

DEDUCTIBLES AND THE DEMAND FOR MEDICAL SERVICES: THE THEORY OF THE CONSUMER FACING A VARIABLE PRICE SCHEDULE UNDER UNCERTAINTY

A REPORT PREPARED UNDER GRANTS FROM THE OFFICE
OF ECONOMIC OPPORTUNITY AND THE NATIONAL CENTER
FOR HEALTH SERVICES RESEARCH AND DEVELOPMENT

**EMMETT B. KEELER
JOSEPH P. NEWHOUSE
CHARLES E. PHELPS**

**R-1514-OEO/NC
DECEMBER 1974**

Rand
SANTA MONICA, CA. 90406

The research reported herein was performed in part pursuant to a grant from the Office of Economic Opportunity, Washington, D.C. This project was also supported in part by PHS Grant Nos. 1 R01 HS-00840-01 and 1 R01 HS-01029-01 from the National Center for Health Services Research and Development. Reports of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

DEDUCTIBLES AND THE DEMAND FOR MEDICAL SERVICES: THE THEORY OF THE CONSUMER FACING A VARIABLE PRICE SCHEDULE UNDER UNCERTAINTY

A REPORT PREPARED UNDER GRANTS FROM THE OFFICE
OF ECONOMIC OPPORTUNITY AND THE NATIONAL CENTER
FOR HEALTH SERVICES RESEARCH AND DEVELOPMENT

EMMETT B. KEELER
JOSEPH P. NEWHOUSE
CHARLES E. PHELPS

R-1514-OEO/NC
DECEMBER 1974

Rand
SANTA MONICA, CA. 90406

PREFACE

This report is concerned with the impact of health insurance deductibles on the demand for medical services. Although the study treatment is theoretical, the findings have implications for empirical estimates of the elasticity of demand for medical services, as well as for the design of insurance policies.

The theory of coinsurance follows straightforwardly from economic theory and has been discussed widely in the literature. (Our own work on the theory and effect of coinsurance includes Newhouse and Phelps 1974b, forthcoming; Newhouse, Phelps, and Schwartz 1974; and Phelps and Newhouse 1972, 1973, 1974.) The theory of deductibles, however, introduces some novel aspects, and these are discussed in this report.

The research was supported by grants from the Office of Economic Opportunity and the National Center for Health Services Research and Development to The Rand Corporation.

SUMMARY

In this report we address the theoretical problem of the individual who faces a price schedule that has a kink in it; i.e., after a certain number of units of the good are bought, the unit price of the good changes. After treating the well-understood one-period case, the discussion generalizes to the multiperiod case in which there is uncertainty about future demand for the good. Examples of kinked or variable price schedules with uncertainty present include deductibles and upper limits in health insurance policies, unemployment compensation, and the spending of influence by politicians. For all of these, the consumer must consider possible future demand when making present decisions. In the case of a health insurance deductible, a purchase today has the "bonus" of reducing tomorrow's deductible, although such a reduction is only advantageous if the individual has a future illness.

Here we treat the problem of a variable price schedule with uncertainty as a dynamic programming problem. Because analytical solutions are impossible to obtain at a full level of generality, we have simplified the problem. For concreteness we have structured the problem around a deductible in a health insurance policy, as follows: On any given day the consumer may or may not be sick. If he is not sick, he moves without cost to the subsequent day. If he is sick, he may purchase a cure. If the disutility of paying for the cure is less than the disutility of the suffering caused by the sickness, the consumer will always seek the cure; therefore, the analysis is most interesting if we assume that the disutility of paying for the cure is greater than the disutility of the suffering for any single episode of the illness. However, if the individual pays for enough cures during the "year" (i.e., if he satisfies the deductible), all future cures are free.

We derive some qualitative properties from the model implied by these assumptions. Such properties include: (a) Once the consumer has decided not to seek a cure, he will continue not to do so for the remainder of the accounting period. (b) Abstracting from the change in premium, increasing the deductible or the length of the accounting period increases expected cost, but at a decreasing rate.

We also compute the "effective price" (defined as the nominal or uninsured price less the "bonus" for reducing the deductible) for an example problem. By plotting the effective price against the size of the deductible, one obtains a logistic (S-shaped) curve. (We expect that such a curve would also be observed in a less simple model.) Intuitively, this means that with very small deductibles, the consumer acts as if he had no deductible at all (because he expects, with high probability, to exceed it); and with very large deductibles, the consumer acts as if he had no insurance at all (because he expects, with high probability, not to exceed the deductible). Thus, we expect that actual demand for medical services plotted against the size of the deductible will resemble a logistic curve.

Our model implies that use of average out-of-pocket price or marginal out-of-pocket price to explain variation in annual medical expenditure, as is common in the empirical literature, is a misspecification if a deductible is present. The direction of the resulting bias may be either positive or negative.

This model also has implications for the design of optimal insurance policies. From the point of view of the individual, optimal insurance balances any inefficiencies and increases in administrative cost generated from subsidizing medical care against the value of reduction in risk from more complete insurance. It is suggested that there is likely to be little difference in either demand or administrative cost between "large" and "very large" deductibles (defined as deductibles in the upper range of the logistic curve); therefore, on grounds of risk, large deductibles are preferred to very large deductibles. Put another way, it would appear that a deductible just above the region in which demand is sensitive is a local optimum. For similar reasons, no deductible or a deductible only slightly greater than zero (i.e., one in the lower portion of the logistic curve) is also likely to be a local optimum. The global optimum is likely to be one of these two local optima, but other considerations (such as the amount of the demand change, the variation in administrative cost, and the degree of risk aversion) need to be known before one can choose between these optima. From the point of view of society in choosing a deductible

for a national health insurance plan, additional considerations may enter, such as the distribution of income and whether medical care should be treated as a private good.

ACKNOWLEDGMENTS

We wish to thank James Hosek, Bridger Mitchell, Finis Welch, and Richard Zeckhauser for numerous suggestions.

CONTENTS

PREFACE	iii
SUMMARY	v
ACKNOWLEDGMENTS	ix
Section	
I. DEDUCTIBLES AND HEALTH INSURANCE POLICIES	1
II. THE EFFECTS OF A MULTIPART PRICE SCHEDULE ON ONE-PERIOD DEMAND	5
The Graphical Exposition of the Theory	5
Implications of the One-Period Model	8
III. A MULTIPERIOD MODEL WITH UNCERTAINTY: A MORE GENERAL FORMULATION	10
Expected Costs in the Simple Model	16
Limitations and Extensions of the Simple Model	23
Implications of our Theory for Estimating the Demand for Medical Care	28
Implications of our Model for Insurance Policy Design	33
IV. CONCLUSION	38
Appendix	
A. THE ELASTICITY OF DEMAND FOR PURCHASES NEAR THE KINK	39
B. A SEQUENTIAL DECISION PROBLEM: THE WILDCATTER	42
C. PROOFS OF PROPERTY 3 AND PROPERTIES 6 THROUGH 10	47
BIBLIOGRAPHY	55

I. DEDUCTIBLES AND HEALTH INSURANCE POLICIES

A principal purpose of health insurance is to spread the risk of major expenditures for medical care. Without such insurance, patients might become destitute as a result of extremely expensive illness or even unable to afford treatment to keep them alive. The value of such risk spreading has made health insurance highly desirable; today about 80 percent of American civilians are protected by some form of private health insurance. There are also two major federal health insurance programs in the United States--Medicare for persons over 65 and Medicaid for some of the low-income population¹--plus state and local government assistance programs. There is substantial support for new legislation covering the entire population. However, both Medicare and Medicaid greatly exceeded their anticipated costs, partly because the demand response to the insurance changes that these programs introduced was underestimated. As a result, those contemplating new legislation are quite interested in estimates of demand.

In recent years a number of articles have appeared in response to this interest; these articles attempt to estimate the demand for medical care as a function of insurance (Feldstein 1970, 1971; Davis and Russell 1972; Rosett and Huang 1973; Newhouse and Phelps 1974b; Phelps and Newhouse 1974). Almost invariably these articles make the assumption that medical insurance is of a form that simply reimburses the consumer a stipulated fraction of his expenditure, say 80 percent. The remaining fraction, the coinsurance rate, is paid by the patient.

This is a convenient assumption, because it permits one to apply standard economic theory in a straightforward fashion; instead of facing a gross price p_h for medical care, the consumer is assumed to face a price Cp_h , where C is the coinsurance rate. Unfortunately, this assumption is known to be grossly at variance with reality, because health insurance frequently includes a deductible and an upper limit, in addition to a coinsurance rate. In the subsequent sections we will

¹The medical deduction on the federal income tax also operates as a public insurance program.

present a model of patient behavior that includes such features. In the remainder of this section, we will define the terms deductible and upper limit, show that such features are frequently present in existing insurance, and discuss why this is so.

A deductible specifies the amount that a consumer must pay before the insurance becomes effective; an upper limit specifies the amount of total expenditure, above which there is no insurance. If there is a deductible, the policy also defines an accounting period, such as a year, at the end of which the consumer must again satisfy the deductible.²

Deductibles are quite common. In 1970, about 40 percent of the private insurance policies had a deductible (Health Insurance Association of America 1972), as did Medicare. The only general National Health Insurance (NHI) program currently in force--the medical deduction on the income tax--has a 3-percent-of-income deductible for medical expenses and a 1-percent-of-income deductible for drugs. Many NHI bills propose \$100 or \$150 annual per person deductibles for nonpoor families.

Upper limits are even more common in private insurance policies than deductibles, but they are usually too high to be of much practical significance. Contrary to normal principles of insurance, they act to cut off payment when it is most needed. Virtually all proposed NHI bills have just the opposite approach to large medical costs. They specify, instead, a maximum amount that any family must pay; above that amount, the plan pays for everything.

The popularity of deductibles in insurance policies is not surprising, because they have a number of desirable features. First, administrative costs appear to be nearly invariant to the size of the claim.³ As a result, insurance that covers larger claims is more attractive than insurance that covers smaller claims, because the payout ratio is higher.

²In some plans the consumer may be permitted to carry forward unreimbursed expenditures that occur near the end of the accounting period to the next period.

³Parish (1974) asserts that the cost of processing a claim is about \$3 per claim, regardless of the size of the claim.

Second, the problem of adverse incentives may be more severe for small claims than for large. Specifically, if price elasticity falls with total expenditure (there is little evidence), insuring large losses is relatively more attractive.⁴

Finally, deductibles are beneficial to risk-averse policyholders because they concentrate the benefits where they are needed most, i.e., where the losses are greatest. This concentration of benefits is the heart of Arrow's (1963) proof that the optimal insurance policy for a given premium has no coinsurance after a deductible. However, the assumptions used to establish this result are known not to hold. Below we derive the implications of more realistic assumptions.⁵ We show the conditions under which Arrow's result will be correct; given the popularity of deductibles, these conditions may frequently be satisfied.

One of our primary purposes in analyzing the behavior of a consumer whose policy includes a deductible clause is to determine what

⁴Insurance coverage may also make it rational for consumers to take more chances and spend more on medical services than they would choose to, a priori. This is another type of adverse-incentives problem, but it is not clear whether coverage of small or large expenditures is more likely to induce such behavior. [See Ehrlich and Becker (1972).]

⁵Arrow assumed that (1) the loading fee (percentage taken by the insurance company) is constant over all states of the world; (2) losses are monetary and freely observable; (3) there is no change in the distribution of losses because of insurance (no problem of altered incentives). The first assumption is unlikely to hold because the loading fee as a percentage probably falls with total expenditure; this, however, serves to strengthen Arrow's result. The second assumption does not hold because health losses are not monetary; consumers do not want to balance income itself, but rather the marginal utility of income in different states of health. [See Zeckhauser (1970), Phelps (1973), and Arrow (1973a).] This implies varying insurance by type of illness, but very few policies do so because of transactions costs. It has proved almost impossible to define categories and to ensure that patients have been correctly placed in them. Indeed, diagnosis is often an important (and costly) medical service. However, weakening the second assumption does not appear to weaken the optimality of a deductible. The third assumption is incorrect because demand elasticity is greater than zero (see below). The implications of relaxing this assumption for the design of optimal insurance are discussed at the end of this report.

the possibilities are for designing better insurance policies. (There are, of course, considerations other than those of an individual consumer that must be taken into account when discussing national health insurance; we will discuss these later.) Before we can even begin to talk about the individual consumer's choice among policies, we must understand the demand response of consumer and physician to various types of policies. The reaction of consumers and physicians to deductibles is difficult to analyze because of (a) the nonconvexity that deductibles introduce into the policyholder's budget line, and (b) the sequential decision problem that confronts him. When deciding whether to make a medical purchase, the patient and his physician must consider not only how much the current purchase will cost, but how it will affect the costs of possible future purchases. The main part of this report is devoted to an analysis of a model that incorporates these two factors. This analysis is followed by a discussion of the implications of the model for estimating demand, and for insurance policy design.

II. THE EFFECTS OF A MULTIPART PRICE SCHEDULE ON ONE-PERIOD DEMAND

Consumers are frequently faced with a multipart price schedule, in which the price per unit changes, depending on the quantity of a good or service demanded. Volume discounts for goods are a simple example. Multipart price schedules can be maintained only if entrepreneurial resale of the good is difficult or unprofitable. This is the case in natural monopolies such as the electric and telephone companies, who use such schedules to price services with falling marginal costs. For these services, the traditional economic approach of marginal cost pricing would cause the business to show a loss.⁶ Service monopolies such as amusement parks and country clubs can also use multipart price schedules to extract more money from consumers by charging admission or membership fees and then making additional charges for rides or for using the greens.⁷ Deductibles in health insurance policies are another example of a multipart price schedule.

The economics of demand response to price schedules are the same, no matter what their cause. Since our focus in this report is on health insurance, we will discuss the economics of demand response in the context of a deductible in a health insurance policy, although our analysis could be applied to any commodity.

A GRAPHICAL EXPOSITION OF THE THEORY

Consider a consumer who has not yet exceeded his deductible, and is trying to decide how much medical care to purchase on the last day of an accounting period. For any illness, ℓ , he may suffer, we assume that he or his physician (acting as an agent) can exercise his preferences between a composite other good, x , and medical services, h .

⁶For a defense of multipart price schedules and a history of the literature development in the field, see Coase (1970).

⁷A thorough analysis of profit-maximizing two-part tariffs in the monopoly case is presented in Oi (1971).

Phelps (1973) shows formally how this might be done, starting from a utility function $U(x, H)$ defined over other goods and perceived health, H , which is related to h by the equation $H = H_0 - \ell + g(h, \ell)$, where ℓ is the random loss of health from illness and $g(h, \ell)$ is the production function of the stock of health.⁸ Let the price of x be 1 and the price of h be p_h . The consumer's budget line and indifference curves for the two different illnesses, $\ell_1 > \ell_0$, might be as shown in Fig. 1.

