Health Economics: A Modern Approach

Randall P. Ellis Boston University

Chapter 2 Production of Health

This chapter develops conventional models of benefits and costs that lead to models of demand and supply. We intermingle critiques of this conventional model and then discuss alternative behavioral mdoesl in 2 which try to capture the decisions of Humans rather than the choices of Econs, the stereotypical perfectly rational optimizer.

2.1 Utility of health and consumer demand for medical care

Conventional economists predominantly model consumers as maximizing an objective function subject to constraints, and health economists often start with the same framework. We first focus on the consumer's problem, holding supply prices fixed. The usual assumption is that health care goods (doctors office visits and lab test, for instance) give no direct utility or losses, but only affect consumers through their effects on health. Using a static (one period) framework, as in Mark Pauly's (1980) classic article we can write consumer utility U as a function of health H and the quantity of all other goods and services purchased, Y.

(2.1) $$U = U(H, Y)$$

A very general formulation for a health production could be

(2.2) $$H = H(H_0, X, Y, T_M, A, I, Z, R, B, G, E, \varepsilon)$$

Where

H_0 = the consumer's initial health status, from the previous period

X = vector of medical care inputs purchased to influence health,

Y = vector of all non-health goods and services, which might also influence health (cigarettes, alcohol, fitness clubs…)

T_M = time spent obtaining medical care

A = Access to health care providers, including distance, quality, availability

I = Income (or wealth)

Z = Patient's demographic variables (age, gender, education, race, etc.)

R = taste parameters, including discount rate and degree of risk aversion

B = Beliefs, influenced perhaps by framing, anchoring, and other concepts from Behavioral Economics

G = genetic composition

E = environment (air, water, food, sanitation, weather, sun)

ε = health shocks, which may be multidimensional

In this framework, health is not purchased directly, but can instead be thought of a produced intermediate product that provides utility. Eventually we will explore many of these dimensions that affect health and hence the demand for health care, but for now, it is better to greatly simplify and only model one health care services *x* and one other good, *y* instead of an array of them. Equation (2.2) then becomes

(2.3) $H = H(H_0, x)$

Substituting equation 2.3 into the consumers direct utility function, yields a utility function that can be written as

(2.4) $U = U(\mathrm{H}(H_0, \mathrm{x}), y) = \widetilde{U}(H_0, \mathrm{x}, y)$

Conventionally, the consumer's problem is to maximize (2.1) subject to the budget constraint (2.5)

(2.5) $I = p_x\, x + p_y\, y$

As discussed in Appendix A, this conventional consumer maximization problem can be represented graphically and used to characterize the consumer's optimal choice.

A more formal solution to this is presented in Text Box 2.1 and shown graphically below, where the optimum is at O*.

Figure 2.1a Constrained utility maximization for an uninsure <u>healthy</u> consumer A ($H^A{}_0 > H^B{}_0$), showing income consumpti curve (ICC^A)

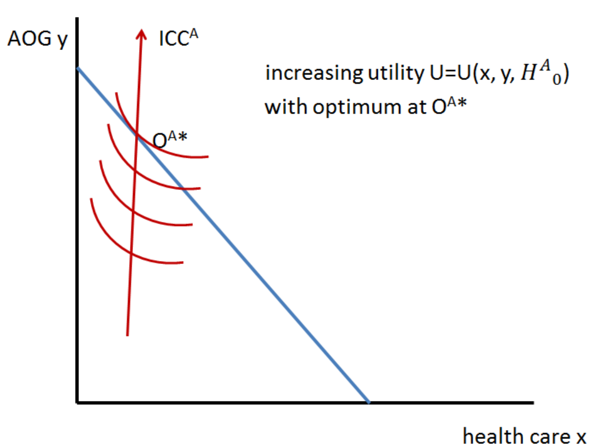


Figure 2.1b Constrained utility maximization for an uninsured <u>sick</u> consumer B ($H^A{}_0 > H^B{}_0$), showing income consumption curve (ICC^B)

