \infty)m(0, \infty)} , \qquad (3.5)$$

where

$$F(100, \infty) = \int_{100}^{\infty} \frac{1}{\sqrt{2\pi\nu}} \exp\,[-(z - \mu)^2/2\nu]x^{\lambda-1}\, dx ,$$

and

$$F(0, 100) = 1 - F(100, \infty) .$$

The difficulty with this approach is that we have no grounds to believe that the density between 0 and 100 dollars is a transformed-normal with parameters $(\hat{\mu}, \hat{\nu}, \hat{\lambda})$. Indeed the data appear to reject this specification, because the above method yields an estimate of p as -0.022 (using estimates of μ, ν, and λ developed from the data above \$100). Even if we were to "round off" this estimate to zero (and constrain p to be zero), our estimates would be inconsistent with the observation that a substantial proportion of the population spends nothing in a given year (Andersen et al., 1973).

Because our data gave us no satisfactory basis for estimating p, we adopted the strategy of using other data to estimate p and then adjusted the estimated density of expenses between 0 and 100 to be consistent with the estimated value of p. We chose the 1970 National Survey done by the Center for Health Administration Studies (CHAS) of the University of Chicago as being the best available source of data for estimating p, and deleted expenses that did not match the coverage of the federal employees' insurance data.* Table 3.2 gives the proportion of insured and uninsured populations that had no expenses. The relatively small difference in the proportion of zero expenses between

*To maintain comparability, observations were restricted to employed individuals, and expenditures for dental care, eye glasses, and drugs were excluded. See Andersen et al. (1973) for a description of the survey.

Table 3.2

PROPORTIONS OF INDIVIDUALS WITH NO MEDICAL EXPENSES

Individuals	Proportion with Zero Expenses	Number with Zero Expenses	Total Sample Size
Insured	27.5	877	3078
Uninsured	36.0	541	1278
Total	32.6	1418	4356

insured and uninsured individuals suggests that the effect of insurance on the probability that an individual will have no medical expenses is at most small, since the observed difference also contains the selection effect of sicker individuals tending to obtain insurance. For this reason, we use 0.275, the proportion of insured people in the CHAS survey with zero expenses, as our estimate of p for both our $50 deductible policy and the other higher-deductible policies we will describe later.

AN ADDITIONAL PARAMETER ESTIMATE NEEDED FOR THE TOTAL EXPENSE DISTRIBUTION

The parameters (μ, ν, λ) have been estimated by maximum likelihood from Eq. (3.3), using only reported expenses above $100. The value of p is taken as $p_0 (= 0.275)$ from external data. The remaining problem is to estimate the density of the total (as opposed to reported) expense distribution between 0 and 100. In order to obtain a probability distribution, we need to introduce an additional parameter for the portion of the expense distribution between 0 and 100. The probability distribution of medical expenses (Y) that we have used is given as follows:

$$
\begin{aligned}
f_Y(y) &= p_0 && \text{if } y = 0\ , \\
&= (1 - p_0)\frac{1}{\sqrt{2\pi\nu}} \exp\left[-(z-\mu)^2/2\nu - \beta(z_{100} - z)\right] y^{\lambda-1} && \text{if } 0 < y \le 100\ , \\
&= (1 - p_0)\frac{1}{\sqrt{2\pi\nu}} \exp\left[-(z-\mu)^2/2\nu\right] y^{\lambda-1} && \text{if } y > 100\ ,
\end{aligned}
\tag{3.6}
$$

where

$$z_{100} = \frac{100^{\lambda} - 1}{\lambda} \quad \text{and} \quad z = \frac{y^{\lambda} - 1}{\lambda} .$$

Note that the multiplicative factor exp $[-\beta(z_{100} - z)]$ applies only to y's between 0 and 100, as depicted in Fig. 3.3 where

$$g(y) = (1 - p_0) \frac{1}{\sqrt{2\pi\nu}} \exp\,[-(z - \mu)^2/2\nu]y^{\lambda-1} .$$

The dashed line gives the density of total expenses, $f_Y(y)$.

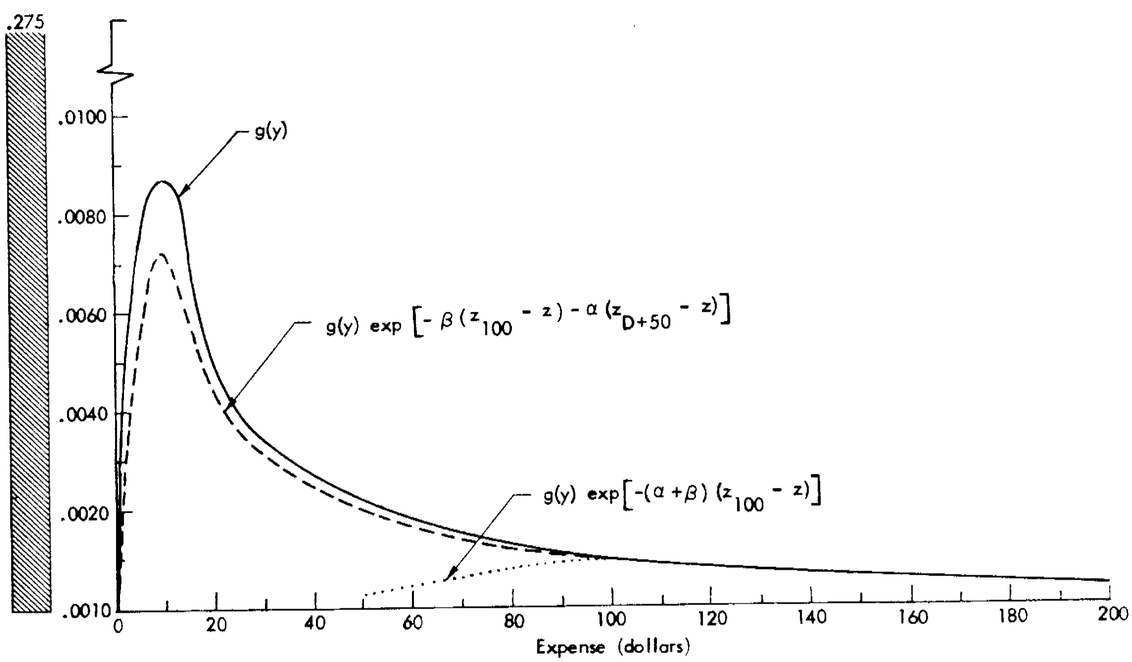

Fig. 3.3--Total expenses and underreporting factor for $50 deductible

The new parameter β is estimated by using the constraint that the density $f_Y(y)$ be a probability distribution and hence integrate to 1. Specifically, let

$$P_\beta(0,\ 100) = (1 - p_0) \int_0^{100} \frac{1}{\sqrt{2\pi\nu}} \exp\,[-(z - \mu)^2/2\nu - \beta(z_{100} - z)]y^{\lambda-1}\,dy$$

be the area between 0 and 100. After some algebra, this simplifies to

$$P_\beta(0,\ 100) = (1 - p_0)\Phi\left(\frac{z_{100} - (\mu + \nu\beta)}{\sqrt{\nu}}\right) \exp\,[-\beta(z_{100} - \mu) + \nu\beta^2/2]\ .$$

To estimate β, we solve the equation

$$p_0 + P_{\hat{\beta}}(0,\ 100) = \frac{m(0,\ 100)}{m(0,\ \infty)} \tag{3.7}$$

for the estimate $\hat{\beta}$.* The resulting value of 0.055 leads to a relatively small adjustment of the estimated probability distribution (see Fig. 3.3). To summarize, Table 3.3 gives the estimated values of the parameters of the \$50 deductible expense distribution, as well as its mean and standard deviation of the distribution.

Table 3.3

DISTRIBUTION OF TOTAL EXPENSES: \$50 DEDUCTIBLE

Parameter	p_0	μ	ν	λ	β	E(Y)	σ(Y)
Estimate	0.275	4.087	15.836	0.11123	0.055	180.08	784.34

THE REPORTED EXPENSE DISTRIBUTION

When computing the expected out-of-pocket costs to the insurance

---

*Strictly speaking, this does not necessarily have a solution. For values in our range of interest, a unique solution always exists.

company, we will need to estimate the distribution of reported expenses. To account for underreporting, we let the density of *reported* expenses (claims) X in the \$50 to \$100 range be

$$f_X(x) = \exp\,[-(\alpha - \beta)(z_{100} - z)]f_Y(x) \qquad \text{if } 50 < x \le 100 ,$$

where

$$z_{100} = (100^\lambda - 1)/\lambda \quad \text{and} \quad z = (x^\lambda - 1)/\lambda . \tag{3.8}$$

As can be seen by inspecting Fig. 3.3, the underreporting factor is then given by $\exp\,[-(\alpha - \beta)(z_{100} - z)]$. Using Eq. (3.8), the distribution of X becomes

$$\begin{aligned} f_X(x) &= \frac{(1 - p_0)}{\sqrt{2\pi\nu}} \exp\,[-(z - \mu)^2/2\nu - \alpha(z_{100} - z)]x^{\lambda-1} && \text{if } 50 < x \le 100 , \\ &= \frac{(1 - p_0)}{\sqrt{2\pi\nu}} \exp\,[-(z - \mu)^2/2\nu]x^{\lambda-1} && \text{if } x > 100 , \end{aligned} \tag{3.9}$$

with the remaining probability going to $0 \le x \le 50$ (no claim). This equation makes it clear that α can be estimated independently of β. The method for estimating α is the same as that for estimating β, namely, equating sample proportions to probabilities. Define

$$P_\alpha(50, 100) = \int_{50}^{100} \frac{(1 - p_0)}{\sqrt{2\pi\nu}} \exp\,[-(z - \mu)^2/2\nu - \alpha(z_{100} - z)]x^{\lambda-1}\, dx .$$

Some algebra leads to

$$P_\alpha(50, 100) = (1 - p_0)\left[\Phi\left(\frac{z_{100} - (\mu + \nu\alpha)}{\sqrt{\nu}}\right) - \Phi\left(\frac{z_{50} - (\mu + \nu\alpha)}{\sqrt{\nu}}\right)\right]$$
$$\times \exp\,[-\alpha(z_{100} - \mu) + \nu\alpha^2/2] .$$

To estimate α, we solve the equation

$$P_{\hat{\alpha}}(50,\ 100) = \frac{m(50,\ 100)}{m(0,\ \infty)}$$

to yield the estimate $\hat{\alpha} = 1.757$. The underreporting factor is $\alpha - \beta$, so it is estimated as $\hat{\alpha} - \hat{\beta} = 1.757 - 0.055 = 1.702$. This estimate implies that only around 15 percent of the individuals with \$50 of expenditures file claims, but this fraction rises rapidly as expenditures move toward \$100 (see the dotted line in Fig. 3.3).