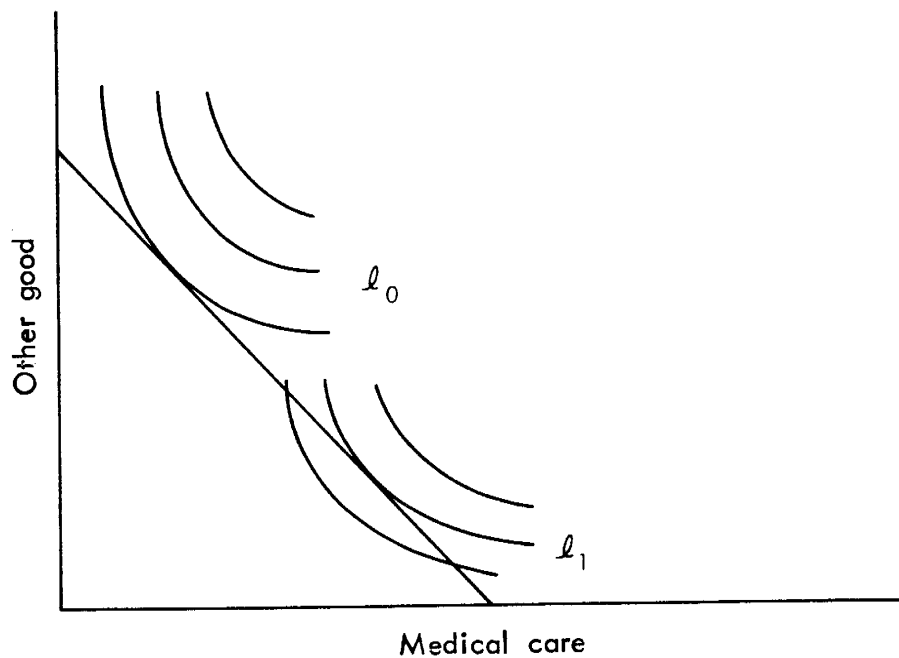


Fig. 1—Derived indifference map with no insurance

Assume now that the consumer owns an insurance contract that specifies some deductible, d (in units of h), and a coinsurance rate, C . It can be easily shown that larger losses shift the consumer's demand

⁸See also Grossman (1972) and Phelps and Newhouse (1973). Note that the consumer generally does not know his real health but can only guess at it. For the purpose of this model, H represents what he thinks is the state of his health. Thus a doctor who convinces the patient that there is nothing really wrong increases H .

for medical care so that (*ceteris paribus*) more medical care is consumed as ℓ increases.⁹ In Fig. 2, the altered budget line is shown, with one solution value A for h in the range where full market prices are paid. That is, the consumer's deductible has not yet been satisfied, given his loss, ℓ_0 . Also shown in Fig. 2 is another loss, ℓ_1 , which alters the difference curves and leads the consumer to purchase sufficient medical care to satisfy the deductible as shown by solution value B.

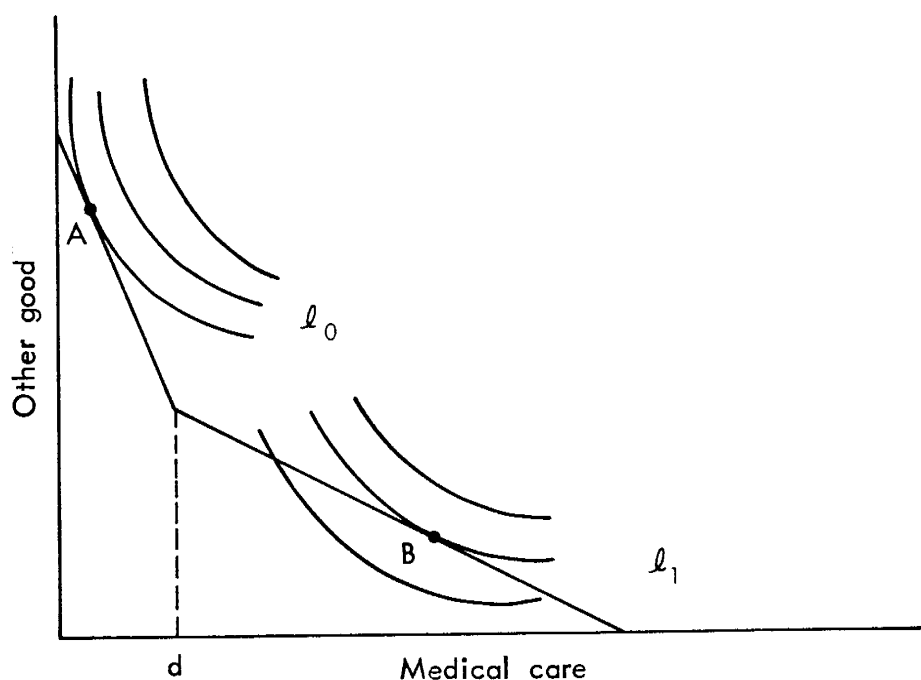


Fig. 2—An insurance policy with a deductible equal to d units of care

Given the kinked budget line shown in Fig. 2, for any loss ℓ , there exists by continuity some pair (C, d) that will place the consumer at a point of indifference between two different pairs of x and h . It is similarly true that for any pair (C, d) there is some loss ℓ_d

⁹See Appendix A of Phelps and Newhouse (1973) for a proof of this proposition. We assume enough differentiability of $U(x, H)$ and $g(h)$ so that $u(x, h, \ell)$ is smooth.

such that any greater loss will "push the consumer over," so that he consumes above d rather than below it, in which case Cp_h is the marginal money price rather than p_h .

IMPLICATIONS OF THE ONE-PERIOD MODEL

We have the interesting result in this one-decision case that consumers with smooth indifference curves would *never* be observed purchasing exactly d units of medical care, and in general would not be near the kink in the budget line.¹⁰ In fact, a consumer who knowingly purchases care near the kink must have an elasticity of close to zero at that point. In Appendix A, we derive the result that the price elasticity of demand satisfies

$$\eta < \frac{\Delta h \{1 + C + [2(1 + C^2)]^{\frac{1}{2}}\}}{2h(1 - C)}, \quad (2.1)$$

where Δh is distance from the actual purchase to the kink, h is the amount purchased, and C is the coinsurance rate after the deductible is satisfied. If, for example, a consumer with \$100 remaining in his deductible and whose $C = .25$ purchases \$98 worth of medical care on the last day of an accounting period, then $\Delta h = 2$, $h = 98$, and $\eta < .037$. We shall subsequently show, however, that this formula may only be applied in the one-period case and thus is of more limited scope than it appears to be at first glance.

Another point to be noted is that it is not enough to find a point where the indifference curve is tangent to the budget line--this only insures a *local* optimum. With a normal utility function, finding the overall optimum is made easy by knowledge of the critical level of illness, ℓ_d . All losses smaller than ℓ_d will produce purchases of h

¹⁰Occasionally, one sees the bizarre, "notched" pricing schedule in which average cost falls in discrete jumps. For example, a customer may get 1 to 5 xerox copies for 3 cents each or 6 to 10 for 2 cents each. Unless disposal costs are significant, only the most dedicated ecoperson would purchase exactly 5 copies. What we show here is that "kinks," as well as "notches," may cause gaps in observed quantities demanded.

less than d , and losses greater than ℓ_d will lead to h greater than d . It is easy to see that this discussion can be generalized to multipart price schedules with more than one kink. In such schedules there will be a critical illness level for each kink.

The analysis of multipart price schedules other than health insurance (such as those discussed earlier) is, in general, simpler because future demand is more certain, and therefore the approximation of only one set of indifference curves should give reasonable results. Thus, for most goods with multipart price schedules, we would not expect to see consumers near a kink; rather, we would expect them to be indifferent, say, between (a) low consumption of electricity, and some consumption of gas or coal, and (b) high consumption of electricity (via electrically heated homes), and low consumption of other fuels. A shift in the point at which rates decreased would change the equilibrium position of some consumers, causing (in equilibrium) shifts to or from electrically heated homes. In other words, falling price schedules would lead to "specialization" in certain factors of production or consumption of certain goods.

III. A MULTIPERIOD MODEL WITH UNCERTAINTY:
A MORE GENERAL FORMULATION

With existing health insurance (such as Medicare), expenses toward a deductible accumulate over time, so that even if expenditure on today's illness does not satisfy the deductible, expenditure on tomorrow's illness may satisfy it (assuming that today is not the last day of the accounting period). If future illnesses could be predicted with certainty, the decision problem for any one illness (loss) would be the same as in the one-period case. If total planned consumption exceeds a critical level with certainty, then the deductible will be used up, and the effective marginal cost for all purchases of medical services will be Cp_h rather than p_h . The "excess" charge $d(1 - C)p_h$ for the units below the deductible can be considered as part of the premium--it produces only an income effect on demand. Similarly, if total planned consumption is less than the critical level with certainty, then the marginal cost is p_h .¹¹

When uncertainty is present, any expenditure in the range below the deductible has the bonus of reducing the remaining deductible and hence reducing the expected costs of future medical care: the greater the chance that future expenditures will exceed the deductible, the cheaper today's visit to the doctor. As a result, uncertainty changes the function that relates marginal price to expenditure from the step-function of the certainty case to a smooth function.

In this section, we will examine this effect in detail in a formal model of the sequential decision process faced by an ill consumer. The main point is that uncertainty about the future implies that decisions cannot be viewed in isolation but must be considered as part of a sequence. Although our model is formulated in terms of health insurance, the results are equally applicable to a worker who has a certain number of sick-leave days and is contemplating taking one, or a worker who is

¹¹The idea of insurance when the future is known with certainty is rather strange. Either the consumer won't want to buy, or the company won't want to sell. What one gains, the other loses.

approaching the end of his unemployment benefits and is confronting a job offer but is uncertain about whether he can obtain a better offer, or a politician uncertain of reelection who is making a decision to "spend" influence.

The Complete Formal Model

Before stating the formal model, we will clear up one ambiguity. In the one-decision case, the composite "other good," x , and health, H , could be considered to be either stocks or flows. In a multistage decision, it makes no sense to maximize utility in each period with stocks as arguments. Instead, we will assume $U(x, H)$ is the valuation on final stocks of other goods and health at the end of the deductible period, and introduce another function $u(e, H)$ to represent the current utility given by expenditure on other goods, e_t , and health status, H_t , at time, t , of the deductible period. (We continue with the notation of H on the assumption that the flow of health status is proportional to the stock. This is a standard assumption; see Grossman 1972.) Thus, if the deductible period ends at step T , the consumer tries to maximize the expected value of

$$\left[\sum_{t=1}^{T-1} u_t(e_t, H_t) \right] + U(x, H) \quad (3.1)$$

subject to

$$\$1 \cdot x + \sum_{t=1}^{T-1} (e_t + p_h h_t) = I, \quad (3.2)$$

where I is the total income over the period, including any initial holdings.

This problem can be formulated as a dynamic program.¹² For example, suppose that the consumer is in the second to last step of the

¹² A somewhat different way of presenting this problem is to use the model of decision theory. [See Raiffa (1968).] An example of how this is done is given in Appendix B.

deductible period. At the start of period $T - 1$, he has health status H_{T-1} , remaining money $\$1 \cdot x_{T-1}$, and unused deductible d_{T-1} . We suppose that the distribution of illnesses, $f(\ell, H_T)$, in the last period depends only on H_T . Using the method of the preceding section, the consumer knows that his optimal response to each of the various possible losses ℓ_T in the last period is $h^0(H_T, x_T, \ell_T, d_T)$. Let $W(H, x, d, t)$ be the expected sum of utility from t to T if the consumer behaves optimally. Then

$$\begin{aligned} W(H_{T-1}, x_{T-1}, d_{T-1}, T - 1) = & \max_{e, h} u_{T-1}(e, H_{T-1} + g(h)) \\ & + \int [U(x_{T-1} - e - p_h h - p_h h^0, H_{T-1} + g(h) \\ & - \ell + g(h^0)) f(\ell, H_{T-1} + g(h))] d\ell, \quad (3.3) \end{aligned}$$

where the first term is utility in period $T - 1$ if the consumer buys h units of care, and spends e on other goods, and the integral gives the expected value of his legacy $U(x, H)$ if he behaves optimally in the last period. When there are many stages left in the deductible period, the consumer simply calculates W backwards from the end of the period to his present position.

Note that a consumer purchasing any durable good whose usefulness varies with the state of the world must solve a similar problem even if there is a fixed price per unit of the good. In other words, the expected value of optimal use in various future contingencies must be weighed against expected values of other uses for the money (Zeckhauser 1973b).

The Simplified Model

Derivation of the properties of a model at this level of generality is well out of reach. We have therefore assumed a simple form of the utility function and a simple model of illness in order to derive qualitative properties. Except where noted, we conjecture that the solution to the more general problem is qualitatively similar. Before

getting into the details, we shall try to justify the two main simplifications that are introduced, namely that the consumer confronts a minor ailment, and has a linear utility function. The situation of most interest for a medical plan with at least a moderate deductible (say the \$150 deductible that is part of many proposed national health insurance bills) is that of a person who has not yet exceeded the deductible, who has some troubling but not overwhelming symptoms, and is trying to decide whether to go to a doctor or not. For the person with a sufficiently severe pain or threat rather than a minor problem, a moderate deductible is not likely to affect the decision to seek care. (He will have a large health stock loss.) Thus, the assumption of a relatively minor ailment is the interesting case analytically; it also appears to be the most common case practically, because "large" health stock losses are relatively infrequent. (Only about 10 percent of the population are hospitalized annually.)

The second assumption--that of a linear utility function or risk neutrality--may appear to be stringent, but it is not. Because utility is a function of both health and wealth, we consider risk aversion with respect to both arguments. Risk aversion to loss of wealth is the traditional concept of risk aversion and follows from the decreasing marginal utility of money. However, we would argue that possession of a contract that fully insures expenses above a deductible reduces the individual's concern about possible loss of wealth. For a small or moderate deductible, therefore, we may approximate his insured behavior by a risk-neutral utility function.¹³ We are not analyzing why the individual bought insurance, but how he acts after he has purchased it.

¹³ Direct evidence on risk aversion is sketchy, but using health insurance enrollment data, Friedman (1974) has estimated that an individual with \$10,000 annual income would pay \$143 to avoid a 10 percent chance of a \$1000 loss. For the same utility function, the same person would pay \$33 to avoid a 10 percent chance of a \$300 loss, or merely \$3 above the actuarial value of the lottery. Thus for a plan with small to moderate deductibles, our assumption of risk neutrality does not seem grossly at odds with Friedman's estimates.

Friedman's estimates of risk aversion would appear to be an upper bound, for several reasons. He attributes to risk aversion several

Risk aversion to loss of health is less familiar than risk aversion to loss of wealth. Health-stock losses, unlike wealth losses, have no natural scale; hence, the health-stock loss for any one illness episode can be scaled so that utility is proportional to the size of the loss. The individual can thus be assumed to be risk neutral in health for any one episode. There remains the issue of risk neutrality toward number of episodes (e.g., indifference between one episode with certainty and an equal chance of two episodes or none). The question is, do individuals adjust to repeated illnesses, or do they suffer more than proportionately from them? Because there is no obvious answer, we believe that the most reasonable assumption is that the individual suffers proportionately to the number of episodes.