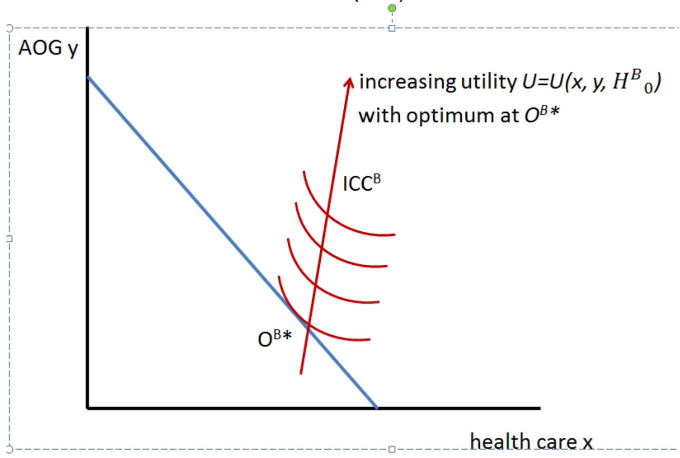


Figure 2.1 illustrates several key features of health care choices, tastes and budget constraints for two stylized individuals. First even though the two individuals have identical budget constraints, the sicker individual without insurance will consume a much higher amount on health care, and as a result have less to spend on all other

goods (AOG). Second, for both the healthy and the sick consumer, the level of spending is relatively insensitive to changes in income. This is revealed by the steepness of the income consumption curve (ICC) for both individuals, which traces out the optimal level of spending on health care and AOG for different levels of income holding prices the same. Third, the relatively sharp kink in the indifference curves near the optimal consumption bundles indicates that the quantity of health demanded is relatively insensitive to changes in the relative price of health care. For the indifference curves shown, one can visualize that even lowering the relative price of health care by 50 percent (a huge change!) changes the quantity of health care demanded by only a few percent. Empirical magnitudes of how inelastic health care is of presented in chapter XXX. Finally, although not drawn in, the property of very income inelastic demand (i.e., small changes in health care use as income falls) cannot be true for very low income levels for an uninsured consumer. In a dynamic model they might be able to borrow, but in a static something very different has to happen for very low income levels, greatly distorting their choice of health care inputs.

A behavioralist critique of this conventional model: independence of prices and consumer utility
Many criticisms of this framework will come up in the course of this textbook. The one that we will focus on here is that consumer utility can be separated entirely from prices and budgets. Behavioral economics has demonstrated repeatedly that consumer happiness is not independent of prices and framing issues resulting from price expectations. For example, for high prices can be interpreted as a signal of quality.(cite XXX) Or the structure of how a good is sold (offering good X at a high price before setting the actual price of X). Or expanding the choice set offered may change consumer preferences, even if the alternative options are never optimal choices. Or consumer preferences (as summarized in an indifference curves) are very fickle and sensitive to many other factors.

Later we will use much richer models, highlighting that consumers rarely make health care choices independently of their doctors and other health care providers. Plus, insurance completely changes the choice process.

(Derive the Price Consumption curve and hence the demand curve here before going on. Show two demand curves on one set of axes.)

Text Box 2.1 A more formal approach, with conventional and behavioral interpretations.

Conventional Analysis

Mathematically this problem can also be solved numerically for properties using the Lagrangian technique to capture constrained optimization, the derivation of the more formal solution of which I leave as an exercise. Letting capital letters denote vectors while lower case letters with subscripts denote scalars the problem is:

$$\max_{X,Y,\lambda} \mathcal{L} = U(\mathrm{H}(H_0, \mathrm{X}), Y) + \lambda[I - P_x X - P_y Y]$$

Optimality implies that the marginal utility of each good divided by its price should be equal to the marginal utility of income:

(B2.1) $\frac{U_{y_i}}{p_1} = \lambda$

The solution also implies that with multiple health care goods, optimality requires that levels of medical care inputs be chosen so that the marginal health products of a dollar spent on each medical service x_i is equal to its price p_i.