We have now fully specified the estimated probability distribution for both total and reported yearly medical expenses for a \$50 deductible policy. Testing the fit of the reported expense distribution is discussed in detail in Appendix D. How this distribution changes as the deductible increases is described in Section 4.

Section 4

THE EFFECT OF VARYING DEDUCTIBLES ON EXPENSES: AN ESTIMATED DEMAND CURVE

Having adjusted the federal employees' claims data to make them comparable with the insurance company premiums, we now turn to comparisons of the premiums and the estimated expense distribution. Briefly, we will proceed as follows: The quotation for a policy with a \$50 deductible premium is compared with the expected payout by the insurer (computed by using the results of Section 3) to derive the amount of the premium attributable to the insurance company's loading factor. We assume that the loading factor is the same proportion of the premium across all policies and thus estimate the insurer's expected payout for each of the higher-deductible policies (\$75, \$100, \$200, \$500, \$1000). The form of the \$50-deductible distribution estimated in Section 3 is assumed to change in a specified way as the deductible increases by varying one new parameter. This new parameter is estimated by equating the insurer's expected payout under the fitted expense distribution to the expected payout implied by the insurance company's premiums. Having estimated all the parameters of the expense distribution, we then compute the expectation of yearly medical expenses for each deductible to give us a demand curve for medical care as the deductible varies.

ESTIMATING THE LOADING FACTOR

Because insurance companies only pay when a claim is submitted, we use the *reported* expense distribution to compute the insurer's expected payout. Let P_D be the premium quote for a deductible D and let L_D be the expected portion of reported expenses *covered* by insurance. The P_D and L_D are related by

$$P_D = L_D(0.8)(1 + \ell)(1.15) \quad , \tag{4.1}$$

where ℓ is the loading factor, the factor "0.8" occurs because the insurer pays 80 percent of all expenses above the deductible up to a

maximum of \$25,000,* and 1.15 represents inflation between 1970 and 1973. The loading factor ℓ (which includes systematic errors in the adjustments to render P and L comparable, as explained in the first section) can be estimated from Eq. (4.1) by taking D = 50. The formula for L_{50}, the insurance company's expected payout, is

$$L_{50} = \int_{50}^{25,000} (x - 50) f_X(x)\, dx$$

$$= \frac{(1 - p_0)}{\sqrt{2\pi\nu}} \left\{ \int_{z_{50}}^{z_{100}} [(\lambda z + 1)^{1/\lambda} - 50] \exp[-\alpha(z_{100} - z) - (z - \mu)^2/2\nu]\, dz \right.$$

$$\left. + \int_{z_{100}}^{z_{25,000}} [(\lambda z + 1)^{1/\lambda} - 50] \exp[-(z - \mu)^2/2\nu]\, dz \right\}, \qquad (4.2)$$

where the values for $(\mu, \nu, \lambda, \alpha)$ are the estimates given in Section 3. From Eqs. (4.1) and (4.2), the estimated loading factor is then computed as

$$\hat{\ell} = \frac{P_{50}}{0.8 L_{50}(1.15)} - 1 . \qquad (4.3)$$

Using the premiums given below, we list the estimated loading factors for two insurance companies in Table 4.1. These loading factors are consistent with existing knowledge of the magnitude of loading factors for small insured groups.† (Recall that there were no special group discounts in these premiums.) We infer that the adjustment

---

*As mentioned earlier, this is a lifetime maximum in the policy for which premiums were obtained, but we treat it as a yearly one in view of the reinstatement of the benefits provision of the policies.

†See Phelps (1973), p. 63.

Table 4.1

ESTIMATED LOADING FACTORS

Quantity	Company A	Company B
Annual premium (P_{50})	239.52	194.52
Covered expenses (L_{50})	157.15	157.15
Loading factor	0.657	0.345

process has been reasonably accurate, but we emphasize that any remaining systematic errors (i.e., multiplicative errors, such as an error in the inflation rate) will be captured by the factor $1 + \ell$ and so will not affect the estimated demand curve.

THE TOTAL AND REPORTED EXPENSE DISTRIBUTION AS THE DEDUCTIBLE VARIES

We have very little empirical evidence to guide us in choosing how the form of the medical expense distribution varies as a function of the insurance policy's deductible. As mentioned in Section 3, the CHAS data provide some evidence that the probability of having no yearly medical expenses varies little, if any, with the deductible size. Thus, our estimate of $p_0(D)$ will remain at 0.275 for all D. For the positive part of the expense distribution, the simplest assumptions are that the distribution changes in an additive or multiplicative fashion. The additive (as well as a transformed additive) assumption gave implausible results, whereas the multiplicative assumption gave plausible results that were consistent with the theory discussed in Section 1. We have therefore used the multiplicative assumption. With a multiplicative assumption, the model is defined as follows: Let Y_D be the expense distribution with deductible D. Thus, $Y_{50} = Y$ in our earlier notation. The distribution of Y_D is then assumed to be the same as $C(D)Y$, where $C(50) = 1$ and $C(D)$ will be estimated from expected insurance company payouts for higher deductibles.

We would expect $C(D)$ to be a decreasing function of D. That is, the density shrinks toward the origin as the deductible increases. This is equivalent to the demand curve's sloping downward. The density

of total expenses Y_D under the multiplicative model is given by

$$f_{Y_D}(t) = p_0 \qquad \text{if } t = 0 ,$$

$$= \frac{1}{C(D)} f_Y[t/C(D)] \qquad \text{if } t > 0 , \tag{4.4}$$

where $f_Y(t)$ is given in Eq. (3.6).

To estimate C(D), we need the distribution of *reported expenses*, because that is what the insurance company payout is based on. When working with reported expenses for higher deductibles, we must make an assumption about how underreporting varies as the deductible increases. We assume that the underreporting factor applies in the range \$50 above the deductible (that is, D to D + 50) and is the same for all deductibles. Thus, analogously to Eq. (3.8), the density of reported expenses X_D is given by

$$f_{X_D}(x) = f_{Y_D}(x) \exp\,[-(\alpha - \beta)(z_{D+50} - z)] \qquad \text{if } D \le x \le D + 50 ,$$

$$= f_{Y_D}(x) \qquad \text{if } x > D + 50 , \tag{4.5}$$

with the remaining probability being at x = 0.

Calculating the expected payout is a little more involved, and so we write out the formulae in full. Analogously to Eq. (4.2), we have (assuming $\beta \ge 0$)

$$L_D = \frac{1 - p_0}{\sqrt{2\pi\nu}} \int_{D/C(D)}^{25,000/C(D)} [C(D)y - D] \exp - \left\{(\alpha - \beta) \max\,[0, z_{(D+50)} - z_{yC(D)}]\right.$$

$$\left. + \beta \max\,[0, z_{100} - z_y] + \left(\frac{z_y - \mu}{2\nu}\right)^2 \right\} y^{\lambda-1}\, dy . \tag{4.6}$$

Thus, the adjustment for underreporting occurs within \$50 of the new deductible, and the β factor introduced to modify the distribution between \$0 and \$100 is kept. Note that this formula uses the distribution of reported expenses X_{50} and finds the appropriate expectation of $[C(D)X - D]$.* For deductibles D between \$50 and \$100, evaluating formula (4.6) is complicated. There are several possible cases to consider, depending on the relative positions of C(D)50, D, C(D)100, and D + 50. For D/C(D) above \$100, and for that part of the distribution more than \$50 above the deductible, the expression for L_D reduces to

$$L_D = \frac{(1 - p_0)}{\sqrt{2\pi\nu}} \int_{D/C(D)}^{25,000/C(D)} [C(D)y - D] \exp\left[-\frac{(z - \mu)^2}{2\nu}\right] y^{\lambda-1}\, dy \ .$$

To illustrate the complexity of the distribution of reported expenses X_D (as well as of total expenses Y_D), we present Fig. 4.1. The dashed line gives the density of Y_D, whereas the dotted line is where the density of X_D departed from that of Y_D. The solid line is the transformed normal density (it would be observed if $\beta = 0$). In this particular situation, in which $50 < D < C(D)100 < 100 < D + 50$, we have, for the nonzero part of the distributions of X_D and Y_D,

$$\begin{aligned} f_{Y_D}(y) &= g_D(y) \exp\left\{-\beta[z_{100} - z_y/C(D)]\right\} && \text{if } 0 < y \le 100C(D) \ , \\ &= g_D(y) && \text{if } y > 100C(D) \ , \end{aligned} \qquad (4.7)$$

---

*The integral is taken over Y rather than Y_D for mathematical convenience. The subscripts on the z variables may be explained as follows: The underreporting adjustment occurs for expenditures C(D)Y such that $D \le Y_D = C(D)Y \le D + 50$. The adjustment to fit the distribution for expenditures less than \$100 is stated in Y-space.