We therefore feel justified in making the mathematically convenient assumption that the utility function is linear (i.e., risk neutrality). If the approximation of risk neutrality toward wealth is a poor one, and the individual is significantly risk averse in money over this range of losses, then the consumer will spend less on care than our model predicts; if he is risk averse to the number of illness episodes, then he will tend to spend more on care than our model predicts.¹⁴

other factors that would lead persons to select better insurance plans; self-selection of sickly persons toward better insurance and the income-tax deductibility of premiums are two such factors. There may be offsetting biases, but we do not have sufficient data to assess the net effects. Nevertheless, we present Friedman's estimates as the only ones available from studies of actual consumer behavior.

¹⁴The decision problem is that of an individual with a deductible in his policy who must make a decision about whether or not to consume care, knowing only the probability of contracting illness in the future. For a sufficiently large health-stock loss, the individual will consume care in the present, and the probability of contracting illness in the future is irrelevant. (This converts the problem to the one-period case.) If the health-stock loss is not that great, aversion to the probability of wealth loss will make the individual less inclined to consume care; by not consuming care, he can avoid any uncertainty about his final wealth outcome. (By the definition of risk aversion, certainty equivalents are preferred.) If the individual is averse to uncertainty about the number of illnesses he would suffer (i.e., his utility falls disproportionately with the number of episodes), he is more likely to consume care, because care counteracts the effect of the illness; thus, by incurring a known money expense, he offsets the possible loss of utility from future illness.

To begin, suppose that in each day of the deductible period the consumer has probability, p , of suffering a loss, ℓ^* . A complete cure costs F , and is instantaneous. If no cure is sought, the patient recovers at the end of the day anyway. (The cure saves him one day of suffering.) The insurance policy has a deductible of DF and zero co-insurance after the deductible.¹⁵ The consumer tries to minimize the expected sum of costs in money and pain. If $\ell^* \geq F$, his decision is easy--he will always see the doctor, whether he has insurance or not. To concentrate on the case in which insurance makes a difference, we will assume for now that $\ell^* < F$. Thus, for the moment, the consumer whose behavior we are studying has only minor illnesses.

The problem can be structured as a simple dynamic program. Let $V(n,d)$ be the expected costs, in money and pain, to a consumer with n days and dF deductible left; V is the negative of the utility function. By Bellman's principle of optimality, the consumer can make his decisions one step at a time. With probability $(1 - p)$ he will not be sick, and go costlessly into day $n - 1$. With probability p he will be sick, and have to decide between suffering or purchasing relief that at the same time reduces his deductible. Thus for all n and d such that

$$1 \leq n \leq N, 1 \leq d \leq D,$$

we have

$$V(n,d) = (1 - p)V(n - 1,d) + p \min \{V(n - 1,d) + \ell^*, V(n - 1,d - 1) + F\}.$$

Once he has used up his deductible, he will naturally always choose to relieve sickness, so $V(n,0) = 0$. If there are no more days left in the accounting period, there are no more costs, so $V(0,d) = 0$. Using these initial conditions, one may compute $V(n,d)$ recursively.

¹⁵Formally, $f(\ell) = \ell^*$ with probability p , 0 with probability $(1 - p)$, independently of H , t . $u_t(e,H) = e + H - H_0$ for all t , and $U(x,H) = x$. $g(1) = \ell^*$ provided $f(\ell) = \ell^*$. $p_h = F$.

EXPECTED COSTS IN THE SIMPLE MODEL

We first state a number of properties of this model.

Property 1: $V(n,d) \leq V(n,d + 1)$.

Property 2: $V(n,d) < V(n + 1,d)$ for $d > 0$.

These two facts are clear from the definition of the model. The first says that with n fixed, a higher deductible cannot decrease expected costs; the second says that with d fixed, a longer time horizon in which to be potentially sick increases expected costs.

Let $A(n,d)$ be the optimal action at (n,d) . $A(n,d)$ can have two values, s (suffer) or r (purchase relief).

Property 3: In the matrix $A(n,d)$, the orderings $\begin{pmatrix} r \\ s \end{pmatrix}$ and $(s \ r)$ are impossible. ($\begin{pmatrix} r \\ s \end{pmatrix}$ means the individual suffers on day n and obtains relief on day $n - 1$, holding constant the remaining deductible; $(s \ r)$ means the individual seeks relief with deductible d and suffers with deductible $d - 1$, holding constant the remaining days.) This means that $A(n,d)$ looks like Fig. 3. The proof of property 3 (as well as the proof of properties 6, 7, and 8, below) is in Appendix C. It follows from property 3 that once the consumer decides to suffer (his deductible is too large relative to the time remaining in the accounting period to make the purchase of relief attractive), then he will suffer for any other illness he happens to have until the end of the period.

Property 4: $V(n,d) < dF$. Also, $\lim_{n \rightarrow \infty} V(n,d) = dF$.

The point of property 4 is that the strategy of "always get the cure" results in a cost of dF if the consumer gets d or more illnesses, and costs less otherwise. Because relief is appropriate whenever the expected loss from suffering exceeds dF , as n becomes large, $V(n,d)$ approaches dF .

Property 5: If $n \leq d$, then $V(n,d) = np\ell^*$.

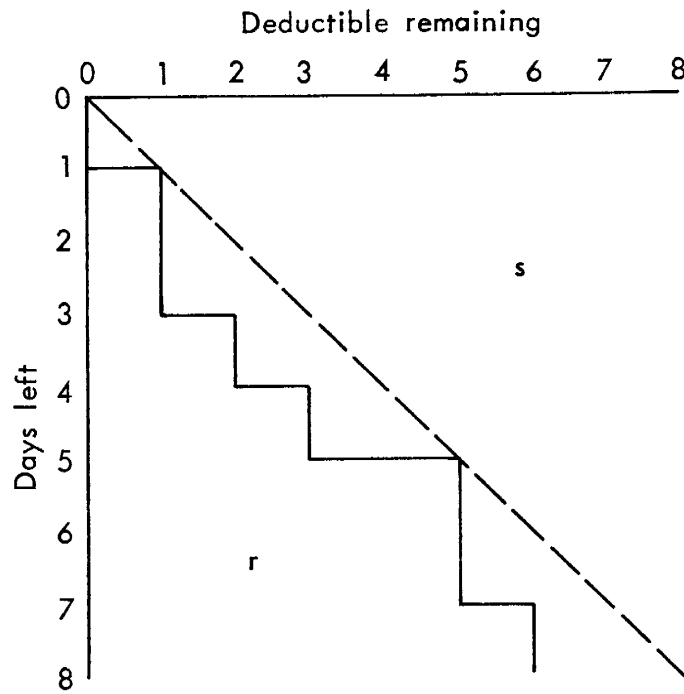


Fig. 3—Indifference locus between purchasing relief (r) or suffering (s), given the size of deductible and the remaining accounting period. The locus lies on or below the 45° dotted line, by Property 5.

This fact follows from the assumption that $\ell^* < F$. Because it is impossible to use up the deductible, it is foolish to try. Thus, the indifference locus drawn in Fig. 3 lies on or below a 45° line extending southeast from the origin.

Property 6: $V(n,d) - V(n,d - 1) \leq V(n,d - 1) - V(n, d - 2),$
for $d \geq 2$.

This fact shows the decreasing marginal cost of care to the consumer with a higher deductible. As property 5 shows, once d gets big enough, the consumer should decide to suffer, so that increasing d past that point has no effect on his decisions.

Property 7: $V(n,d) - V(n-1,d) \geq V(n,d-1) - V(n-1,d-1),$
for $d, n \geq 1.$

The marginal cost of extra days increases with the amount of the deductible left.

Property 8: $V(n,d) - V(n-1,d) \leq V(n-1,d) - V(n-2,d),$
for $n \geq 2.$

This fact shows the decreasing marginal cost to the consumer of extra days. As property 4 shows, the marginal cost of extra days eventually approaches 0. We can summarize most of the above facts in terms of derivatives if we act as if n and d are continuous: $V_d, V_n \geq 0$; $V_{dd}, V_{nn} \leq 0$; $V_{nd} \geq 0.$

Marginal Cost and Demand in the Simple Model

Table 1 shows $V(n,d)$ (labeled "Expected Costs"), as well as the expected use of cures and the true price of cures to the consumer for $0 \leq n \leq 15$ and $0 \leq d \leq 5$ for parameter values $\ell^* = \$5$, $F = \$7.50$, and $p(\ell^*) = 0.4$. The Expected Costs portion of the table shows what the consumer would pay in a situation (n,d) for a cure that would give him immunity against illness until the end of the period.

The optimal action can be read from the Expected Number of Visits for Care portion of Table 1. If the entry is zero, the optimal action is to suffer, but whenever the entry is nonzero, the optimal action, given an illness, is to get relief. For example, suppose $d = 3$, $n = 12$; then the expected number of visits is 3.88. This is greater than zero, so a consumer in that position will see a doctor if he becomes ill. In that case, he will be at $d = 2$, $n = 11$ in the next day. If he stays well for the next 6 days, he arrives at $d = 2$, $n = 5$. Expected use is zero, so that he does not treat an illness on that day, or any other day until the end of the period. Note that with no deductible left, the expected use is precisely the expected number of illnesses. For example, since $p = 0.4$, with 10 days left, the expected use is 4. With a large deductible left, subsidized care is so inaccessible that for

Table 1
PROPERTIES OF THE MODEL, WITH GIVEN PARAMETERS^a

Days Left	Deductible Remaining (in number of visits)					
	0	1	2	3	4	5
Expected Costs (\$) = V(n,d)						
0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	2.00	2.00	2.00	2.00	2.00
2	0.0	4.00	4.00	4.00	4.00	4.00
3	0.0	5.40	6.00	6.00	6.00	6.00
4	0.0	6.24	8.00	8.00	8.00	8.00
5	0.0	6.74	10.00	10.00	10.00	10.00
6	0.0	7.05	11.70	12.00	12.00	12.00
7	0.0	7.23	12.84	14.00	14.00	14.00
8	0.0	7.34	13.59	16.00	16.00	16.00
9	0.0	7.40	14.09	18.00	18.00	18.00
10	0.0	7.44	14.42	19.44	20.00	20.00
11	0.0	7.46	14.63	20.43	22.00	22.00
12	0.0	7.48	14.76	21.11	24.00	24.00
13	0.0	7.49	14.85	21.57	25.84	26.00
14	0.0	7.49	14.90	21.88	27.13	28.00
15	0.0	7.50	14.94	22.09	28.03	30.00
Expected Number of Visits for Care						
0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.40	0.0	0.0	0.0	0.0	0.0
2	0.80	0.0	0.0	0.0	0.0	0.0
3	1.20	0.72	0.0	0.0	0.0	0.0
4	1.60	1.31	0.0	0.0	0.0	0.0
5	2.00	1.83	0.0	0.0	0.0	0.0
6	2.40	2.30	1.13	0.0	0.0	0.0
7	2.80	2.74	2.00	0.0	0.0	0.0
8	3.20	3.16	2.69	0.0	0.0	0.0
9	3.60	3.58	3.28	0.0	0.0	0.0
10	4.00	3.99	3.80	1.71	0.0	0.0
11	4.40	4.39	4.27	2.95	0.0	0.0
12	4.80	4.80	4.72	3.88	0.0	0.0
13	5.20	5.20	5.15	4.62	1.95	0.0
14	5.60	5.60	5.57	5.23	3.42	0.0
15	6.00	6.00	5.98	5.77	4.54	0.0
Effective Price of Care (\$)						
0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	7.50	7.50	7.50	7.50	7.50
2	0.0	5.50	7.50	7.50	7.50	7.50
3	0.0	3.50	7.50	7.50	7.50	7.50
4	0.0	2.10	6.90	7.50	7.50	7.50
5	0.0	1.26	5.74	7.50	7.50	7.50
6	0.0	0.76	4.24	7.50	7.50	7.50
7	0.0	0.45	2.85	7.20	7.50	7.50
8	0.0	0.27	1.89	6.34	7.50	7.50
9	0.0	0.16	1.24	5.09	7.50	7.50
10	0.0	0.10	0.81	3.59	7.50	7.50
11	0.0	0.06	0.53	2.48	6.94	7.50
12	0.0	0.04	0.34	1.70	5.93	7.50
13	0.0	0.02	0.22	1.15	4.61	7.50
14	0.0	0.01	0.14	0.78	3.23	7.34
15	0.0	0.01	0.09	0.52	2.25	6.63

^aTo calculate these numbers, we assume that
pain (ℓ^*) = \$5, fee (F) = \$7.50, probability
of pain ($p(\ell^*)$) = 0.4.

decision purposes it is as if it didn't exist. Just as he would without insurance, the consumer always decides to suffer. In between the extremes, the usage looks like a fairly steep logistic curve. This is shown in Fig. 4, which plots expected use against a varying deductible for accounting periods of 6, 10, and 13 days.

The Effective Price of Care portion of Table 1 is the nominal out-of-pocket price (F) minus the value of reducing the deductible. This latter value, which we will call the *bonus*, is the amount that the consumer would pay (ex ante) to exchange his insurance policy for one with the deductible reduced by one visit. In the model, the bonus is equal to $V(n-1, d) - V(n-1, d-1)$. The properties of effective price are shown in Fig. 5. The falling lines of Fig. 5 are