(B2.2) $\frac{\partial H/\partial x_1}{p_1} = \cdots = \frac{\partial H/\partial x_n}{p_n} = \frac{\lambda}{U_{\mathrm{H}}}$

Where $\gamma \stackrel{\text{def}}{=} \frac{\lambda}{U_{\mathrm{H}}}$ is the shadow price (or value) of an increase in health.

For well-insured individuals, total spending on health care, $m \stackrel{\text{def}}{=} p_x\, x$, will be small and the marginal utility of income λ will change only slightly with x. (Cases where spending is large relative to income are considered later in Chapter XXX.) For the same reasons that economists often use consumer surplus to the average consumer to approximate welfare change, even though they know it is inexact at an individual level, it is often convenient to focus on the case of only one dimension of medical services, x, and to work with the benefit of health care function B(x). A formal definition would be

(B2.3) $\mathrm{B(x)} \stackrel{\text{def}}{=} \frac{U(H(x)) - U(H(0))}{\lambda}$

A Behavioralist Critique: Health is not uni-dimensional, consumers are heterogeneous

This conventional framework and equation (B2.2) justifies thinking of the production of "health" as something that can be modeled and chosen independently of consumer preferences: income and taste variation may result in different levels of Y and H being chosen, but there is still only one efficient way of producing H (known by doctors!). This ignores that some health care inputs may be painful, or consumers may vary in their relative value of particular dimension of health, x, or y. The total benefit formulation takes this model a step further and also ignores variation in the marginal utility of income λ.

Given that the collection of goods in all other goods in the vector Y is diverse, it is convenient to assume there is only one such good, *y*. For most purposes, it loses nothing in generality to normalize the price this of one "all other good" (AOG) composite to be 1, and hence $p_y = 1$. The variable y then denotes spending on AOG which is convenient to put on figures as a reminder. Now x can also be thought of as a composite health care service with many components, each of which will have a market price, p_x. Later we will examine that the prices charged for each of these services can vary across consumers and perhaps be chosen by health care providers, but with only one health care good x, it is convenient to focus on total health care spending , m:

(2.6) $$m = p_x\, x.$$

Using these simplifications, the above budget constraint can be written as simply:

(2.7) $I = m + y$ = spending on medical care + spending on all other goods

Doesn't this mean that prices are held constant? Yes, as written. But we will soon be introducing insurance which is easier to introduce without worrying about the prices of each good or service. There is still much to be done before introducing price choices.

Note that after normalizing by the marginal utility of income, this benefit of medical care function measures benefits in dollar terms. This leads to the much more familiar and convenient partial equilibrium formulation of the consumer optimization problem when each unit of x can be purchased at a fixed price p_x. In this case the maximization problem is

(2.8) $$\max_x \mathrm{B}(x) - p_x \mathrm{x}$$

A tremendous number of very useful results can be derived from this simple specification. The classic one is the efficiency condition:

(2.9) $$\mathrm{B}'(\mathrm{x}) - p_x = 0$$

This consumer maximization problem is very similar to the social optimization problem which can be written as

(2.10) $$\max_x \mathrm{B}(x) - \mathrm{C}(x)$$

Clearly if there is an optimum then it requires:

(2.11) $$\mathrm{B}'(\mathrm{x}) - \mathrm{C}'(\mathrm{x}) = 0 \text{ and } \mathrm{B}''(\mathrm{x}) - \mathrm{C}''(\mathrm{x}) < 0$$

Marginal benefit equal to the marginal cost at the social optimum, and a sufficient condition is that the total benefit function is concave while the total cost function is nonconcave. If we assume that the price is equal to the marginal cost, $P_x = \mathrm{C}'(\mathrm{x})$, and that the consumer acts as a pure price taker, then the marginal benefit function also becomes the consumer's inverse demand function. (We explore cases where consumers do not act as price takers in chapter XXX.) We will also find it convenient to assume that the marginal cost of medical care is constant, so that $\mathrm{C}'(\mathrm{x}) = c$