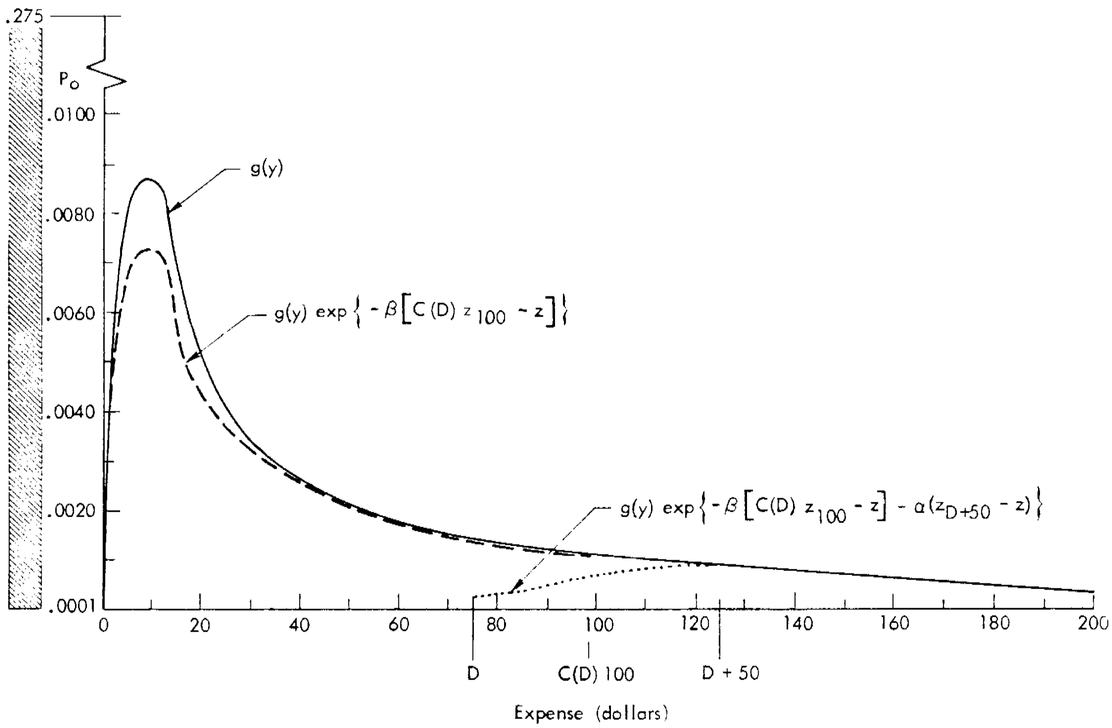

Fig. 4.1--Total expenses and underreporting factor for non-\$50 deductible

and

$$f_{X_D}(x) = g_D(x) \exp\left\{-\beta[z_{100} - z_x/C(D)] - (\alpha - \beta)[z_{D+50} - z]\right\}$$

$$\text{if } D \le x \le 100C(D) ,$$

$$= g_D(x) \exp\left\{-(\alpha - \beta)[z_{D+50} - z]\right\} \qquad \text{if } 100C(D) < x \le D + 50 ,$$

$$= g_D(x) \qquad \text{if } D + 50 < x , \tag{4.8}$$

where

$$g(x) = \frac{1 - p_0}{\sqrt{2\pi\nu}} \exp\left\{-[z_x/C(D) - \mu]^2/2\nu\right\} x^{\lambda-1} C(D)^{-\lambda} .$$

With the above formula for L_D, we find values of C(D) such that from Eqs. (4.1) and (4.3),

$$L_D = \frac{P_D}{0.8(1 + \hat{\ell})(1.15)} \tag{4.9}$$

for each insurance company. A trial-and-error search routine was used for each of the values D = 75, 100, 200, 500, 1000 to estimate C(D) from Eq. (4.9). Table 4.2 gives the resulting estimates of C(D).

Table 4.2

ESTIMATES OF MULTIPLICATIVE FACTORS

	Multiplicative Factor C(D)	
Deductible (D)	Company A	Company B
50	1.0	1.0
75	0.93	0.98
100	0.89	0.92
200	0.85	0.84
500	0.72	0.72
1000	0.64	0.66

THE DEMAND CURVE

The purpose of our calculations is to estimate the distribution of total medical expenditures for the range of deductibles: \$50, \$75, \$100, \$200, \$500, \$1000. We have now estimated all the parameters of these distributions, namely, μ, ν, λ, β, p, C(D). The density of expenses for an arbitrary deductible D is given in Eq. (4.7) and is illustrated in Fig. 4.1. To give some idea of how the densities compare as the deductible increases, we present them all on one graph in Fig. 4.2 for company B (loading factor = 0.345). The expense

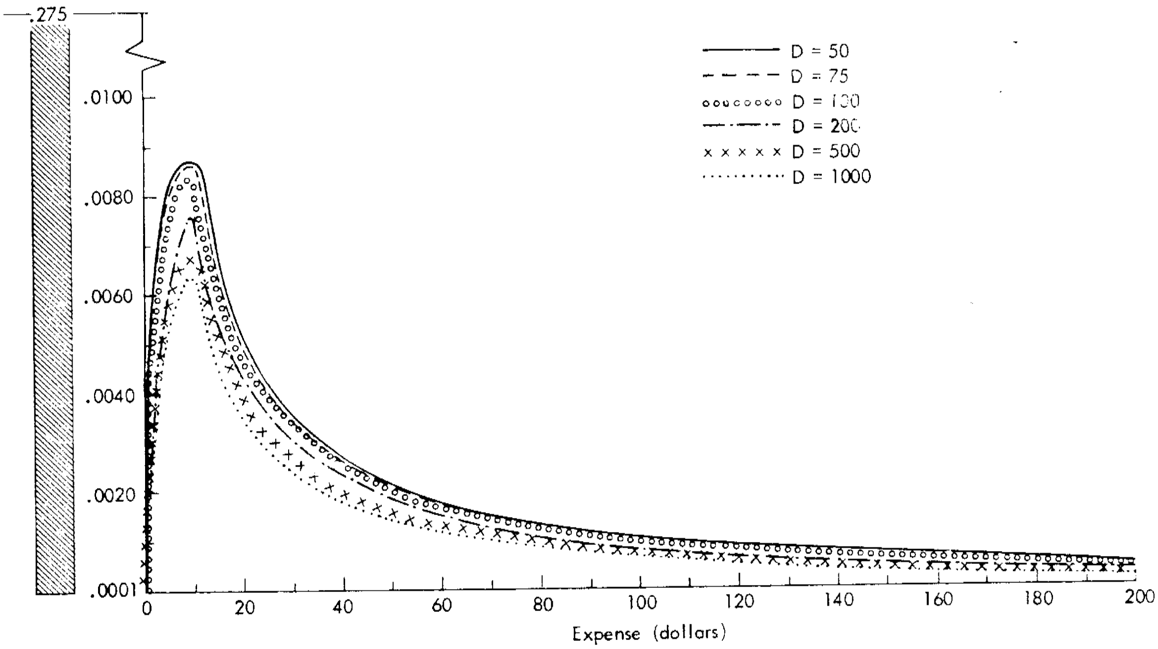

Fig. 4.2--Densities of total expense distributions in the \$0 to \$200 range

distribution is fairly heavy-tailed. This has a substantial effect on $E(Y_D)$, the expected value of yearly medical expenses for a policy with deductible D. For example, if the expense distribution were truncated at \$10,000 (all occurrence of expenses above \$10,000 reduced to \$10,000), the expected yearly expenses for an individual with a \$50 deductible policy from company A would be reduced from \$180.08 to \$164.97.

Payout maximums in insurance policies (\$25,000 in our case) will be binding for a few persons, raising an issue concerning demand above \$25,000. Many individuals do not have sufficient financial resources to pay bills on the order of, say, \$50,000. Such individuals have zero probability of spending such a large sum out-of-pocket. Their care may be financed through some third source, such as a research grant, or they may elect to forgo some quite expensive treatment. To account for such individuals, we have

computed expected yearly expenses under two assumptions. First, we have straightforwardly estimated demand, assuming that behavior is unchanged by the upper limit. Second, we have truncated the distribution of Y_D so that an individual cannot pay more than \$10,000 out of his own pocket (including the premium). This calculation assumes that no additional demand is forthcoming once \$10,000 is exceeded, and thus is a lower bound on the demand that the medical care system faces. For example, the distribution of Y_{100} is truncated for company A at \$29,727, because an individual's share (with a 20-percent coinsurance rate) is 193.80 (the premium) + 100 + 0.2(25,000 - 100) + 29,726.92 - 25,000 = 10,000. Table 4.3 gives both of these calculations for company A and company B. Fortunately the two estimates are not very different.

Table 4.3

EXPECTED YEARLY MEDICAL EXPENSES FOR VARYING DEDUCTIBLES

	Untruncated		Truncated at \$10,000[a]	
Deductible	Company A	Company B	Company A	Company B
50	180.08	180.08	178.08	178.08
75	167.47	176.47	165.90	174.61
100	160.27	165.67	158.91	164.16
200	153.06	151.26	151.90	150.15
500	129.65	129.65	129.00	129.00
1000	115.25	118.85	114.81	118.36

[a]The truncation points correspond to total expenditures ranging from \$29,153 to \$29,765.

Figure 4.3 shows a graph of $E(Y_D)$ in the untruncated case for companies A and B. Several findings are of note:

1. The demand curves estimated from the data for the two companies are very similar, despite markedly different absolute levels of premiums. This finding lends credibility to the results.