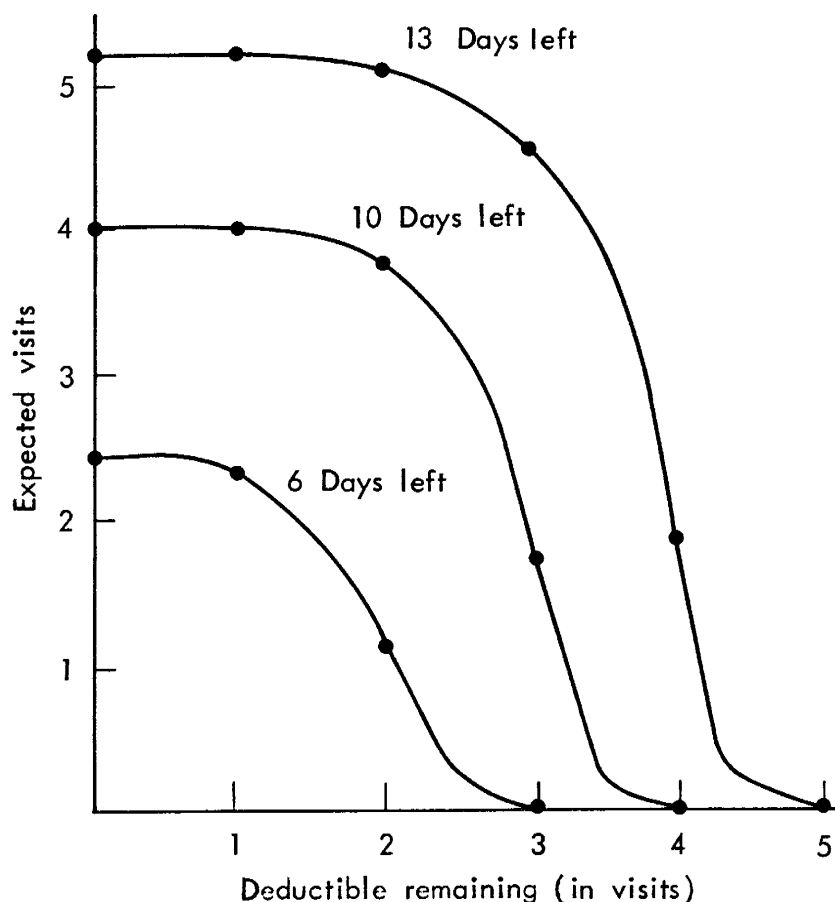


Fig. 4—Expected visits for medical care given different accounting periods

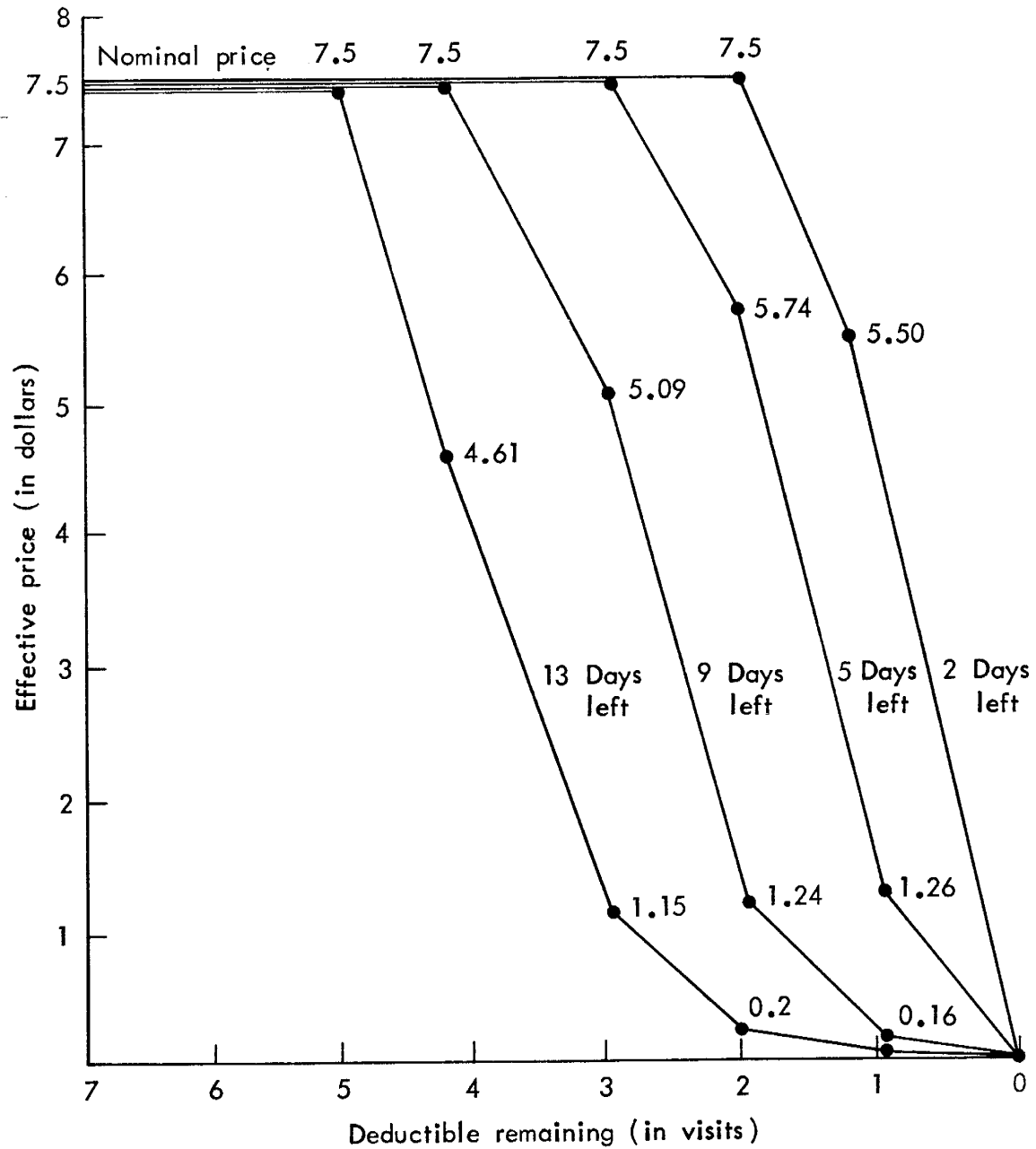


Fig. 5—Effective price depends on deductible and number of days left .
The values are taken from Table 1.

based on the rows of the Effective Price of Care portion of Table 1 corresponding to 2, 5, 9, and 13 days left.¹⁶ The effective price never exceeds the nominal price of \$7.50. As the unused deductible falls because of consumption during an accounting period, the effective price also falls, although the rate at which it falls depends on the number of days left in the accounting period. The more time left in the accounting period, the more the effective price falls (property 7).

In the simple model we are discussing, the demand function takes the form of a step function. Whenever there is an illness, demand is one unit at any effective price under \$5 and no units at any effective price above \$5. Thus, in Table 1, if the effective price exceeds \$5, there is no use. If we relax the assumption that all illnesses are of equal severity and suppose that the consumer has a distribution of possible illnesses of different severity on a given day, the effective price would act as the cutoff point between the illnesses he would relieve at the nominal price of \$7.50 or those he would not. For example, from Table 1 we may conclude that if he had 13 days remaining and just one unit of deductible left, he would spend the \$7.50 to buy relief from a \$.02 illness but not from a \$.01 illness.

Provided that the pattern of illness and cures reverts to the simple form of one loss and one possible treatment on all future days, we could use the simple model to analyze a quantity-demanded decision on any given day. Suppose that the consumer has a very severe illness, and could take one or two units of treatment with diminishing marginal effect but with constant prices. Then the bonus would be an easily computed increasing function of the size of the expenditure.¹⁷

¹⁶The continuous fall of effective prices as medical services are used limits the scope of our earlier result on expenditures near the kink. Only when we are close to the end of the deductible period will the budget line have a sharp bend, and expenditures near the kink be unlikely.

¹⁷Precisely, the bonus from the purchase of two units would be $V(n - 1, d) - V(n - 1, d - 2)$. If the consumer could purchase any amount of care between 0 and 2 units, the exact bonus would be difficult to compute, but we could approximate it by interpolating from our integer values.

Effects of Higher Variance of Sickness Costs

What happens to demand for care in our model if the variance of sickness costs rises? Property 9 in Appendix C shows that in this model (even without risk aversion!) a higher variance of sickness costs, given a constant mean (i.e., the probability of illness times ℓ^* or F is constant), makes using up the deductible more attractive. Therefore, demand for care will generally be higher. Property 9 is illustrated in Table 2, which is a variant of Table 1 with ℓ^* and F four times as big and p one-fourth as big. Thus, one visit that costs \$30 in the High Variance column is equivalent to four visits that cost \$7.50 in the Low Variance column. Although the mean cost of illnesses NpF is the same as in Table 1, the variance $Np(1 - p)F$ is one and one-half times as big. Table 2 illustrates that relief is sought sooner with the larger variance of expenses, and expected "costs" (disutility) are lower.¹⁸ Intuitively, the reason for this is that an episode of illness, if it occurs, moves the consumer closer to satisfying the deductible and therefore has more information content. Moreover, this is not only true of any illness that does occur, but is also true for any illness that might occur in the future. For both of these reasons the bonus associated with a single expenditure is larger.

LIMITATIONS AND EXTENSIONS OF THE SIMPLE MODEL

Our simple model indicates that individuals with a "moderate" deductible insurance plan should be sharply divided in their doctor-visiting habits. Those who can expect a lot of sickness should use up their deductible, but those who feel themselves to be basically healthy will rarely go to the doctor for minor problems. It seems unlikely

¹⁸Another way of generating equivalent higher variance programs is to make the period k times as long and the probability of an illness in any day = p/k . Property 10 in Appendix C shows that in this case, also, relief is sought earlier, and "costs" (disutility) are lower.

If we divide the days into smaller and smaller time parts--letting N increase and p decrease proportionally so that Np is constant, in the end we have a Markov renewal program to decide whether to treat the illnesses which arrive as a Poisson distribution with mean Np . The optimal policy for the renewal program is the limit of the optimal policies for the dynamic programs. [See Fox (forthcoming).] It is this property that makes discretization and computer solution useful in continuous programs.

Table 2
HIGH AND LOW VARIANCE OF ILLNESSES

Days Left	High Variance ^a		Low Variance ^b	
	Deductible Remaining (in number of visits)		Deductible Remaining (in number of visits)	
	1	2	4	8
Expected Costs (\$) = V(n,d)				
0	0.0	0.0	0.0	0.0
3	6.00	6.00	6.00	6.00
6	12.00	12.00	12.00	12.00
9	16.88	18.00	18.00	18.00
12	20.43	24.00	24.00	24.00
15	23.03	30.00	28.03	30.00
18	24.92	36.00	29.38	36.00
21	26.29	41.27	29.81	42.00
24	27.30	45.44	29.94	48.00
27	28.03	48.73	29.98	54.00
30	28.56	51.31	30.00	57.90
Expected Number of Visits for Care				
0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0
9	0.5	0.0	0.0	0.0
12	0.9	0.0	0.0	0.0
15	1.3	0.0	4.54	0.0
18	1.6	0.0	6.78	0.0
21	2.0	0.8	8.28	0.0
24	2.3	1.4	9.57	0.0
27	2.6	1.9	10.79	0.0
30	3.0	2.4	12.00	8.62
Effective Price of Care (\$)				
0	0.0	0.0	0.0	0.0
3	0.87	1.0	1.0	1.0
6	0.67	1.0	1.0	1.0
9	0.49	0.99	1.0	1.0
12	0.36	0.91	0.79	1.0
15	0.25	0.81	0.30	1.0
18	0.19	0.68	0.10	1.0
21	0.13	0.55	0.03	1.0
24	0.11	0.43	0.01	1.0
27	0.07	0.33	0.00	0.76
30	0.05	0.27	0.00	0.29

^aIn this case, $\ell^* = \$20$, $F = \$30$, $p(\ell^*) = 0.1$.
The deductible is the number of visits that must be made to satisfy the deductible.

^bAs in Table 1, $\ell^* = \$5$, $F = \$7.50$, $p(\ell^*) = 0.4$.
The deductible is the number of visits that must be made to satisfy the deductible.

that doctor-visiting is so sharply divided in plans with a deductible. When then is missing?

The most important cause for this difference between the real world and the model is probably the assumption that ℓ^* is always less than F . This assumption was made because the opposite case is trivial analytically--the patient will always seek care.¹⁹ Nevertheless, the trivial analytical case may occur fairly often, because illnesses vary widely in their severity, and a substantial portion may be sufficiently severe or threatening so that ℓ^* is greater than F .²⁰ Because the benefit from seeking care exceeds the physician's gross fee for many illnesses, there is undoubtedly less variance in demand for physician services among individuals than this model implies.²¹

However, our assumption that $\ell^* < F$ will sometimes occur; if this were not the case, we would observe no demand response to variation in a deductible. However there is such a response; in fact, work now in progress that estimates empirically the effect of varying levels of deductibles on demand shows considerable responsiveness for certain levels of deductibles (Newhouse and Murphy, forthcoming).

There is much interest in how the aggregate demand for medical care of a group of insured individuals changes as both their deductible

¹⁹Incorporating "serious" illnesses ($\ell^* > F$) into the model would show that such illnesses have an effect on demand even if they do not occur. Since the patient knows he may unavoidably have to see the doctor in the future, the effective price is always lower than the nominal price, though this was not the case in the simple model (as in the upper right-hand region marked "s" in Fig. 3). The consumer may approximate the bonus from spending on today's minor illness by calculating the expected future savings from today's expenditure should severe illnesses push his expenditure over the deductible. Even that calculation will underestimate the bonus because it does not take into account the fact that decisions on future expenditures do not have to be made today.

²⁰Our assumption that the cure is always fully and instantaneously effective artificially increases the incentive to seek care in our model. On the other hand, two of our assumptions reduce this incentive--that the distribution of diseases is fixed, i.e., that patients do not revise upward their estimated probability of future illness given an illness, and that untreated diseases do not have aftereffects. All of these assumptions seem less important distortions of reality than the mild illness assumption.

²¹Uninsured people do, after all, go to the doctor.

and coinsurance rate varies. To obtain estimates of this demand curve, numerical results would have to be obtained for a wide variety of deductibles, coinsurance rates, and demographic parameters, but we can infer that such a demand function would be qualitatively similar to the ones pictured in Fig. 6. The shape of Fig. 6 comes from combining an inelastic demand for care caused by severe illnesses with the results

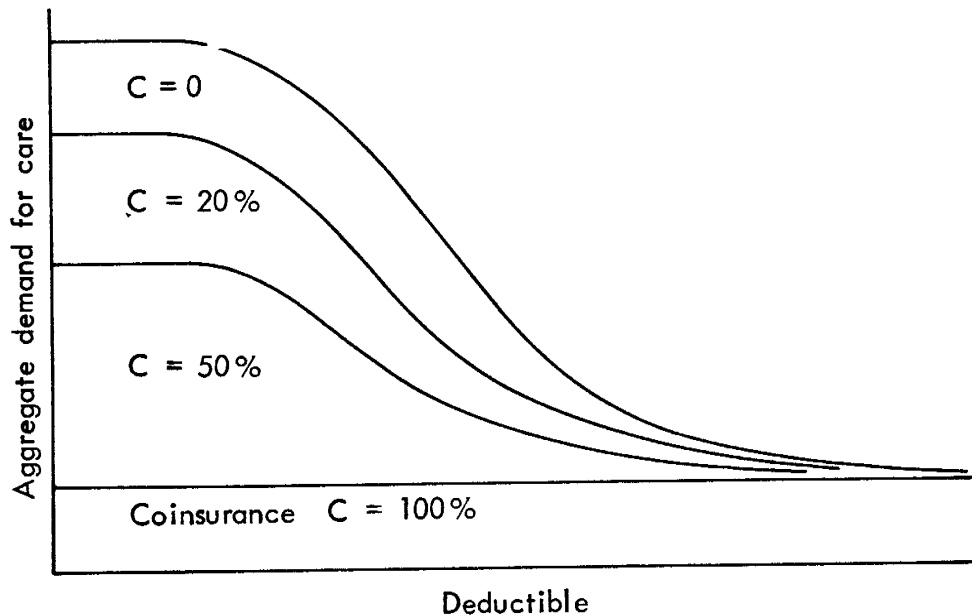


Fig. 6—Aggregate demand for care as a function of insurance parameters. The coinsurance rate C applies after the deductible has been satisfied.

of the simple model for minor illnesses.²² When the coinsurance rate is 100 percent, the consumer has no insurance and the size of the deductible is irrelevant. As the coinsurance rate falls, treatment of marginal illnesses will become increasingly desirable in the range of

²²Figure 6 was supported by running the model with the mild illness and an additional small probability of severe illness. The pattern of expected use was similar to those pictured. The results of Newhouse and Murphy (forthcoming) are also of this shape.

small deductibles.²³ Just as in Fig. 5, demand falls abruptly as the size of the deductible is increased for a certain region. Unlike Fig. 5, demand does not fall all the way to the no insurance demand level; the reason is that even with a large deductible, some people will have a very severe illness early in the accounting period, exceed their deductible, and then treat all of their minor illnesses until the next accounting period starts. Thus, the slope of the aggregate demand curve is not exactly zero even in the range of large deductibles.