A graphical example of the same result is shown in Figures 2.1 and 2.2 below. The case shown it for a constant marginal cost, which can alternatively reflect a constant marginal production cost (the social problem) or a constant purchase price (the consumer problem). Figure 2.2 adopts the convenient assumption that the marginal benefit and cost functions can also be written as $b(x) = B'(x)$, $\mathrm{c}(\mathrm{x}) = \mathrm{C}'(\mathrm{x})$. The case shown is also the convenient one where the marginal cost of medical care is constant, so that $\mathrm{C}'(\mathrm{x}) = c(x) = c$. If the objective is to maximize the difference between the benefit and cost function, then the solution will be at a point such as m^1 where the slopes of the total curves are equated, or where the two marginal curves intersect. If the price to the consumer is zero, or costs are irrelevant to the consumer, then the consumer maximum will occur at a point such as $m^0$where the slope of the total benefit curve is zero, or the marginal benefit curve intersects the horizontal axis. It will be useful to remember x^1 as the consumption point when the consumer has to pay the full marginal cost hence $x^1 = x(1)$ and $x^0 = x(0)$ as the consumption point when the price is zero. Also, we will use $\beta \equiv \frac{x^1}{x^0}$, which is spending on medical care it is at full price as a proportion of spending on medical care when it is free.