2. There is substantial responsiveness in demand to deductibles that vary between \$50 and \$200 per person per year (1975 dollars), and even considerable responsiveness to variation between \$200 and \$500.

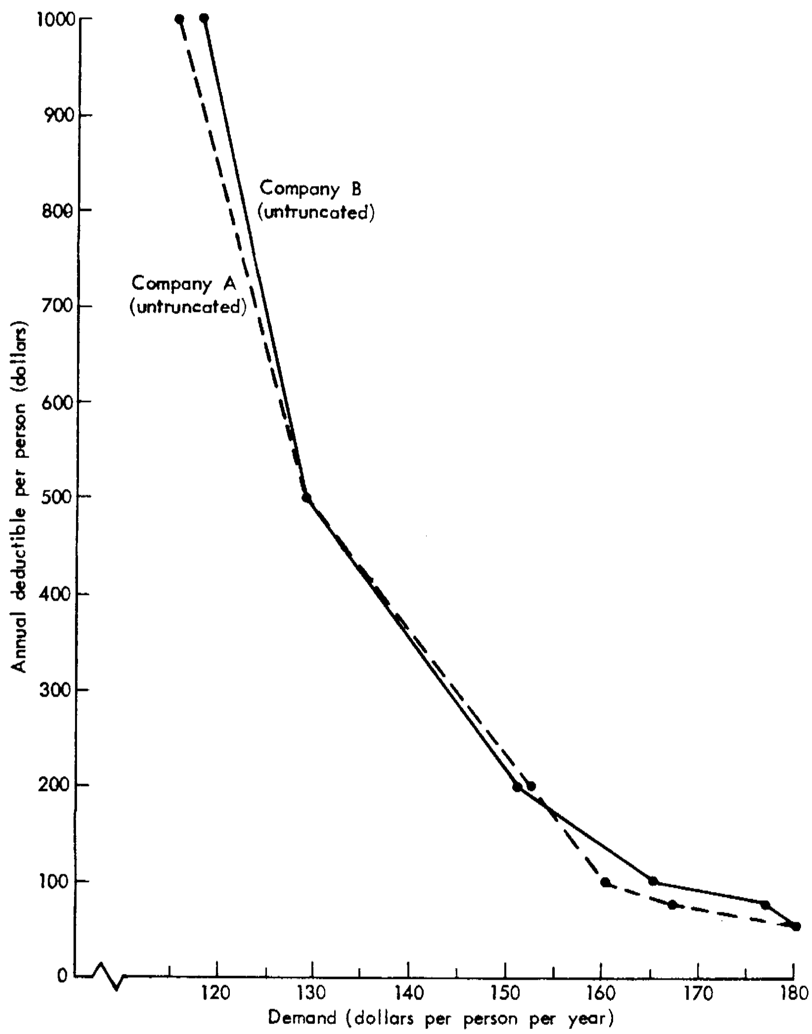

Fig. 4.3--Demand as a function of deductible

Demand at a $200 deductible is estimated to be around 85 percent of demand at a $50 deductible; raising the deductible to $500 reduces this demand to 72 percent. However, an additional increase to $1000 only reduces demand an additional 7 percentage points to 65 percent of the $50 level. Thus, in terms of Fig. 1.2, we estimate point A (the deductible above which demand becomes relatively unresponsive) to be at least $200 per person per year, and there is some support for a considerably higher figure. Demand shows a steadily diminishing response as the deductible rises above $75, consistent with the theory of demand presented in Keeler, Newhouse, and Phelps (1977).

3. For company B there is some evidence that demand becomes less responsive to variation in a deductible as the deductible decreases below \$75. We infer that point B of Fig. 1.2 (the deductible level below which demand becomes unresponsive) lies at or below \$75.

4. Insofar as one can make a comparison, our estimates are consistent with estimates of the responsiveness of demand to coinsurance. Those estimates show that demand at no insurance is roughly 50 percent of demand at full insurance (Newhouse, Phelps, and Schwartz, 1974). That we estimate demand at a \$1000 per person per year deductible to be around two-thirds of demand at a \$50 deductible is quite consistent with this range.

5. There is no evidence that the demand curve is changing sign in the region of \$50. Thus, there is no support in these data for the notion that deductibles deter preventive care, thereby raising medical expenses in the long run. However, the data are not well suited to test this hypothesis because one would like to observe individuals on a plan for a substantial period of time, and because one would like to observe behavior when care was free ($D = 0$). Because of movements into and out of plans, and because the lowest level deductible we observe is \$50, these conditions do not hold in our sample. Nonetheless, if positive deductibles had a substantial effect in raising expenditure, we would not expect to observe a steadily increasing response to demand as the deductible fell toward zero.

6. The relative effect of a deductible on the insurer's payout is substantially larger than its effect on demand. As the deductible rises, the insurer's payout is reduced not only because of reduced demand, but also because the insurer no longer covers some bills at all. Furthermore, on those bills that are covered, the insurer's share is reduced by the larger deductible. Table 4.4 shows the premium data of Table 1.1 recast in index form, showing how premiums fall with deductible for the two companies. The table could be interpreted as follows: If the cost to the federal budget of a tax-financed plan with a \$50 deductible and 20 percent coinsurance above the deductible were \$100 billion, then a plan with a \$200 deductible (and 20 percent coinsurance above the deductible) would require around \$65 billion in taxes, and a program with a \$1000 deductible would require around

$20 billion. In presenting these figures, we do not wish to imply that the federal budget share should be minimized; we only wish to point out that a change in deductible levels has a substantial impact on the budget.

Table 4.4

EFFECT OF INSURER'S SHARE OF VARIATION IN DEDUCTIBLE
(Index with $50 deductible = 100)

Deductible	Company A	Company B
50	100	100
75	89	94
100	81	84
200	66	65
500	37	37
1000	20	21

CONCLUSIONS

This report has explored the question of the effect of variation in a deductible on demand. Unfortunately the assumptions that we had to make to derive these estimates are quite strong, most notably that the proportional adjustment to the claims distribution accounted for all important omitted factors in reconciling the differences between the claims distribution and the premium data, and that the claims distribution changes in a multiplicative fashion as the deductible varies. The necessity for these assumptions only points up the need for better data.

Much better data from the Health Insurance Study will be available in the 1980s to confirm or reject these findings (Newhouse, 1974). That study observes families with deductibles that range from zero to 15 percent of income (maximum $1000) and collects data on all utilization (not just utilization above the deductible). Moreover, the minimum period of participation is 3 years (some participate 5 years), so that the deleterious effects of not consuming preventive care, if they exist, should appear.

Unfortunately results from the Health Insurance Study are some years away. As a result, the estimates presented in this report, tenuous though they may be, are significant. But improving their quality is important unfinished business.

Appendix A

ADJUSTMENTS TO THE AETNA FEDERAL EMPLOYEE DATA

The premium data were based on a hypothetical population assumed to live in the Los Angeles area and thus the federal employee claims data, a nationwide sample, had to be adjusted for cost differences between Los Angeles and the remainder of the country.* This was accomplished as follows: We found that over 50 percent of all claims in the federal employee sample occurred in nine states. We kept claims originating from these states in the sample and excluded claims from all other states. We then determined in which metropolitan areas of these states the largest percentage of federal employees lived. We used a cost index for inpatient and outpatient prices for that area as a deflator for claims from that state. This index was normalized so that the deflator for California was 1.0. As a result, all costs are approximately at the Los Angeles area level. Table A.1 gives a list of the states included in the calculations and the deflators that were applied to both inpatient and outpatient expenditures before the total expenditure distributions were computed. After data from the other states were deleted and the other exclusions described below were applied, there remained a sample of 11,737 claimants (i.e., those who reported expenditure in excess of the $50 deductible).

The federal employee insurance policy contains a high and low option. Both options specify a $50 deductible; however, the low option has a 25 percent coinsurance rate above the deductible, whereas the high option plan has a 20-percent coinsurance rate above the deductible. Pregnancy-related expenses and neuropsychiatric expenses are covered like any other expenditure. In the high option plan, the lifetime maximum is $50,000, and in the low option plan, it is $20,000. If the insured is hospitalized, the high option plan pays the first

---

*Note that cost differences encompass both differences in nominal price and differences in regional usage patterns not attributable to insurance plan or demographic characteristics.

Table A.1

PRICE DEFLATORS USED FOR FEDERAL EMPLOYEE DATA

State	Percent[a]	Deflator applied to-- Inpatient Expenditures	Outpatient Expenditures
Florida	(4)	.895	1.099
Maryland	(6)	.721	.692
Virginia	(8)	.785	.627
Texas	(7)	.699	.796
Washington	(4)	.552	.547
Oklahoma	(5)	.616	.662
New York	(3)	1.113	1.014
Washington, D.C.	(3)	.800	.600
California	(18)	---	---
Total	58	---	---

SOURCE: *Medicare: Health Insurance for the Aged--Geographic Index of Reimbursement by State and County, 1970*, U.S. Department of Health, Education, and Welfare, Social Security Administration. The distribution of federal employees enrolled in this plan is not the same as the distribution of all federal employees.

[a]The percentage of the sample of claimants from each state is shown in parentheses.

$1000 in its entirety (i.e., the deductible and coinsurance are only applied after the $1000 expenditure has been incurred); the low option plan pays the first $500 in its entirety.

Unpublished work of Bridger Mitchell, and additional evidence presented below, indicate that there is self-selection between the high and low option plans. To eliminate any effect of self-selection, we combined claims from both of the plans.*

We compared the combined claims distribution to the premiums quoted on a policy with 20-percent coinsurance above the deductible.