Continuous Models

In the simple model, the consumer is limited to a binary decision--to buy or not to buy a stipulated service. Because of the discrete nature of doctor visits, this seems more reasonable than a continuous choice of treatment.²⁴ However, for aggregate demand curves, which have the most policy interest, it makes little difference whether individual consumers face a binary or a continuous decision. Conceptually, the case of a continuous choice of treatment and continuous distribution of disease severity is not much more complicated, but to compute results, we must approximate the continuous space of parameters by a finite space.

In our later discussion, we assume that the results for the simple model are qualitatively the same as those in the continuous case. We believe this because we find the results plausible, and because they held in all the variants of the simple model that we tried out.

An Application to Unemployment Compensation: An Extension of the Model

The simple model can be applied to an unemployed worker who is unsure of his ability to get another job quickly. Unemployment compensation plans have a fixed cutoff date that is based on previous employment experience. Each week the unemployed worker receives benefits, and the period of future benefits is reduced by one week. Unlike the case of

²³ Demand is not infinite when the coinsurance rate is zero because the money paid to the doctor is only a part of the costs of seeking care, and inconvenience and time costs remain. [Phelps and Newhouse (1974) and Acton (1973) provide estimates of the time-price elasticity of demand for physician visits.]

²⁴ A consumer cannot buy five cents' worth of a doctor's advice.

medical insurance, the "bonus" from drawing unemployment compensation benefits (i.e., not accepting a job) is negative, and represents the expected costs of replacing the worker's unemployment insurance package by one which is one week shorter. The model implies that instead of a step function that drops to zero exactly when the benefit period ends, uncertainty about the worker's ability to find another job makes effective compensation benefits fall smoothly to zero as the end of the benefit period is approached. For an individual who feels that he will have difficulty finding a new job, effective benefits toward the end of the period may be quite low, so that the pressure to find a job may be as great as it would be if he had no unemployment benefits. This feature complements other forces leading to an increasing tendency to end the search process (E. S. Phelps, et al., 1970).

IMPLICATIONS OF OUR THEORY FOR ESTIMATING THE DEMAND FOR MEDICAL CARE

The results of our model have important implications for estimating demand curves for medical care services. Typically, researchers have attempted to explain variation in the annual consumption of medical services. The price variable is usually defined to be out-of-pocket expense divided by total expenditure. Our model implies that this is a misspecification, if the policy includes a deductible.

In our model the consumer makes sequential decisions each time he becomes ill. Thus, for the i th person's j th illness episode, the demand is

$$q_{ij} = f(p_{ij}, Y_i, Z_j) + u_{ij},$$

where q_{ij} is the quantity demanded by the i th person in the j th episode; p_{ij} is the effective price for the j th episode, as given by a schedule of the kind shown in Fig. 5; Y_i, Z_j are vectors of other variables affecting demand; and u_{ij} is a random error term, with $E(u_{ij}) = 0$, $\text{Var}(u_{ij}) = \sigma^2 \Omega$.

Given a deductible and an accounting period, our model specifies that each p_{ij} is a function of demand prior to the j th episode and of time remaining in the period, i.e.,

$$p_{ij} = f \left(\sum_{k=1}^j q_{ik}, t(j) \right),$$

where $t(j)$ is the time at which the j th episode occurs.

The usual mode of analysis seeks to explain variation in demand aggregated temporally; i.e.,

$$q_i = \sum_{j=1}^T q_{ij},$$

where T is the number of episodes occurring during the year. Information about variation in T (or variation in health stock loss) is obviously germane to explaining variation in q_i , although existing data generally can measure T only imperfectly because of difficulties in measuring health stock loss. For ease of exposition we shall assume that there is no variation in T across individuals, because unmeasured variation in T is not the problem on which we wish to focus. With this assumption, explanation of variation in total demand, q_i , is equivalent to explanation of variation in average demand per episode, q_i/T . We shall speak as if the variable to be explained is q_i/T .

In the case of a linear model, the correct price to use is the average of the effective prices

$$p_i = \sum_{j=1}^T p_{ij}/T.^{25}$$

²⁵ Even if the p_{ij} were observable, it would be inappropriate to use Ordinary Least Squares estimators of demand under these circumstances. Because each p_{ij} is a function of a past episode's demands, each is correlated with past error terms u_{ik} , $i = 1, \dots, j$. The aggregate error term is the average of all u_{ik} 's, so that p_i would be correlated with the demand curve's average error

$$u_i = \sum_{j=1}^T u_{ij}/T.$$

Instrumental variable estimators would be required to remove this correlation.

Existing demand studies do not use p_i as a measure of price; instead, they use the fraction of annual expenditures paid out of pocket, \bar{p}_i (Feldstein 1971; P. Feldstein and Severson 1964; Davis and Russell 1972; Fuchs and Kramer 1972; Rosett and Huang 1973) or the marginal price for the last unit purchased, p_{mi} (Newhouse and Phelps 1974b). Both \bar{p}_i and p_{mi} measure the true price p_i with error if there are deductibles present in the insurance policies of some persons in the sample.

We shall first analyze the inconsistency caused by using \bar{p}_i as an explanatory variable. For explanatory convenience, we assume that there is no coinsurance ($C = 0$) after the deductible D , and that the market price for care is P per unit. Figure 7a plots various prices against total expenditure, showing how the true aggregate price $\sum p_{ij}/T$ always lies below the annual average fraction paid out of pocket \bar{p}_i .

Figure 7b shows how the measurement error in price ($\bar{p}_i - p_i$) changes with total expenditure when \bar{p}_i is used as an explanatory variable. The error is always positive, and rises with expenditure until total expense reaches the deductible; the error then falls asymptotically to zero as expenditure approaches infinity. The size of the

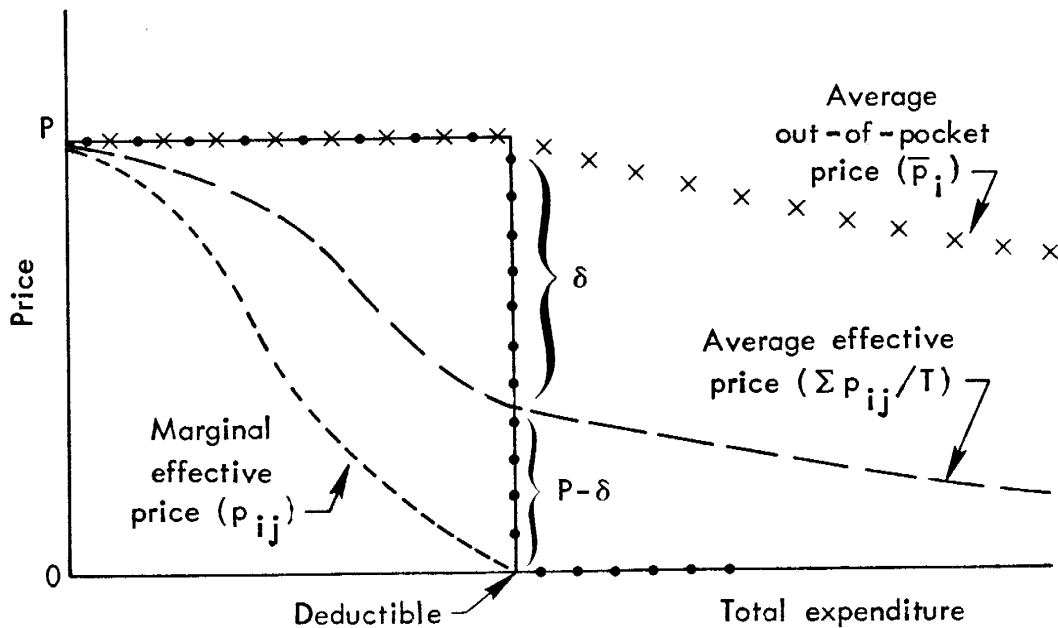


Fig. 7a — Relationship among total expenditure and average out-of-pocket price, average effective price, and marginal effective price.
The dotted solid line represents the nominal marginal price.

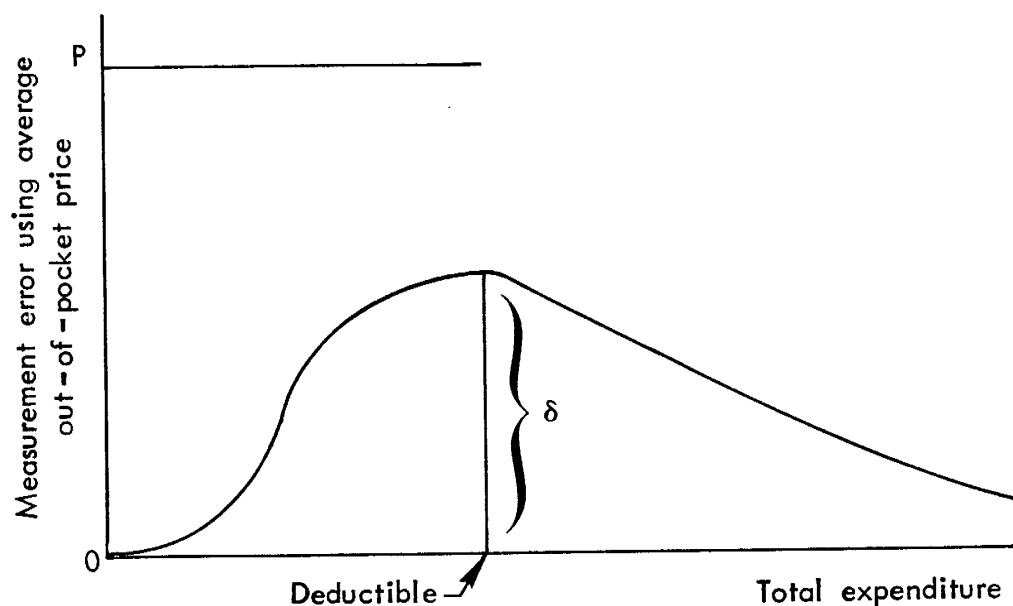


Fig. 7b — Measurement error using average out-of-pocket price as a function of total expenditure

maximum error δ depends on how rapidly p_{ij} falls with total expenditure (as determined by the dynamic programming model).

Because the error in measuring price is not random, but is generally correlated with the true price, the usual errors-in-variables result does not hold--the estimated price coefficient can be inconsistent either toward or away from zero, depending on the sample, especially on the distribution of observations above and below the deductible (Newhouse and Phelps 1974a).

A similar result occurs when the marginal price p_{mi} is used to explain demand. For expenses below the deductible, the marginal price and the average annual price \bar{p}_i are identical (and = P); but for expenses above the deductible, the marginal price is the coinsurance rate times the gross price (here equals 0). As a result, using the marginal price results in a different error pattern, as shown in Fig. 7c. The measurement error is discontinuous at D , and the error becomes negative for total expense exceeding the deductible. This error term is also generally correlated with the theoretically desirable

$$p_i = \sum_{j=1}^T p_{ij} / T,$$

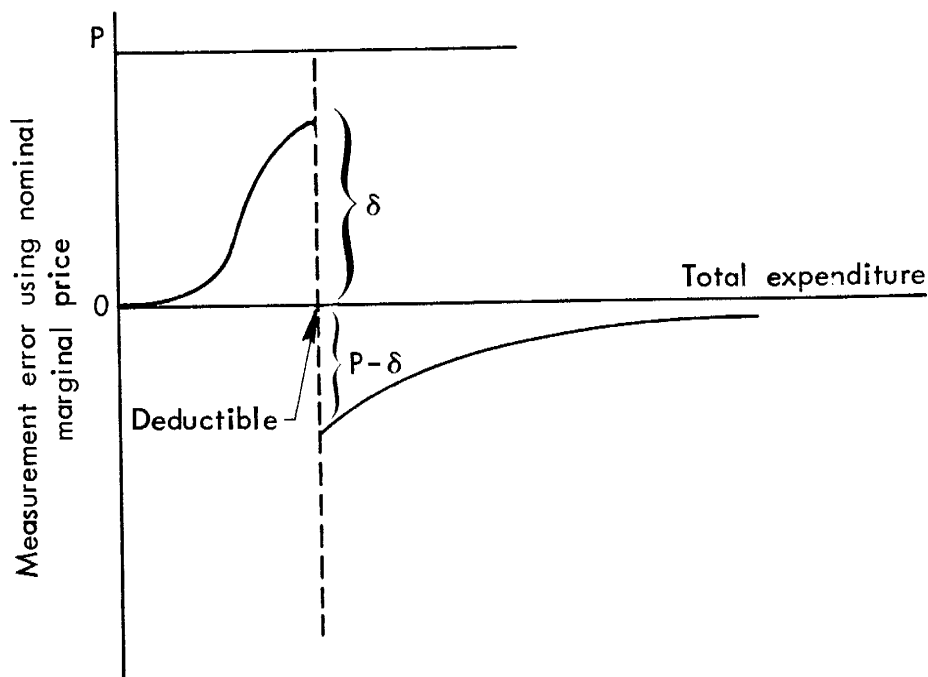


Fig. 7c—Measurement error using nominal marginal price as a function of total expenditure

so that inconsistencies in the price coefficient of indeterminate sign result, depending on the sample distribution of expenses, deductibles, and coinsurance.

Our analysis suggests that the correct method for analyzing medical demand when deductibles are present is by episode of illness, rather than by any temporal aggregation. Aggregate demand would then be a function of number of episodes, as well as demand for treatment per episode (Newhouse 1974). Our model also suggests that if analysis by episode of illness is not practical, individuals with deductibles in their policies should be excluded from the sample.²⁶ In this case, it will often be necessary to treat insurance as endogenous in order to

²⁶In the case of hospitalization, where common deductibles are almost always exceeded, admission decisions may be analyzed as a function of total out-of-pocket price, while length-of-stay decisions can be assumed to be made on the basis of the nominal marginal price, if the marginal price is not likely to change again as a function of total expenditure. In this situation, the inconsistency we have been discussing does not occur, and policyholders with deductibles may be kept in the sample. In effect, ℓ^* always exceeds F .

guard against a nonrandom distribution of insurance (i.e., that policyholders with deductibles systematically differ from those without deductibles). Preliminary work with the 1963 Center for Health Administration Studies Survey data shows that excluding policyholders with deductibles in their policies can lead to substantial changes in estimated price elasticities (changes both toward and away from zero) as compared with including such individuals and defining price as p_{mi} .²⁷ This result indicates that the inconsistency caused by not properly treating deductibles could be significant.

IMPLICATIONS OF OUR MODEL FOR INSURANCE POLICY DESIGN

Length of Period

One problem in insurance policy design is how the deductible should vary if the length of the insurance period is varied. Unfortunately, the simple model does not shed much light on this question. Figure 8 shows the tradeoff between the size of the deductible and the length of the insurance period when the percentage of episodes for which care is sought is held constant. As one might expect, given the simple model, except for very short periods and small deductibles (i.e., deductibles near the origin), the relation between D and time is linear; if the length of the accounting period is multiplied by a factor t , the deductible must also be multiplied by t to keep demand the same.²⁸

However, it can be shown that linearity does not hold if rarer, more severe illness episodes are possible. In this case, an annual policy with the same (actuarially fair) premium as four consecutive quarterly policies may have an effective deductible less than twice as big as the quarterly deductible, and the variance of out-of-pocket payments

²⁷For a description of the Survey, see Andersen and Anderson (1967).