---

*We cannot account for the possibility of adverse selection by type of federal employee plan. However, the major alternative to the insurance plan we consider is a Blue Cross-Blue Shield plan that is quite similar; thus, we doubt that there is significant adverse selection by type of insurance plan.

To account for the effect of 25-percent, rather than 20-percent, coinsurance in the low option plan, we inflated the low option claims by 2 percent. This accounts for the elasticity of demand implied by the premium data shown in Table 2.1 (Phelps and Newhouse, 1974b).

The policy for which we obtained premiums does not cover pregnancy-related expenses. Therefore, we eliminated claims related to pregnancy from the federal employee claims. It is not possible to treat the difference in coverage of neuropsychiatric disorders as cleanly as the difference in coverage of pregnancy-related expenses, because the premium data include 50-percent coverage for such expenditures. We did not attempt a complex adjustment, but simply eliminated those claimants who submitted claims for neuropsychiatric disorders from the federal employee file. This should not significantly affect our estimates, because only 4 percent of all claimants were eliminated, so that even a markedly different distribution of claims among this group would not change the overall distribution very much. However, it means that the premium data are somewhat overstated.

An adjustment was then made for the complete coverage of the initial $1000 of room and board expenses ($500 in the low option plan). Such coverage would lead to higher demand among federal employees than among those insured by plans for which we had premiums. Our method was to estimate the amount by which inpatient demand would be lower, if the complete initial coverage of the federal employee plan were replaced by a $50 deductible, together with a 20- or 25-percent coinsurance above the deductible.

To estimate the amount of the reduction in demand, we disaggregated hospital expenses into admission rates, average lengths of stay, and average dollars per patient day. This was necessary because a change in these quantities induced by complete initial coverage has different effects on the expense distribution. A decrease in admission rates means that some hospitalizations do not occur at all. As a result, some expenses in the federal employee claims distribution would become zero. However, changed length of stay and changed dollars per day mean that expenses are somewhat less than they otherwise would be; i.e., the expense distribution shrinks toward zero. We used data contained in a

study by Heaney and Riedel (1970) to calculate the amount of the necessary change for admissions and length of stay. The data in the study indicate that changes of the following magnitude are likely.

In the data collected by Heaney and Riedel, the insurance changed from indemnity benefits, in which the insurance company paid $15 per day plus ancillary services, to service benefits with full coverage. We estimate that the $15 per day plus ancillary services benefit was the equivalent of a 31-percent coinsurance rate for admissions and a 57-percent coinsurance rate for length of stay (see Phelps and Newhouse, 1974b). Interpolating the observed changes in use linearly to the 20-percent-to-zero range of coinsurance, we derive the figures in Table A.2. Data from a study by Williams (1966) are also available. They show a larger effect of insurance on admissions and a smaller effect on length of stay than the data we used, but our conclusions would not change had our estimates been based on Williams' data.

Table A.2

ESTIMATED CHANGES IN HOSPITALIZATION RATES FROM CHANGE IN COINSURANCE

	Percent Change[a]
Admission rates	7.5
Length of stay	4.4
Dollars per patient day	Not available

SOURCE: Phelps and Newhouse (1974b), based on data in Heaney and Riedel (1970).

[a]Percent increase in use as the coinsurance rate changes from 20 percent to zero.

Based on the numbers shown in Table A.2, we have made "most likely" estimates as follows: We have assumed that there would be 7.5-percent fewer admissions than there actually are in the high option plan and 5-percent fewer admissions in the low option plan, and that these decreases in admissions would come randomly from the entire hospital expense distribution. In addition, we have assumed that anyone who is admitted to a hospital in the high option plan would spend 7.5

percent less if the complete initial coverage were not present (through a shorter stay, selection of a cheaper hospital, or fewer tests). That is, the change in the length-of-stay component (4.4 percent) and the dollars-per-patient-day component were assumed to total 7.5 percent, so that the hospital expense distribution was deflated by 7.5 percent. The corresponding figure for the low option plan was 5 percent. Ideally, the probability of no hospital expenses would have been increased 7.5 percent in the high option plan and 5 percent in the low option plan to account for an admissions effect. Because of a computational oversight, this was not done; instead, the probability of no expenses (in the entire distribution) was increased 5 percent. Because these adjustments are small, the error should be small, and it was felt that the cost of recalculation was not justified. Estimates based on these "most likely" adjustments are presented in the text. Some sensitivity analysis showed that the estimates are not sensitive to moderate variation in these figures.

We also adjusted for age and sex differences between the groups. Our procedure was to estimate expenses for the high and low option employee groups as a function of age and sex, and then to correct the claims distribution for the different age-sex mix. We regressed expenditure upon age (entered as a fifth-degree polynomial) and sex (entered as a zero-one dummy, one equals female). No age-sex interaction terms were entered. The resulting estimated equation for the high option plan was (absolute values of t-statistics in parentheses) as follows:

$$\text{Expense} = \underset{(2.22)}{4608} - \underset{(2.06)}{470.2(\text{Age})} + \underset{(2.07)}{19.70(\text{Age})^2} - \underset{(2.07)}{0.3930(\text{Age})^3}$$

$$+ \underset{(2.08)}{.003770(\text{Age})^4} - \underset{(2.09)}{.00001384(\text{Age})^5} - \underset{(1.50)}{30.41(\text{Sex})} .$$

The analogous equation for the low option plan was

$$\text{Expense} = \underset{(1.33)}{4238} - \underset{(1.38)}{529.9(\text{Age})} + \underset{(1.49)}{26.44(\text{Age})^2} - \underset{(1.58)}{.6208(\text{Age})^3}$$

$$+ \underset{(1.64)}{.006944(\text{Age})^4} - \underset{(1.69)}{.00002975(\text{Age})^5} + \underset{(2.05)}{60.13(\text{Sex})} .$$

Among employees, the mean age in the high option plan is 50, and 44 in the low option plan. (Note the implied adverse selection.) The high option plan is 31-percent female and the low option plan is 25-percent female. Using the above equations, we next estimated expenses for both the high and low option plans at their respective means for age and sex. We then divided these predicted expenses by predicted expenses at age 40, 50-percent female. The resulting ratios are 1.16 for the high option plan and 0.97 for the low option plan. The claims distributions were then deflated by these values.

The premium data are for the Los Angeles area in 1973, whereas the claims data are for 1970. As described in the text, we have used the medical-care-services component of the Consumer Price Index to convert both streams of data to (approximately) 1975 prices.

The distribution of federal employee claims includes claims from annuitants and dependents. These claims have been eliminated; annuitants' claims account for roughly 15 percent of the total claims, and dependents' claims account for just over half of them.

Finally, the federal employees have a considerably different income distribution from that of the group covered by the policy for which the premium was quoted. In the group for which premiums were supplied, half the employees were earning less than $5000 in 1973, and the remainder between $5000 and $10,000. No data are available on the income distribution of the federal employees insured through Aetna, but average earnings for federal employees covered by the General Schedule were $13,204 in 1973. An adjustment for this income difference was made as follows: A permanent-income elasticity of 0.2 was assumed.[*] The federal employees enrolled in the Aetna plan were assumed to have the same distribution of earnings as those covered by the General Schedule.[†] When comparison was made with the group for which premiums

---

[*]Newhouse and Phelps (1976) calculate an earned-income elasticity closer to 0.1; however, this measure is probably biased downwards as an estimate of permanent-income elasticity. An upward adjustment to 0.2 was made, which is roughly consistent with the findings of Andersen and Benham (1970).

[†]This distribution is given in United States Civil Service Commission (1973), Table 4.

were supplied, half of that group was assumed to earn $2500 and the other half, $7500. Purely on the basis of this difference in income, one would expect that the federal employees would spend 22.4 percent more on medical care. As a result, the distribution of claims for the federal employees has been deflated by that factor.

Appendix B

ACCOUNTING FOR THE UNDERREPORTING OF EXPENSES

Our data on the distribution of expenditures come from insurance policies with a \$50 deductible provision and a 20-percent coinsurance rate. For yearly expenses only slightly above the \$50 deductible, there is every reason to believe that some individuals will *not* file claims with the insurance company; e.g., a person with \$55 in annual medical expenses would only receive a \$4 (80 percent of \$55 - \$50) reimbursement, so he may not bother to file a claim (or he may have lost earlier bills and thus may not be able to prove a \$55 expense). Our procedure was first to estimate from the claims distribution the upper limit below which a significant amount of underreporting takes place--about \$100 in yearly expenses. We investigated various functional forms for modeling underreporting between \$50 and \$100 and concluded that the "exponential decay" in the transformed expenses given in Eq. (3.9) was satisfactory. We then applied this functional form for underreporting to get the estimated distribution of reported expenses for higher deductibles (e.g., for a deductible of \$200, the underreporting factor is applied to expenses between \$200 and \$250). We now describe these procedures in detail.

THE UPPER CUTOFF POINT

Figure B.1 gives some indication of apparent underreporting in the \$50 to \$150 range. The horizontal axis gives yearly expenses, and the vertical axis gives the expected and actual number of individuals whose expenses fall in the intervals \$50-\$55, \$55-\$60, ..., \$145-\$150. The expected number of individuals in an interval is calculated as follows: Maximum likelihood estimation is used in Eq. (3.3) for the N = 11,737 claims above \$100 to obtain an estimate of the density function of total expenses, $f_Y(y)$ in Eq. (3.2). The expected number of claims in interval (i, i + 5) is denoted by n_i and is defined as

$$n_i = N \int_i^{i+5} f_Y(y)\ dy \ . \tag{B.1}$$

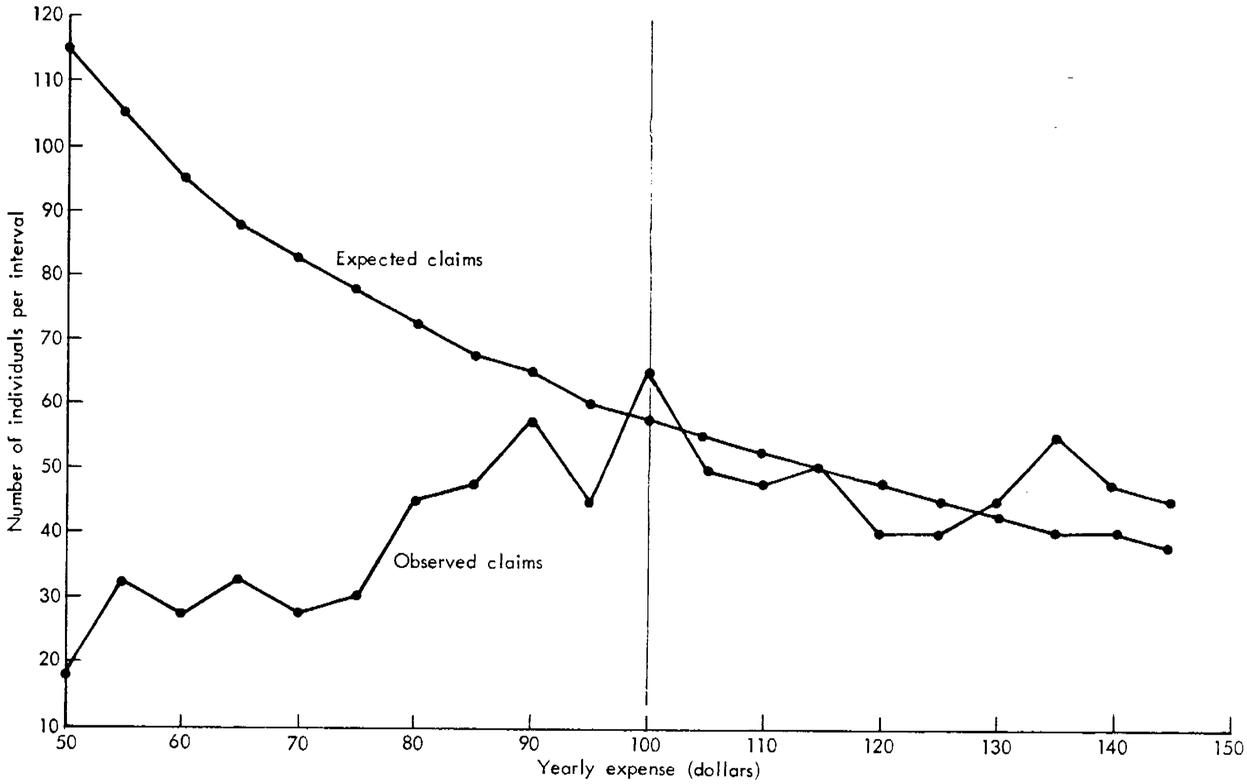

Fig. B.1--Expected and observed claims by $5 intervals

Table B.1 gives the values used for Fig. B.1. Notice from the figure that there is no evidence of systematic underreporting for claims above $100. From this and other computations using different cutoff points, we chose $100 as the upper limit for underreporting.

FUNCTIONAL FORM OF THE UNDERREPORTING FACTOR

We decided to search for an underreporting function u(x) so that the density of reported expenses between 50 and 100 is

$$u(x)(1 - p_0) \frac{1}{\sqrt{2\pi\nu}} \exp\left[-(z - \mu)^2/2\nu\right] y^{\lambda-1} \qquad \text{if } y > 0 \ . \qquad \text{(B.2)}$$

The reasoning just described leads us to require that $u(x) = 1$ for $x \geq 100$; in addition, we require that $u(x) > 0$ for $x > 50$ and that $u(x)$ be an increasing function. For $D = 50$, the situation with our

Table B.1

OBSERVED AND EXPECTED NUMBERS OF CLAIMS BETWEEN $50 AND $150

Interval	Observed Number	Expected Number	Ratio
50-55	17	114	0.15
55-60	33	104	0.32
60-65	27	96	0.28
65-70	32	89	0.36
70-75	27	82	0.33
75-80	29	77	0.38
80-85	46	72	0.64
85-90	47	68	0.69
90-95	57	64	0.89
95-100	45	60	0.75
100-105	65	57	1.14
105-110	49	54	0.90
110-115	48	52	0.93
115-120	49	49	1.00
120-125	41	47	0.88
125-130	40	45	0.89
130-135	44	43	1.03
135-140	54	41	1.32
140-145	48	39	1.22
145-150	45	38	1.19

data, we tried a number of parametric underreporting functions. These included $\exp[-\alpha(z_{100} - z_x)]$; $1 - \alpha(z_{100} - z_x)$; $\alpha \log(1 - z_{100} - z_x)$; $1 - \alpha(100 - x)$; and others. We used these functions in Eq. (B.2) after we had estimated the free parameter α from the data between 50 and 100 to predict observed numbers of claims using Eq. (B.1). Plots of the residuals of this fit led us to choose the underreporting factor given in the text, on the grounds of adequacy of fit and mathematical convenience. Thus, the density of reported expenses is given by $f_X(x)$ in Eq. (3.9). That is,

$$f_X(x) = \frac{(1 - p_0)}{\sqrt{2\pi\nu}} \exp\left[-(z - \mu)^2/2\nu - \alpha(z_{100} - z)\right] x^{\lambda-1} \qquad \text{if } 50 < x \le 100 \ ,$$

$$= \frac{(1 - p_0)}{\sqrt{2\pi\nu}} \exp\left[(z - \mu)^2/2\nu\right] x^{\lambda-1} \qquad \text{if } x > 100 \ .$$

As explained in the text, the underreporting factor for expenses between \$50 and \$100 is

$$\exp\left[-(\alpha - \beta)(z_{100} - z)\right] .$$

For higher deductibles, we use this same functional form:

$$\exp\left[-(\alpha - \beta)(z_{D+50} - z)\right] .$$

Appendix C

ALGORITHM FOR CALCULATING THE MAXIMUM LIKELIHOOD ESTIMATES

Maximum likelihood estimates (MLE) were calculated by treating the Aetna data as being censored above \$18,000 (u) and truncated below \$100 (ℓ). Letting $\underset{\sim}{y} = (y_1, \ldots, y_n)$ be the observations between ℓ and u, and letting m be the number of observations above u, with $N = n + m$ being the total sample size, the likelihood function for $\underset{\sim}{y}$ and m is given by

$$L(y, m; \mu, \nu, \lambda, N) = \frac{[1 - \Phi(\xi_u)]^m}{[1 - \Phi(\xi_\ell)]^N (2\pi\nu)^{n/2}} \exp\left[- \frac{\sum_{i=1}^{n} (z_i - \mu)^2}{2\nu} \right] \prod_{i=1}^{n} y_i^{\lambda-1} ,$$

where

$$z_{\lambda y_i} = \frac{y_i^\lambda - 1}{\lambda} \text{ (usually abbreviated } z_i) ,$$

$$\xi_u = \frac{z_{\lambda u} - \mu}{\sqrt{\nu}} , \qquad \xi_\ell = \frac{z_{\lambda \ell} - \mu}{\sqrt{\nu}} ,$$

$$\Phi(\xi) = \int_{-\infty}^{\xi} \varphi(t)\, dt , \qquad \varphi(t) = \frac{1}{\sqrt{2\pi}} \exp\,(-t^2/2) .$$

To get the MLE of (μ, ν, λ), take the logarithm of L:

$$\log L = m \log\,[1 - \Phi(\xi_u)] - N \log\,[1 - \Phi(\xi_\ell)] - \frac{n}{2} \log\,(2\pi)$$

$$- \frac{n}{2} \log \nu - \frac{\sum_{i=1}^{n} (z_i - \mu)^2}{2\nu} + (\lambda - 1) \sum_{i=1}^{n} \log y_i$$

and the partial derivatives of log L, which are

$$\frac{\partial}{\partial\mu} \log L = \frac{m\varphi(\xi_u)}{\sqrt{\nu}\,[1 - \Phi(\xi_u)]} - \frac{N\varphi(\xi_\ell)}{\sqrt{\nu}\,[1 - \Phi(\xi_\ell)]} + \frac{1}{\nu} \sum_{i=1}^{n} (z_i - \mu) ,$$

$$\frac{\partial}{\partial\nu} \log L = \frac{m\varphi(\xi_u)\xi_u}{2\nu[1 - \Phi(\xi_u)]} - \frac{N\varphi(\xi_\ell)\xi_\ell}{2\nu[1 - \Phi(\xi_\ell)]} - \frac{n}{2\nu} + \frac{\sum_{i=1}^{n} (z_i - \mu)^2}{2\nu^2} ,$$

$$\frac{\partial}{\partial\lambda} \log L = \frac{-m\varphi(\xi_u)\,\frac{\partial}{\partial\lambda} z_{\lambda u}}{\sqrt{\nu}\,[1 - \Phi(\xi_u)]} + \frac{N\varphi(\xi_\ell)\,\frac{\partial}{\partial\lambda} z_{\lambda\ell}}{\sqrt{\nu}\,[1 - \Phi(\xi_\ell)]}$$

$$- \frac{1}{\nu} \sum_{i=1}^{n} (z_i - \mu) \frac{\partial}{\partial\lambda} z_i + \sum_{i=1}^{n} \log y_i ,$$

where

$$\frac{\partial}{\partial\lambda} z_{\lambda y} = \frac{1}{\lambda^2} [(\lambda z + 1) \log (\lambda z + 1) - \lambda z] .$$

Maximum likelihood estimates of (μ, ν, λ) for log L were calculated by using the Fletcher-Powell (1963) method of function minimization with the Davidon algorithm for estimating the second-order partial derivatives at each iteration based on the analytic first derivatives. The function that is minimized is the negative of the log-likelihood function.