²⁸Note that Fig. 8 also shows, in a somewhat different way, the sensitivity to varying the deductible in a small range and the lack of sensitivity outside that range. This can be seen by choosing a value for the length of the period and then varying the deductible D . Quite large variations in D may make essentially no difference in the fraction of illnesses treated, while for the small range of D between the 90-percent and 0-percent lines, there is a large change in demand.

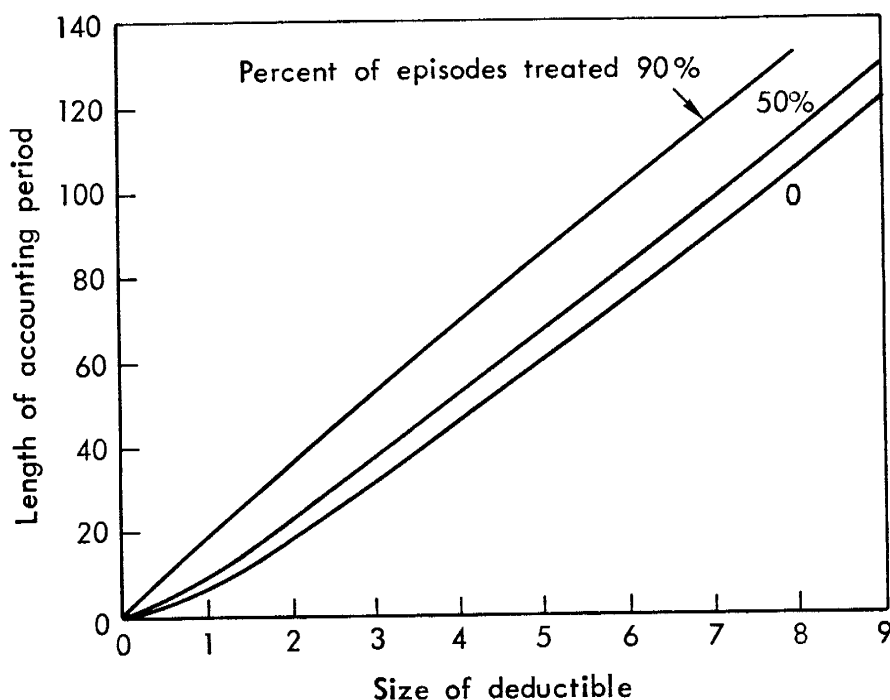


Fig. 8—Expected fraction of episodes treated as a function of deductibles and length of period. The figures are based on the results shown in Table 1.

is lower in the annual policy (Keeler, Relles, and Rolph, forthcoming). Thus, risk-averse consumers will want longer accounting periods. Similarly, in the choice between a family deductible and individual deductibles, the somewhat larger, but not proportionally larger, family deductible should be chosen.²⁹

Optimal Size of Deductible

The simple model is much more instructive for choice of deductible size than for length of the accounting period. The purpose of insurance is to reduce risks. A good policy will balance the value of risk reduction with any allocative inefficiencies resulting from lower prices, as well as additional administrative costs resulting from additional

²⁹This statement ignores the administrative complexities introduced by the possibility of a family unit change in the middle of an accounting period.

claims.³⁰ The simple model shows how to find policies that make this tradeoff efficiently; i.e., give maximal risk reduction for each level of price subsidy.

According to the model, demand (and hence any distortion) as a function of deductible size approximates a logistic curve. While the location of the curve will vary from individual to individual, each person should have (a) a region of "small" deductibles where demand is inelastic because the deductible has no practical effect, (b) a middle region where demand varies with the deductible, and (c) a region of "large" deductibles in which demand is again inelastic. As a result, there should be little difference in the effects of large deductibles on marginal decisions to seek care. For example, a \$300 annual per person deductible may preserve incentives almost as well as a \$3000 deductible (for evidence, see Newhouse and Murphy, forthcoming). Because the risk a person bears increases with the size of the deductible, he should under these conditions prefer one of the "smaller" large deductibles to a "larger" large deductible.³¹

It is unlikely that this conclusion would change with a less simple model. In particular, the existence of a very small probability of a future "catastrophic" expenditure is unlikely to alter this conclusion, because there should be little effective lowering of price from such a small probability event.³² For example, a patient is not likely to be

³⁰Our discussion assumes that only small changes in the total resources devoted to medical care are contemplated or that the supply curve is approximately perfectly elastic. For a treatment of the pure coinsurance case with no deductible that relaxes these assumptions, see Arrow (1973b).

³¹It is possible that demand is sensitive to small coinsurance rates above a large deductible; if so, it may be preferable that the consumer pay a small fraction of the bill. There is no empirical evidence on how demand responds to small coinsurance rates in the region of "expensive" illness.

³²It is possible that induced technological change in the so-called catastrophic end of the expenditure distribution would be sufficiently great to modify this conclusion; however, it is unlikely that substantial out-of-pocket payments could be optimal because of the risk the consumer would bear. [See Zeckhauser (1973a) for a discussion of the issue.]

more inclined to see a physician for a cold on the slight chance that he might need renal dialysis sometime during the year. Modification of the model to take account of administrative costs should also not affect the superiority of the large deductible over the very large, because the number of claims filed will be small in any case.

An additional implication can be drawn. Ignoring administrative costs for the moment, the same analysis applies to the region of small deductibles; once demand becomes insensitive to the level of a deductible, it improves welfare to reduce it--all the way to zero. Inclusion of administrative costs could raise the preferred deductible to a low, but positive level (i.e., one on the relatively inelastic portion of the demand curve). There are therefore likely to be two local optima: (1) a deductible just above the region in which demand is sensitive to variation in the deductible (Type 1) and (2) either no deductible or a deductible in the region below which demand is sensitive (Type 2). It is theoretically possible, although unlikely, that an optimal deductible would be in the middle range of deductibles where demand is elastic.

Assuming that either a Type 1 or Type 2 deductible is preferred to other deductibles, how does one choose between them? From the individual's point of view, the choice between these two deductibles depends on the balance among allocative efficiency, risk aversion, and administrative cost: (1) The greater any allocative inefficiency induced by the subsidy inherent in Type 2 (small) deductible, the more preferable is a Type 1 deductible.³³ The amount of inefficiency will depend, inter alia, on the magnitude of the difference in demand between Type 1 and Type 2 deductibles. (2) The less the consumer's aversion to risk, the more preferable is a Type 1 deductible. (3) The greater the marginal administrative costs, the more preferable is a Type 1 deductible.

Thus, theory can only lead to some likely restrictions; it cannot establish definitive conclusions. We can say that (from the individual's point of view) very large deductibles are not likely to be optimal, and that deductibles in the range of very elastic demand are also not likely

³³It is theoretically possible that the allocative inefficiency is greater with a Type 1 deductible than a Type 2 deductible if the medical care sector is a monopoly and other markets are competitive. [See Crew (1969).]

to be optimal; but we cannot, without additional empirical evidence, make confident statements about whether a moderately large deductible is better or worse than a very small or even no deductible.

All of our analysis concerns choice by an individual. From society's point of view, other considerations could enter into the choice of a deductible as part of a national health insurance plan. First, some believe that the consumer is not rational, in the sense that he does not make good choices. Specifically, it is argued that the consumer does not know enough to seek efficacious preventive care and can be induced to do so by financial means, i.e., by eliminating a deductible. However, there is little evidence to support these claims, particularly the link between additional health care services and improved health status (Newhouse, Phelps, and Schwartz 1974). Second, national health insurance is seen by many as a means of redistributing income. The amount of redistribution across income classes will depend importantly on the means of financing used (e.g., general revenue versus payroll); however, there is no doubt that health insurance *can* be used as a means of redistributing income, and that the potential amount of income redistribution is greater as deductibles fall toward zero. But, whether health insurance is a desirable means for redistribution (as opposed to income transfers) is questionable. Third, national health insurance can be used as a vehicle to implement public control of resource allocation in medical care (i.e., public determination of budgets of health institutions and providers). The absence of a deductible in a national plan is not a necessary condition for public sector control, but such control is more likely to be desirable with no deductible than with a large deductible because the usual market restraints will not be present. Whether the medical marketplace should be abolished and the role of the public regulator made even more prominent is a major public policy issue, but outside the scope of this report.

IV. CONCLUSION

In this report, we have examined the theory of a consumer who faces a price schedule that varies with the number of units bought, and who has an opportunity to make more than one decision to purchase a good. We have developed our theory in the context of health insurance deductibles, but the model is of greater generality.

The comparative statics of the formal dynamic programming model that determines the effective price of medical care when a deductible is present were too complicated to work out. Nevertheless, our simulations with a simplified model show that deductibles are likely to have significant effects.

The relationship between demand and the size of a deductible in the simplified model approximates a logistic curve, a result we believe will hold under less restrictive assumptions. If one is analyzing annual physician visit expenditures among individuals with deductibles, there is an inconsistency in the estimated effect of price of indeterminate sign from using either the average or marginal price as an explanatory variable. This inconsistency can be overcome if expenditures by episode are analyzed.

Because demand for care and administrative costs are likely to be insensitive to variation in the deductible above a certain level, deductibles larger than a certain amount will not be optimal. They impose a cost of risk-bearing for no return. A deductible just above the level where demand is quite sensitive may be optimal. However, below a certain level, further allocative effects from decreasing the deductible are again small. Deductibles in this range, or even full-coverage, may be optimal if claim processing or other administrative costs are not large; such deductibles promise the greatest risk reduction, but at a price in terms of allocative efficiency.

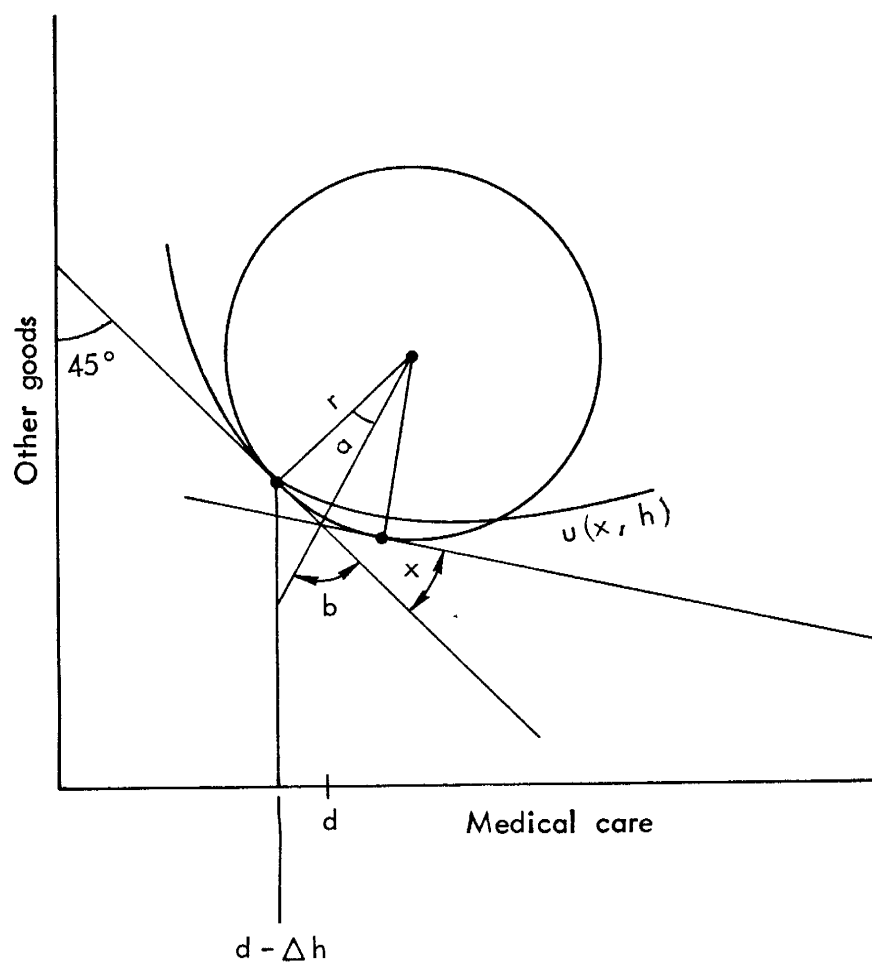


Fig. A.1 — Geometry of purchases near the kink

Since $p_h = p_x = 1$, $\lambda = u_x$ and these expressions simplify to

$$\eta = 1/hf''(h) = R/[1 + f'(h)^2]^{3/2} \cdot h.$$

At $d - \Delta h$, $f'(h) = -1$, since the indifference curve is tangent to the budget line. Thus

$$\eta = R/h2\sqrt{2}.$$

Finally, if the indifference curve misses the lower price line, $R < r$ and so

$$\eta < \frac{\Delta h}{2h \tan [(\pi/4 - \tan^{-1} C)/2]} .$$

This expression simplifies to

$$\eta < \frac{\Delta h \{1 + C + [2(1 + C)^2]^{1/2}\}}{2h(1 - C)} .$$

Some representative values are tabulated in Table A.1.

Table A.1

UPPER BOUND FOR ELASTICITIES IMPLIED
BY OBSERVED PURCHASES

$\Delta h/h$	Coinsurance (C) =			
	0	0.2	0.25	0.5
0.1	0.121	0.165	0.181	0.308
0.05	0.060	0.083	0.090	0.154
0.01	0.012	0.017	0.018	0.031

Appendix B

A SEQUENTIAL DECISION PROBLEM: THE WILDCATTER

In this appendix, we demonstrate the techniques of sequential decisionmaking through the analysis of a simple problem. We will use the problem of an oil wildcatter as an example because we think that it will be easier to grasp the principle involved than if we were to use a health insurance example. The substantive problem is similar in health insurance, however.

The Problem

An oilman owns some potentially oil-holding fields. He can perform some geological tests, or he can just start to drill. The tests will reduce his uncertainty about whether or not there is oil under the land. His questions are: Should he do nothing, test, or drill? To make a rational choice, the oilman has to decide how much various outcomes are worth to him, and what their probabilities of occurrence are. His original problem can then be broken down into a number of simpler problems.

Utility of Outcomes

First, the oilman has to decide how to evaluate uncertain outcomes. Suppose that with no geological test, he estimates that his chances of a wet hole are 0.1 and of a dry hole, 0.9. Moreover, suppose that a wet hole is worth \$1.1 million, drilling costs \$100,000, and the geological tests cost \$20,000. The money payoffs and probabilities are shown in the decision tree illustrated in Fig. B.1.