Appendix D

TESTING THE FIT OF THE REPORTED EXPENSE DISTRIBUTION

The probability distribution of reported expenses is given in Eq. (3.9) and is defined for expenses greater than \$50. For expenses less than \$50, and for unreported expenses between \$40 and \$100, let

$$Q(0,\ 100) = p_0 + (1 - p_0)\Phi\left(\frac{z_{100} - \mu}{\sqrt{\nu}}\right) - P_\alpha(50,\ 100) \ .$$

Q(0, 100) is then the probability of not filing a claim. A chi-square goodness-of-fit test is used to test the fit of the data to the reported expense distribution.

Expenses are divided into 63 intervals. Table D.1 lists the intervals and, within each interval, the observed and expected number of claims. Figure D.1 is a plot of the observed (solid line) and expected (dashed line) frequencies against expense intervals. The expected number of claims for expenses between a and b is defined as

$$n_{ab} = N \int_a^b f_X(x)\ dx \ ,$$

where $f_X(x)$ is given in Eq. (3.9) and N is the number of employees enrolled in the insurance plan. The expected number of employees not filing a claim is therefore

$$n_0 = (N)\ [Q(0,\ 100)] \ .$$

The hypothesis that observed expenses come from the reported expense distribution is rejected at the 0.0001 level ($\chi^2 = 132.2$ with 58 degrees of freedom). Though the difference is significant under a chi-square test, as shown in Table D.1, much of the difference

Table D.1

OBSERVED AND EXPECTED NUMBER OF CLAIMS FOR ALL EMPLOYEES

Upper Limit for Interval	Observed Number	Expected Number	Contribution to Chi-Square
50	20257	20264	0.0025
75	576	524	5.0895
100	909	957	2.3595
125	967	1071	10.1657
150	862	852	0.1270
175	742	699	2.6490
200	673	587	12.5167
225	530	502	1.5250
250	495	436	8.0414
275	405	382	1.3258
300	359	339	1.1893
325	301	303	0.0104
350	258	272	0.7587
375	234	247	0.6363
400	203	224	2.0242
425	201	205	0.0805
450	167	188	2.3991
475	168	173	0.1735
500	146	160	1.2954
600	453	537	13.1228
700	363	413	6.0795
800	284	328	5.8005
900	259	266	0.1773
1000	208	220	0.6324
1100	207	184	2.7550
1200	168	157	0.8054
1300	156	135	3.3829
1400	106	117	0.9862
1500	114	102	1.4109
1600	108	90	3.7046
1700	77	79	0.0777
1800	85	71	2.8575
1900	53	63	1.6875
2000	54	57	0.1517
2100	55	51	0.2518
2200	66	47	8.1292
2300	43	42	0.0092
2400	44	39	0.7485
2500	40	35	0.6178
2600	28	32	0.6052
2700	20	30	3.2431
2800	40	27	5.6955
2900	31	25	1.2282
3000	28	24	0.8526
3500	109	95	2.0433
4000	77	68	1.1819
4500	54	50	0.2710
5000	43	38	0.6042
5500	29	30	0.0119
6000	24	23	0.0154
7000	34	34	0.0001
8000	23	23	0.0010
9000	19	16	0.5433
10000	6	12	2.6760
11000	9	9	0.0213
12000	4	6	0.9415
13000	3	5	0.7827
14000	2	4	0.9293
15000	2	3	0.3972
16000	1	2	0.8758
17000	0	2	2.0135
18000	2	2	0.0755
25000	6	10	1.4371

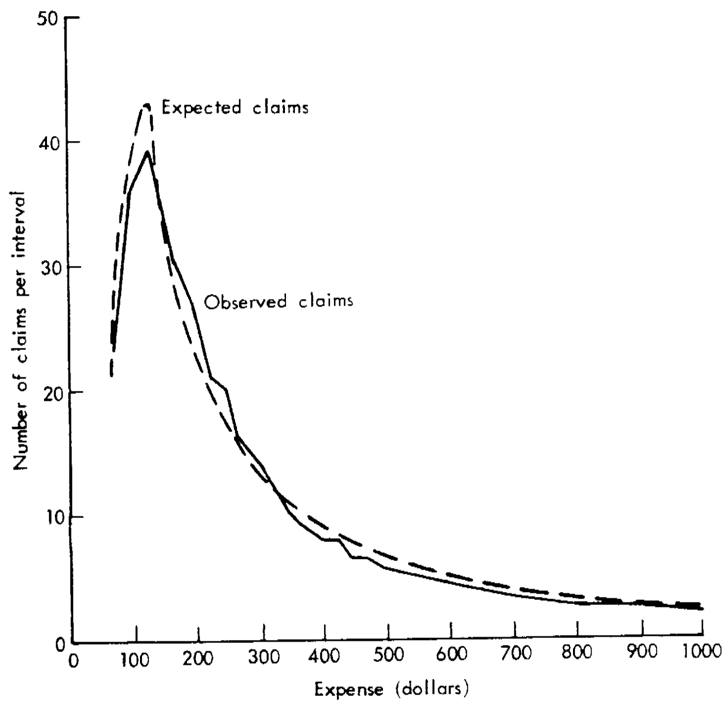

Fig. D.1--Observed and expected number of claims for all employees

(contribution to χ^2) is due to a few intervals. Figure D.1 shows these large differences and shows the fluctuations in the observed frequency of claims. A better fitting model of the reported expenses is not possible with the three-parameter model given here because of the large fluctuations. But, for purposes of this report, the three-parameter model provides a good estimation of the reported expense distribution. (The observed large fluctuations make it unlikely that *any* smooth parametric distribution could fit the data well.)

Further examination of the claims data showed significant differences in the distribution of expenses reported by employees with individual coverage and the distribution of those with family coverage. For example, the proportion of employees not filing a claim is 0.81 for those with individual coverage and 0.41 for those with family coverage. The chi-square test was used to test the fit separately for individual and family plans. Whereas the distribution for family

plans yielded highly significant differences statistically, the distribution for individual plans showed differences that were only marginally significant ($\chi^2_{63} = 83.0$, $p = .038$).* Many of the large contributions to chi square in testing the fit for the total sample are apparent when testing the fit for the family plans. Hence, we give a reestimation of the parameters (μ, ν, λ, p_0, β, α) for only the employees on individual plans and their resultant demand curve. These results should give the reader some feel for the robustness of our conclusions from the analysis of the total sample.

THE EXPENSE DISTRIBUTION FOR EMPLOYEES WITH INDIVIDUAL COVERAGE

Maximum likelihood estimates for μ, ν, and λ were calculated for the employees with individual coverage by using Eq. (3.3) and the algorithm described in Appendix C. Equation (3.7) was solved to estimate β.† The value of $p_0 = .244$ was taken from the 1970 CHAS survey for individuals with "self only" coverage. To estimate α, Eq. (3.10) was solved. Table D.2 gives the estimates parameters of the expense distribution for the employees with individual plans and its mean and

Table D.2

DISTRIBUTION OF TOTAL EXPENSES FOR INDIVIDUAL COVERAGE--$50 DEDUCTIBLE

Parameter	p_0	μ	ν	λ	β	α	E(Y)	(Y)
Estimate	.244	4.08	15.84	.108	-.03	2.15	204	895

standard deviation. Note that μ and ν are almost identical to the values given in Table 3.3 for the total sample.

The chi-square test was used to test the fit of the estimated reported expense distribution for employees with individual coverage.

---

*Strictly speaking, since the individual-plan data are used in the overall fit, the "effective" number of degrees of freedom is somewhat between 63 and 58.

†For the individual plans, sample sizes are $m(0, 50) = 14{,}369$, $m(50, 100) = 364$, $m(100, \infty) = 3001$, and thus $m(0, \infty) = 17{,}734$.

Figure D.2 plots the observed (solid line) and expected (dashed line) number of claims against expense. The hypothesis that the estimated reported expense distribution fits the observed expenditures cannot be rejected at the 0.05 level (χ^2 = 77.5 with 58 degrees of freedom). Although the reported expense distribution is fitted more precisely for the employees with individual coverage than for the total sample, changes in the distribution above 100 are minimal (as evidenced by minor changes in the parameters μ, ν, and λ). Changes in the distribution between 0 and 100 are more significant because of a large change in the value of p_0.

The demand curve for medical care as the deductible varies was estimated as in Section 4. An estimate for L_{50} (= 177.20) was calculated from Eq. (4.2), and the loading factor was then estimated by

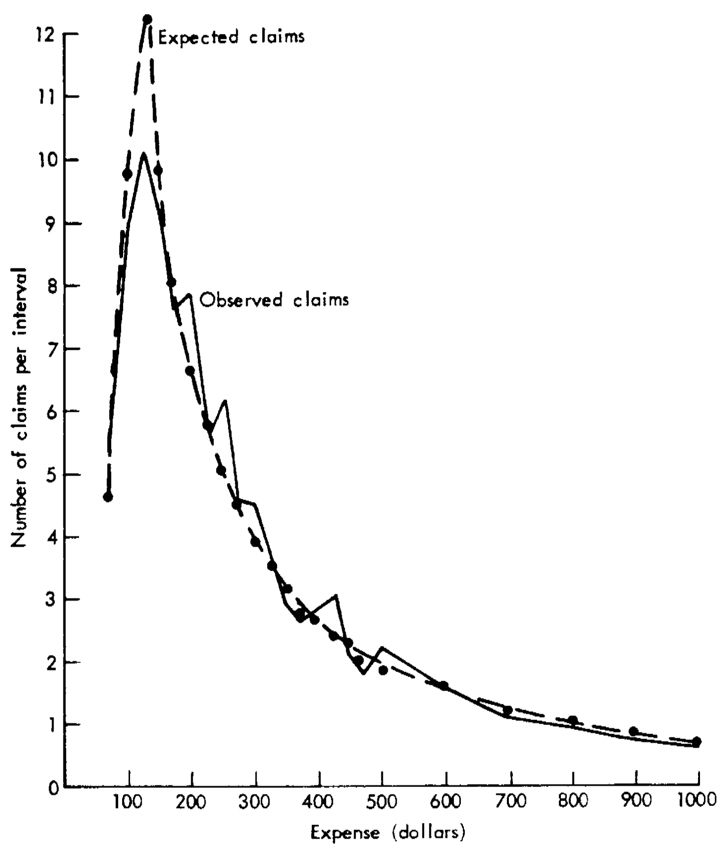

Fig. D.2--Observed and expected number of claims for employees with individual coverage

Eq. (4.3) ($\hat{\ell}$ = .469 and .193 for insurance companies A and B, respectively). Values for C(D) were calculated from Eqs. (4.6) and (4.9) and are shown for insurance companies A and B in Table D.3.