A sufficiently rich oilman may well decide to drill and not test, figuring that his average money gain is $0.1 (1,000,000) + 0.9 (-100,000) = 10,000$, which is positive. But if the oilman is not so wealthy, he may be *risk-averse*, i.e., he may not want to take fair gambles or even better than fair if he "can't afford to lose" \$100,000. By mentally considering which of a variety of such gambles he would take, the oilman

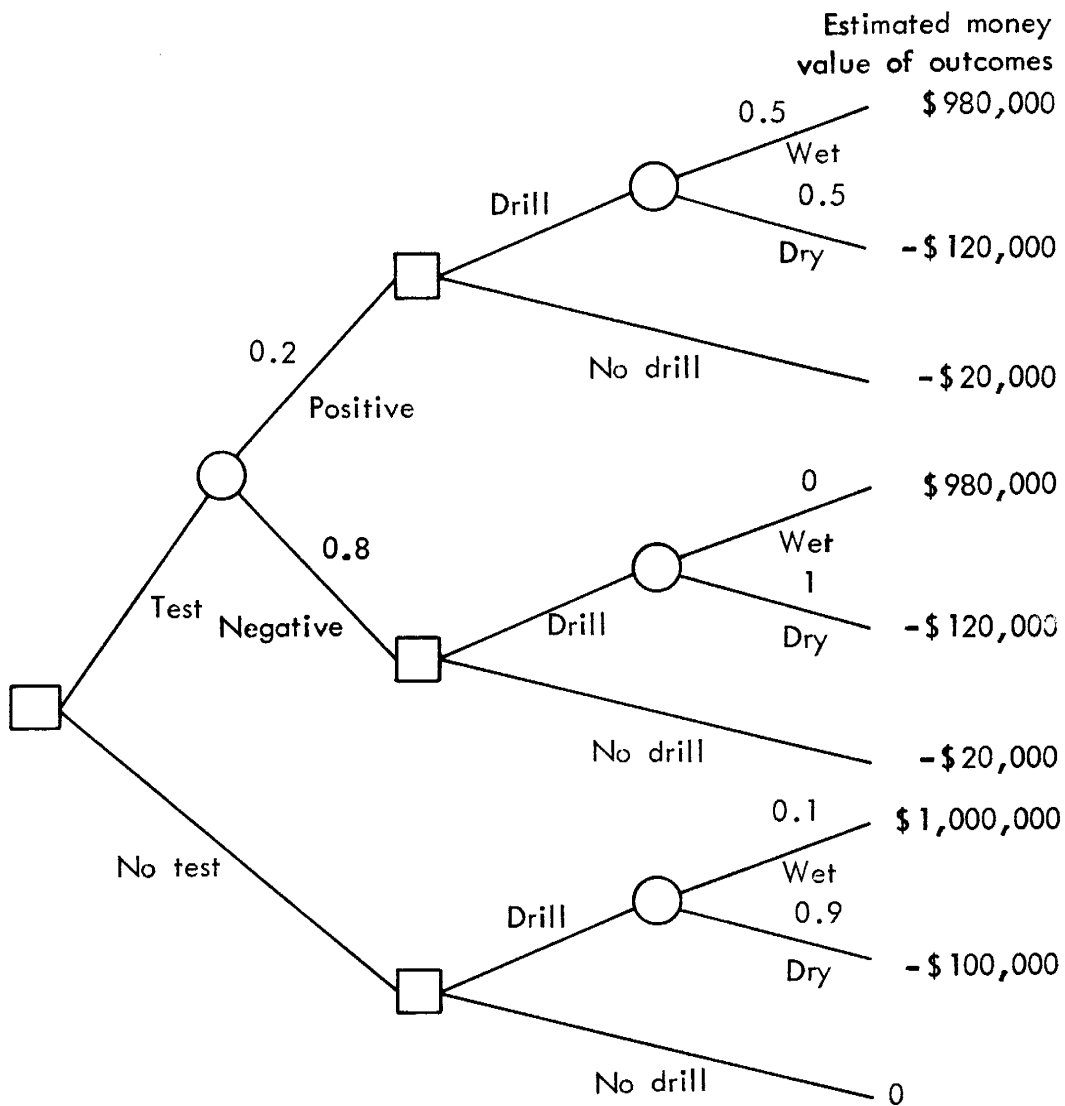


Fig. B.1—The oil wildcatter's problem (the numbers above the branches of the tree represent probabilities)

can compute his utility function for money. Thus, if he is indifferent between \$1 million for sure and a 50-50 chance at \$3 million or nothing, he could say the utility of \$3 million is twice that of \$1 million to him.

The advantage of utility measures over money values in analyzing decisions with uncertainty is that utility can be averaged. Suppose, for example, the oilman gives a utility value of 10 to \$1 million,

5 to the status quo, and 4 to $-\$100,000$, as illustrated in Fig. B.2.³⁵ Suppose, also, that testing is impractical. If the probability of a wet hole is 0.1, the utility of drilling is $0.1 (10) + 0.9 (4) = 4.6$. Since 4.6 is less than 5, he should not drill.

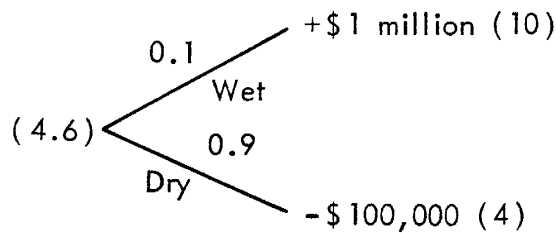


Fig. B.2 —The utility of a lottery

Probabilities

Before decisions can be made, every relevant future development must be assigned a probability. In many cases we can show that the correct action is the same no matter what probability is estimated for some relatively unimportant event.

Often it will be wise to spend resources on improving estimates, as indeed it is in our example. At different places in the tree, probabilities for what seems like the same event may differ. Thus after an encouraging test, the chances of a dry hole are less than they would be if no testing were done. In technical terms, the probabilities are conditional on the circumstances preceding them. This is illustrated in Fig. B.1, where the probability of a wet hole after a positive test differs from the probability after a negative test or no test.

³⁵The utility numbers do not have intrinsic meaning. For example, we might assign to status quo a value of 0, and to $\$1$ million a value of 20. Then all the rest could be computed, and $-\$100,000$ would have a value -4 . Any positive linear transformation of utility preserves the needed qualities.

Solving the Problem

In Fig. B.3, the monetary outcomes have been translated into utilities (shown in the last column). The nodes where probabilities must be estimated are marked by a circle. The other (square) nodes indicate where decisions must be made. Once utilities and probabilities have been assigned, the rest of the analysis is straightforward. Starting from the last nodes, we work backward toward the front of the tree, making the best decision at each node. For example, at the top the

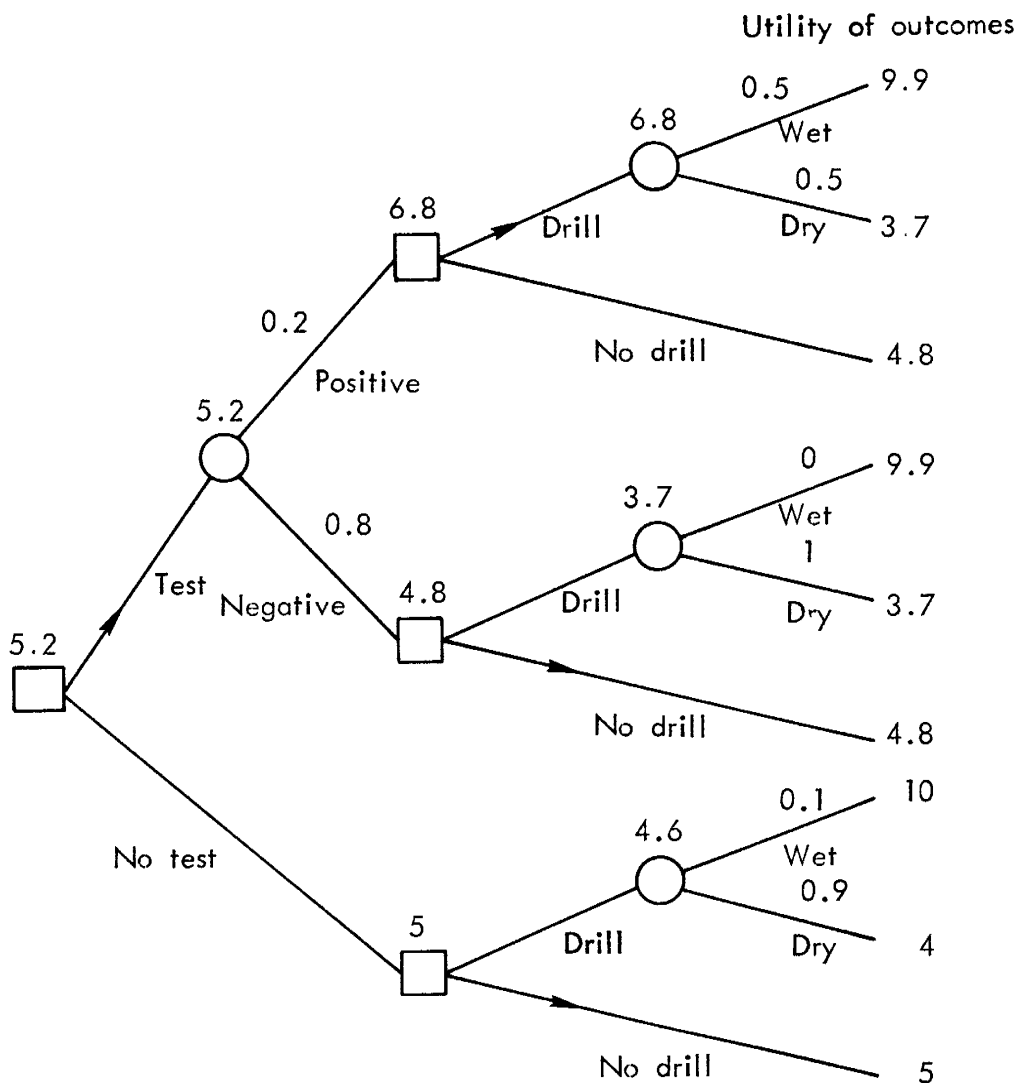


Fig. B.3 — The solution of the oil wildcatter's problem

value of drilling after a positive test is 0.5 (value of a wet hole) + 0.5 (value of a dry hole) = 6.8 . Since not drilling after a positive test is worth only 4.8 , one should drill in that circumstance. We can label the value of a positive test result as 6.8 , and mark the best decision at that point with an arrow (on the drill edge). For prior decisions, the only thing that matters is the value 6.8 , and what follows that node can be ignored until the plan is put into operation. Similarly, in the negative test branch of the tree, drilling has a value of 3.7 , and not drilling has a value of 4.8 . Not drilling is better, so we can label the negative test node with a 4.8 and put in the arrow on the not drilling branch. The value after a test is the weighted average of the possible results, and $0.2(6.8) + 0.8(4.8) = 5.2$. Looking at the lower part of the tree, we see that if there is no test, it is best not to drill, and so the status quo value of 5 is the value of no test. Finally, since $5.2 > 5$, the wildcatter should test.

Let us summarize the steps in making a sequential decision:

1. Consider the whole problem before making any decisions.
2. Estimate the probabilities of future developments, and value the outcomes.
3. Work backwards (i.e., leftwards) from the last (chronological) decision, making decisions one at a time.

At each node, one can assume that a decision is worth what it would be if one makes the best decisions from that node rightward. Thus, at any node, it is only necessary to consider the next rightward nodes of the decision tree. Eventually, one will have worked from right to left and back to the current decision.

In the case of deductibles and health insurance, instead of weighing the probabilities of a dry hole and the value of a wet and dry hole, one must consider the probabilities of illness and the potential value of consumption on future days. This is precisely what the dynamic programming model does.

Appendix C

PROOFS OF PROPERTY 3 AND PROPERTIES 6 THROUGH 10

Emmett B. Keeler

Let $A(n,d)$ be the action taken when the consumer becomes sick on day n , with d unfree visits left. $A(n,d)$ can take the values s (suffer) or r (get relief). To avoid ambiguity, we assume that if both actions are optimal, the consumer suffers. We assume $0 < \ell^* < F$ and $n \geq 0$.

Property 3: Let the path traced through Fig. 3 (text) by the consumer's actions be the action path. In the matrix $A(n,d)$, illustrated by Fig. 3, the action paths $(s \ r)$ or $\begin{pmatrix} r \\ s \end{pmatrix}$ are impossible. (For example, once having chosen s , r is never subsequently chosen.) Thus the matrix can be partitioned as shown in Fig. 3.

Proof. Since $0 < \ell^*$, for all n ,

$$A(n,0) = r. \tag{1}$$

Since $\ell^* < F$, for $d \geq 1$,

$$A(1,d) = s. \tag{2}$$

Let k be the smallest value of $n + d$ for which $A(n - 1, d) = r$, and $A(n, d) = s$; or $A(n, d - 1) = s$ and $A(n, d) = r$. We shall show by contradiction that k does not exist. From (1) and (2), we have $k \geq 3$.

Case 1. Suppose k occurs with $A(n - 1, d) = r$, and $A(n, d) = s$. Since k is a minimum, $A(n - 1, d - 1)$ must be r . In this case, we will show that the actions $A(n - 1, d - 1) = s$, $A(n - 1, d) = s$, and $A(n, d) = r$ have the same expected cost and thus are also optimal. Consider the possibilities shown in the table below.

Illness			Action Path Taken, in Days $\binom{n-1}{n}$		Costs, Given Action Path Taken	
$n - 1$	n	Probability	$\binom{r}{s}$	$\binom{s}{r}$	$\binom{r}{s}$	$\binom{s}{r}$
Yes	Yes	p^2	$(d - 1, n - 2)$	$(d - 1, n - 2)$	$(\ell^* + F)$	$(F + \ell^*)$
Yes	No	$p(1 - p)$	$(d - 1, n - 2)$	$(d, n - 2)$	F	ℓ^*
No	Yes	$p(1 - p)$	$(d, n - 2)$	$(d - 1, n - 2)$	ℓ^*	F
No	No	$(1 - p)^2$	$(d, n - 2)$	$(d, n - 2)$	0	0

The only difference between action paths $\binom{r}{s}$ and $\binom{s}{r}$ is when there is an illness in exactly one of the two periods, but the expected costs of the two action paths are identical. By the principle of optimality, s and r must be equally good for $A(n - 1, d)$; and by our assumption on ties, $A(n - 1, d) = s$, in contradiction to the Case 1 hypothesis.

Case 2. Suppose that k occurs with $A(n, d - 1) = s$ and $A(n, d) = r$. Since k was the minimum value of $n + d$, we must have $A(n - 1, d - 1) = s$. Just as in Case 1, we may show that $\binom{r}{s}$ is equally good, and hence that $A(n, d) = s$ is also optimal. Again our assumption that ties go to suffering contradicts the hypothesis. Q.E.D.

It may seem that the tie-breaking rule is ad hoc and being used too strongly in the proof. If we don't assume the rule, the theorem changes to "there exists an optimal matrix A such that neither of the impossible action paths appears." The proof of property 3 goes through, because whenever there is an optimal impossible action path, we can replace it by its reversal, which must be also optimal.

Property 6: If $d \geq 2$,

$$V(n, d) - V(n, d-1) \leq V(n, d-1) - V(n, d-2). \quad (3)$$

Proof. By induction on d and n . First suppose that $d = 2$. If $A(n, 1) = s$, then $A(n, 2) = s$ by property 3, and so both sides of (3) are $p\ell^*$. If $A(n, 1) = r$, then $V(n, 1) = (1 - p)V(n - 1, 1) + pF$. Since

$V(n,0) = 0$, we must show $V(n,2) \leq 2V(n,1)$. This is true for $n = 0$. Suppose that it is true for $n - 1$. Then

$$\begin{aligned} V(n,2) &\leq (1 - p)V(n - 1,2) + pV(n - 1,1) + pF \\ &\leq 2(1 - p)V(n - 1,1) + pV(n - 1,1) + pF \\ &\leq 2(1 - p)V(n - 1,1) + pF + pF \\ &= 2V(n,1). \end{aligned}$$

Now, suppose that (3) is true for all $d' < d$, and for all (n',d) with $n' < n$. If $A(n,d - 1) = s$, then $A(n,d) = s$, and so the left-hand side is $np\ell^* - np\ell^* = 0$. If $A(n,d - 1) = r$, then $A(n,d - 2) = r$. It is always true that $V(n,d)$ is less than or equal to its value if $A(n,d)$ is constrained to equal r . Thus $V(n,d) - V(n,d - 1) \leq (1 - p)[V(n - 1,d) - V(n - 1,d - 1)] + p[V(n - 1,d - 1) - V(n - 2,d - 2)]$. Also, $V(n,d - 1) - V(n,d - 2) = (1 - p)[V(n - 1,d - 1) - V(n - 1,d - 2)] + p[V(n - 1,d - 2) - V(n - 1,d - 3)]$. The $(1 - p)$ part of the inequality follows from induction on n , and the p part from induction on d .

Property 7: If $d, n \geq 1$,

$$V(n,d) - V(n - 1,d) \geq V(n,d - 1) - V(n - 1,d - 1). \quad (4)$$

Proof. When $d = 1$, the right-hand side is 0, so (4) is true in that case. If $A(n,d) = s$, the left-hand side is $p\ell^*$. Since the one-day gap can never be more than $p\ell^*$, (4) holds in this case. If $A(n,d - 1) = s$, by property 3, $A(n,d) = s$. The only remaining case is that in which $A(n,d - 1)$ and $A(n,d)$ both equal r . Then

$$V(n,d) - V(n - 1,d) = -pV(n - 1,d) + pV(n - 1,d - 1) + pF.$$

By property 6, this is

$$\begin{aligned} &\geq -pV(n-1, d-1) + pV(n-1, d-2) + pF \\ &= V(n, d-1) - V(n-1, d-1). \end{aligned}$$

Property 8:

$$V(n, d) - V(n-1, d) \leq V(n-1, d) - V(n-2, d) \quad \text{for } n \geq 2. \quad (5)$$

Proof. When $d = 0$, both sides of the inequality are 0. Suppose that $d > 0$. If $A(n-1, d) = s$, then the right-hand side is $p\ell^*$ and the inequality is true. If $A(n-1, d) = r$, by property 3, $A(n, d) = r$. Then

$$V(n, d) - V(n-1, d) = -pV(n-1, d) + pV(n-1, d-1) + pF,$$

which, by property 7,

$$\begin{aligned} &\leq -pV(n-2, d) + pV(n-2, d-1) + pF \\ &= V(n-1, d) - V(n-2, d). \end{aligned}$$

Property 9: Let k be a positive integer. Let $v(n, d)$ be the expected utility of a consumer with a one-day illness probability of p . The argument d is the number of visits the consumer must make to satisfy the deductible. Let his illness cost $k\ell^*$, and a cure, kF . Let $V(n, kd)$ be the expected utility of a consumer with probability kp of being ill, a cost ℓ^* of being ill, and a cure of cost F . (kd is the number of visits that must be made to satisfy the deductible.) Then for all $n, d \geq 0$,

$$v(n, d) \leq V(n, kd). \quad (6)$$

Proof. By induction. Both sides are 0 when n or d is 0. Suppose that (6) is true for all $n' < n$ and (n, d') with $d' < d$.

$$\begin{aligned} v(n, d) = & (1 - p)v(n - 1, d) + p \min[v(n - 1, d - 1) \\ & + kF, v(n - 1, d) + k\ell^*], \end{aligned} \quad (7)$$

$$\begin{aligned} V(n, kd) = & (1 - kp)V(n - 1, kd) + kp \min[V(n - 1, kd - 1) \\ & + F, V(n - 1, kd) + \ell^*]. \end{aligned} \quad (8)$$

By property 6, V is concave in d , so that

$$\begin{aligned} k[V(n - 1, kd - 1) - V(n - 1, kd)] & \geq V(n - 1, kd - k) \\ & - V(n - 1, kd). \end{aligned} \quad (9)$$

Add $(kp - p)[V(n - 1, kd)]$ to the first term of (8) and subtract it from the second term to get

$$\begin{aligned} V(n, kd) = & (1 - p)V(n - 1, kd) + p \min[k[V(n - 1, kd - 1) \\ & - V(n - 1, kd)] + kF \\ & + V(n - 1, kd); V(n - 1, kd) + k\ell^*]. \end{aligned} \quad (10)$$

Applying (9) to the first term in min,

$$\begin{aligned} V(n, kd) \geq & (1 - p)V(n - 1, kd) + p \min[V(n - 1, kd - k) \\ & + kF, V(n - 1, kd) + k\ell^*]. \end{aligned} \quad (11)$$

By induction, the right-hand side of (7) is less than or equal to the right-hand side of (11), and so $v \leq V$.

Property 10. Let k be a positive integer and let $p \leq 1/k$. Let $V(nk, d)$ be the cost of being at (nk, d) when the one-day probability of being sick is p . Let $v(n, d)$ be the cost of being at (n, d) when the probability of being sick is pk . Then, for all $n, d \geq 0$,

$$V(nk, d) \leq v(n, d). \quad (12)$$

Proof. By induction. Both sides are 0 when either n or d is 0. Suppose that (6) is true for all $n' < n$ and (n, d') with $d' < d$. We will show that $V(nk, d) \leq v(n, d)$ irrespective of whether the optimal action at (n, d) is to suffer, or to buy relief. Since V is the cost of the optimal action, $V(nk, d) \leq V_s(nk, d)$ where V_s is the result of suffering from nk through $nk - k + 1$. But $V_s(nk, d) = V(nk - k, d) + pk\ell^*$. Therefore, $V(nk, d) \leq V_s(nk, d) = V(nk - k, d) + pk\ell^* \leq v(n - 1, d) + pk\ell^* = v(n, d)$, if $A(n, d) = s$. On the other hand, $V(nk, d) \leq V_r(nk, d)$, where V_r is the result of getting relief from nk through $nk - k + 1$. But $V_r(nk, d) = (1 - p)^k V(nk - k, d) + k(1 - p)^{k-1}p(V(nk - k, d - 1) + F) + \dots + p^k(V(nk - k, d - k) + kF)$. (Those terms, if any, in which the exponent of $p \geq d$ have the value of dF .)

Jensen's inequality states that for any concave function g , any set of values x_1, x_2, \dots, x_n in the domain of g , and any non-negative set of numbers $\alpha_1, \alpha_2, \dots, \alpha_n$ such that $\sum \alpha_i = 1$, we have $g(\sum \alpha_i x_i) \geq \sum \alpha_i g(x_i)$. By property 6, the function $g(j) = V(nk - k, d - j) + jF$ for $0 \leq j \leq d$ and $g(j) = dF$ for $j > d$ is concave in j for all $j \geq 0$. We can write $[V_r(nk, d) - (1 - kp)V(nk - k, d)]/kp$ as

$$\frac{(1 - p)^k - (1 - kp)}{kp} g(0) + \frac{k(1 - p)^{k-1}p}{kp} g(1) + \dots + \frac{p^k}{kp} g(k).$$

Therefore, Jensen's inequality implies

$$\begin{aligned} [V_r(nk, d) - (1 - kp)V(nk - k, d)]/kp &\leq g\left(\frac{(1 - p)^k - (1 - kp)}{kp} \cdot 0 + (1 - p)^{k-1} \right. \\ &\quad \left. \cdot 1 + \frac{(k - 1)(1 - p)^{k-2}p}{2} \cdot 2 + \dots + p^{k-1}\right) \\ &= g(1) = V(nk - k, d - 1) + F. \end{aligned}$$

In summary, $V(nk, d) \leq V_r(nk, d) \leq (1 - kp)V(nk - k, d) + kpV(nk - k, d - 1) + kpF \leq (1 - kp)v(n - 1, d) + kp v(n - 1, d - 1) + kpF = v(n, d)$, if $A(n, d) = r$.

BIBLIOGRAPHY

- Acton, Jan P., *The Demand for Health Care Among the Urban Poor with Special Emphasis on the Role of Time* (Santa Monica, California: The Rand Corporation, R-1151-OEO/NYC), April 1973.
- Andersen, Ronald, and Odin W. Anderson, *A Decade of Health Services* (Chicago: University of Chicago Press, 1967).
- Arrow, Kenneth J., *Optimal Insurance and Generalized Deductibles* (Santa Monica, California: The Rand Corporation, R-1108-OEO), February 1973a.
- Arrow, Kenneth J., "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 53:5, December 1963, pp. 941-973.
- Arrow, Kenneth J., *Welfare Analysis of Changes in Health Coinsurance Rates* (Santa Monica, California: The Rand Corporation, R-1281-OEO), November 1973b.
- Coase, Ronald, "The Theory of Public Utility Pricing and Its Application," *The Bell Journal of Economics and Management Science*, 1:1, Spring 1970, pp. 113-128.
- Crew, Michael, "Coinsurance and the Welfare Economics of Medical Care," *American Economic Review*, 59:5, December 1969, pp. 906-908.
- Davis, Karen, and Louise B. Russell, "The Substitution of Hospital Out-patient Care for Inpatient Care," *Review of Economics and Statistics*, 54, 1972, pp. 109-120.
- Ehrlich, Isaac, and Gary S. Becker, "Market Insurance, Self Insurance, and Self Protection," *Journal of Political Economy*, 80:4, July-August 1972, pp. 623-648.
- Feldstein, Martin S., "Hospital Cost Inflation: A Study of Nonprofit Price Dynamics," *American Economic Review*, 61, 1971, pp. 853-872.
- Feldstein, Martin S., "The Rising Price of Physicians' Services," *Review of Economics and Statistics*, 52, 1970, pp. 121-133.
- Feldstein, Paul, and Ruth Severson, *Report of the Commission on the Cost of Medical Care* (Chicago: The American Medical Association, 1964).
- Fox, Bennett L., "Discretizing Dynamic Programs," *Journal of Optimization Theory and Applications*, forthcoming.
- Friedman, Bernard, "Risk Aversion and the Consumer Choice of Health Insurance Option," *Review of Economics and Statistics*, 56:2, May 1974, pp. 209-214.

- Fuchs, Victor R., and Marcia J. Kramer, *Determinants of Expenditures for Physicians' Services in the United States, 1948-1968* (New York: National Bureau of Economic Research, 1972).
- Grossman, Michael, *The Demand for Health* (New York: Columbia University Press, 1972).
- Health Insurance Association of America, *A Current Profile of Group Medical Expense Insurance in Force in the United States, December 1970* (New York: The Association, February 1972).
- Keeler, Emmett B., Daniel Relles, and John E. Rolph, *Family and Individual Deductibles* (Santa Monica, California: The Rand Corporation, R-1393-OEO), forthcoming.
- Newhouse, Joseph P., "Issues in the Analysis and Design of the Experimental Portion of the Health Insurance Study," *Proceedings of the Social Statistics Section*, American Statistical Association, 1974.
- Newhouse, Joseph P., and Maureen Murphy, *An Empirical Estimate of the Impact of Deductibles on the Demand for Medical Care Services*, forthcoming.
- Newhouse, Joseph P., and Charles E. Phelps, "New Estimates of Price and Income Elasticities for Medical Care Services," in Richard N. Rosett, ed., *The Role of Health Insurance in the Health Services Sector*, Universities-National Bureau of Economic Research Conference Series, forthcoming.
- Newhouse, Joseph P., and Charles E. Phelps, *On Having Your Cake and Eating It Too: Econometric Problems in Estimating the Demand for Health Services* (Santa Monica, California: The Rand Corporation, R-1149-NC), April 1974a.
- Newhouse, Joseph P., and Charles E. Phelps, "Price and Income Elasticities for Medical Care Services," in Mark Perlman, ed., *The Economics of Health and Medical Care: Proceedings of a Conference Held by the International Economic Association at Tokyo* (London: Macmillan, 1974b).
- Newhouse, Joseph P., Charles E. Phelps, and William B. Schwartz, "Policy Options and the Impact of National Health Insurance," *The New England Journal of Medicine*, 290:24, June 13, 1974, pp. 1345-1359.
- Oi, Walter Y., "A Disneyland Dilemma: Two-part Tariffs for a Mickey-Mouse Monopoly," *Quarterly Journal of Economics*, 80:1, February 1971, pp. 77-96.
- Parish, Ned F., Statement of the National Association of Blue Shield Plans Before the Ways and Means Committee, U.S. House of Representatives, May 13, 1974.

- Phelps, Charles E., *The Demand for Health Insurance: A Theoretical and Empirical Investigation* (Santa Monica, California: The Rand Corporation, R-1054-OEO), July 1973.
- Phelps, Charles E., and Joseph P. Newhouse, *Coinsurance and the Demand for Medical Services* (Santa Monica, California: The Rand Corporation, R-964-OEO/NC), April 1973.
- Phelps, Charles E., and Joseph P. Newhouse, "Coinsurance, the Price of Time, and the Demand for Medical Services," *Review of Economics and Statistics*, 56:3, August 1974, pp. 334-343.
- Phelps, Charles E., and Joseph P. Newhouse, *Effects of Coinsurance on the Demand for Physician Services* (Santa Monica, California: The Rand Corporation, R-976-OEO), June 1972.
- Phelps, Edmund S., et al., *Microeconomic Foundations of Employment and Inflation Theory* (New York: W. W. Norton & Company, Inc., 1970).
- Raiffa, Howard, *Decision Analysis* (Reading, Massachusetts: Addison-Wesley, 1968).
- Rosett, Richard N., and Lien-fu Huang, "The Effect of Health Insurance on the Demand for Medical Care," *The Journal of Political Economy*, 81, 1973, pp. 281-305.
- Scitovsky, Anne. A., and Nelda M. Synder, "Effect of Coinsurance on Use of Physician Services," *Social Security Bulletin*, 35, 1972, pp. 3-19
- Zeckhauser, Richard J., "Coverage for Catastrophic Illness," *Public Policy*, 21:2, Spring 1973a, pp. 149-172.
- Zeckhauser, Richard J., "Medical Insurance: A Case Study of the Trade-off Between Risk Spreading and Appropriate Incentives," *Journal of Economic Theory*, 2, March 1970, pp. 10-26.
- Zeckhauser, Richard J., "Time as the Ultimate Source of Utility," *Quarterly Journal of Economics*, 87:4, November 1973b, pp. 668-673.