Table D.3

ESTIMATES OF MULTIPLICATIVE FACTORS

Deductible (D)	Multiplicative Factor C(D)	
	Company A	Company B
50	1.0	1.0
75	0.93	0.98
100	0.89	0.92
200	0.83	0.83
500	0.70	0.70
1000	0.60	0.63

The expected yearly medical expenses for varying deductibes were estimated by using the density of expenses defined in Eq. (4.7). Table D.4 shows the expected yearly medical expenses for varying deductibles for insurance companies A and B.

Table D.4

EXPECTED YEARLY MEDICAL EXPENSES FOR VARYING DEDUCTIBLES

Deductible	Untruncated		Truncated at \$10,000[a]	
	Company A	Company B	Company A	Company B
50	204.07	204.07	200.80	200.80
75	189.78	199.98	187.17	196.91
100	181.62	187.74	179.35	185.23
200	169.37	169.37	167.56	167.57
500	142.84	142.84	141.81	141.81
1000	122.44	128.56	121.81	127.82

[a]The truncation points range in total expenditures from \$29,153 to \$29,765.

A comparison of Table D.4 with Table 4.3 shows that the general shape of the demand curves estimated for employees with individual coverage are very similar to those for all employees. The major difference is a shift to the right of the demand curve for individual plans. Figure D.3 gives a graph of $E(Y_D)$ in the untruncated case for company B for all employees (solid line) and for employees with individual coverage (dashed line). The findings noted on pages 32 through 35 also hold for the employees with individual coverage.

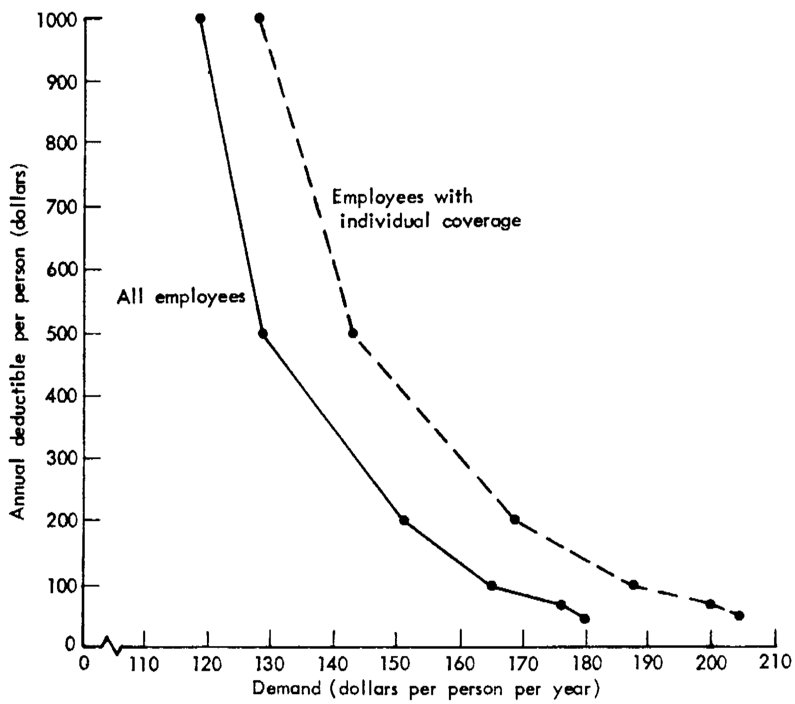

Fig. D.3--Demand as a function of deductible
(company B, untruncated)

REFERENCES

Andersen, Ronald M., and Lee Benham, "Factors Affecting the Relationship between Family Income and Medical Care Consumption," in Herbert Klarman (ed.), *Empirical Studies in Health Economics*, The Johns Hopkins Press, Baltimore, 1970.

Andersen, Ronald M., et al., *Expenditures for Personal Health Services: National Trends and Variations, 1953-1970*, Government Printing Office, Washington, D.C., 1973 (DHEW Publication No. (HRA) 74-3105).

Arrow, Kenneth, J., "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, Vol. 53, No. 5, December 1963, pp. 941-973.

-----, "Welfare Analysis of Changes in Health Coinsurance Rates," in Richard Rosett (ed.), *The Role of Health Insurance in the Health Services Sector*, National Bureau of Economic Research, New York, 1976.

Box, G.E.P., and D. R. Cox, "An Analysis of Transformations (with Discussion)," *Journal of the Royal Statistical Society*, Series B, Vol. 26, 1964, pp. 211-252.

Bureau of Labor Statistics, *Handbook of Labor Statistics, 1973*, Government Printing Office, Washington, D.C., 1973.

Davis, Karen, and Louise B. Russell, "The Substitution of Hospital Outpatient Care for Inpatient Care," *Review of Economics and Statistics*, Vol. 54, 1972, pp. 109-120.

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

Fletcher, R., and M.J.D. Powell, "A Rapidly Convergent Descent Method for Minimization," *Computer Journal*, Vol. 6, No. 2, 1963, pp. 163-168.

Friedman, Bernard, "Risk Aversion and the Consumer Choice of Health Insurance Options," *Review of Economics and Statistics*, Vol. 56, No. 2, May 1974, pp. 209-214.

Heaney, Charles T., and Donald C. Riedel, *From Idemnity to Full Coverage: Changes in Hospital Utilization*, Blue Cross Association, Research Series 5, Chicago, October 1970.

Hinkley, D. V., "On Power Transformation to Symmetry," *Biometrika*, Vol. 62, 1975, pp. 101-111.

Jackson, O.A.Y., "Some Results on Tests of Separate Families of Hypotheses," *Biometrika*, Vol. 55, 1968, pp. 355-364.

Keeler, Emmett B., Daniel Morrow, and Joseph P. Newhouse, "The Demand for Supplementary Health Insurance, or Do Deductibles Matter?" *Journal of Political Economy*, Vol. 85, August 1977.

Keeler, Emmett B., Joseph P. Newhouse, and Charles Phelps, "Deductibles and the Demand for Medical Services: The Theory of a Consumer Facing a Variable Price Schedule Under Uncertainty," *Econometrica*, Vol. 45, April 1977, pp. 641-655.

Newhouse, Joseph P., "A Design for a Health Insurance Experiment," *Inquiry*, Vol. 11, No. 1, March 1974, pp. 5-27.

-----, *The Erosion of the Medical Marketplace*, The Rand Corporation, R-2141-HEW, August 1977.

Newhouse, Joseph P., and Charles E. Phelps, "Price and Income Elasticities for Medical Care Services," *The Economics of Health and Medical Care: Proceedings of a Conference Held by the International Economies at Tokyo*, Macmillan & Co. Ltd., London, 1974.

-----, "New Estimates of Price and Income Elasticities," in Richard Rosett (ed.), *The Role of Health Insurance in the Health Services Sector*, Columbia University Press, New York, 1976.

Newhouse, Joseph P., Charles E. Phelps, and William B. Schwartz, "Policy Options and the Impact of National Health Insurance, *New England Journal of Medicine*, Vol. 290, No. 24, June 13, 1974, pp. 1345-1359.

Phelps, Charles E., *Demand for Health Insurance: A Theoretical and Empirical Investigation*, The Rand Corporation, R-1054-OEO, July 1973.

Phelps, Charles E., and Joseph P. Newhouse, "Effects of Coinsurance: A Multivariate Analysis," *Social Security Bulletin*, Vol. 35, No. 6, June 1972, pp. 20-29.

-----, "Coinsurance, the Price of Time, and the Demand for Medical Service," *Review of Economics and Statistics*, Vol. 56, No. 3, August 1974a, pp. 334-342.

-----, *Coinsurance and the Demand for Medical Services*, The Rand Corporation, R-964-1-OEO.NC, October 1974b.

Roemer, Milton I., et al., "Copayments for Ambulatory Care: Penny Wise and Pound Foolish," *Medical Care*, Vol. 13, June 1975, pp. 457-466.

Rosett, Richard N., and Lien Fu Huang, "The Effect of Health Insurance on the Demand for Medical Care," *Journal of Political Economy*, Vol. 81, March/April 1973, pp. 281-305.

United States Civil Service Commission, Bureau of Manpower Information Systems, Manpower Statistics Division, "Pay Structure of the Federal Civil Service," Government Printing Office, Washington, D.C., March 31, 1973 (Publication SM 33-73).

Williams, Robert, "A Comparison of Hospital Utilization by Costs and by Types of Coverage," *Inquiry*, Vol. 3, September 1966, pp. 28-42.