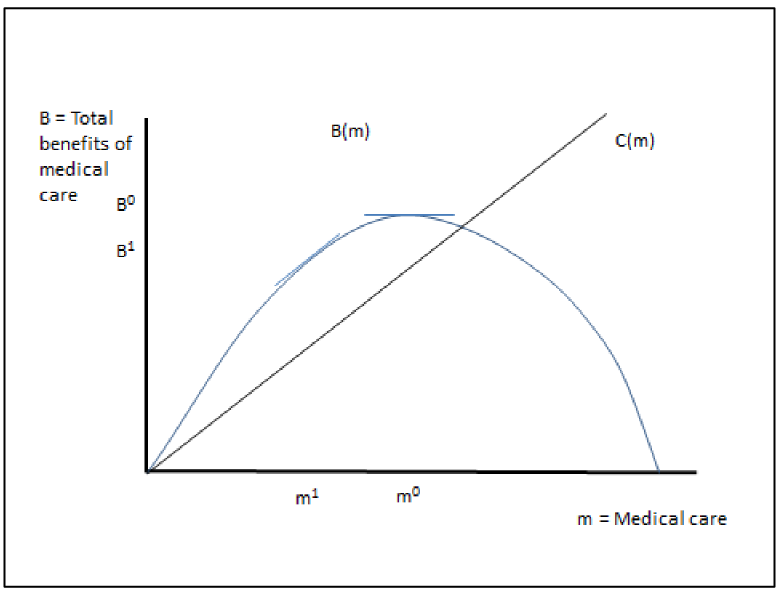


Figure 2.4 Consumer's total benefit and total cost functions

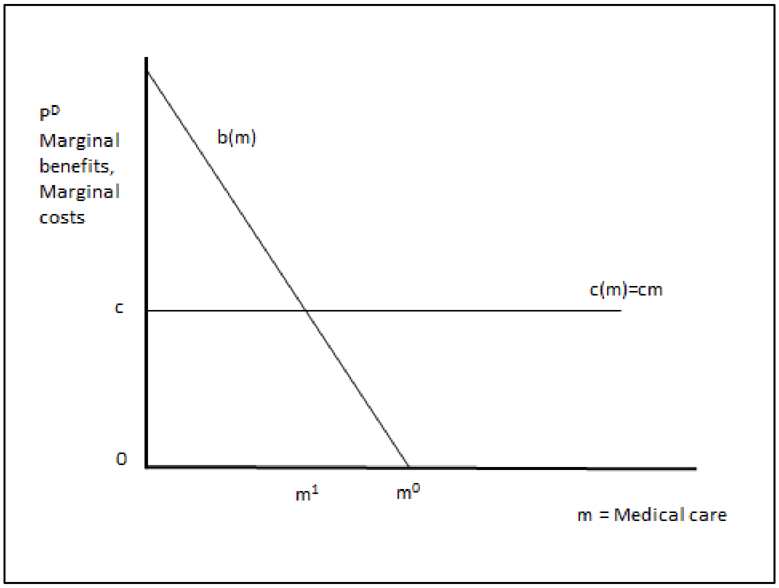


Figure 2.2 Consumers marginal benefit and cost functions

When focusing on the consumers consumption problem and how it is affected by insurance, it will often convenient to think in terms of how much the consumer pays as a fraction of the marginal cost of a unit of medical care. This is equivalent to assuming that the marginal cost of medical dare is one. The essence of most insurance programs is that the consumer will pay only a fraction of the marginal cost, which we will consistently call θ (theta). Figure 2.3 reflects this convenient normalization, which we will use extensively. A numeric example using this framework is included as problem 2.1 at the end of the chapter.

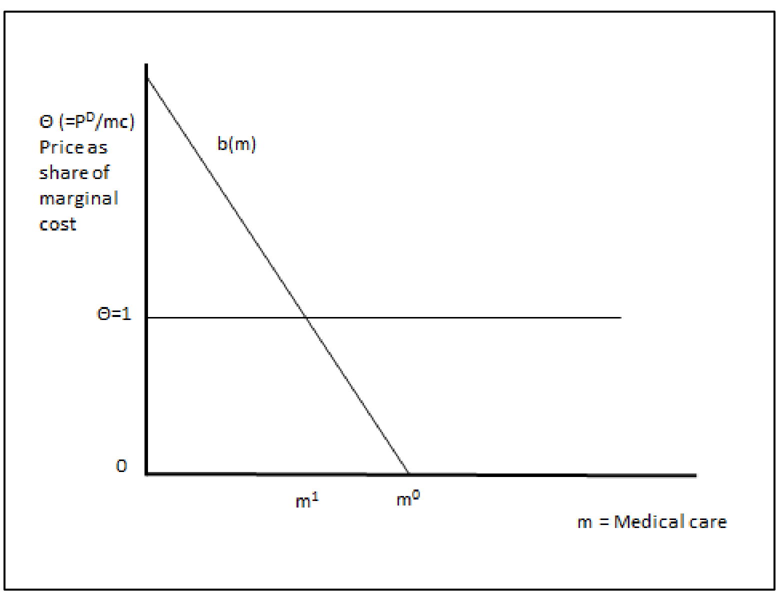


Figure 2.3 Consumer's marginal benefit and marginal cost functions after normalizing prices by MC

If one assumes that the consumer knows the marginal benefit function, acts as a price taker and solely determines the quantity of medical care received, then the marginal benefit function b(m) will also be the individual demand curve. As with conventional goods, the change in Consumer Surplus (ΔCS) can be a useful approximation of the true Compensating Variation (CV), that is, the dollar value of a change in the utility from changing the consumer's consumption away from the private optimum where b(m)=c(m), which happens at m^1. Using the notation we have just developed,

$$(2.12) \qquad \Delta CS \; = \; [\, B(m^1) - \, B(m^0)] - \; [\, C(m^1) - \, C(m^0)]$$

Or equivalently, using integrals,

$$(2.13) \qquad \Delta CS \; = \; [\, \int_{m^0}^{m^1} [b(m) - c(m)] dm$$

Figure 2.4 illustrates this concept of consumer surplus. Note that so far we have been using b(m) which is the inverse demand function $P^D(m)$, whereas it is often more convenient to work with the conventional demand curve, which is $m(P^D)$ or equivalently $\mathrm{m}(\theta)$ if the insurance copayment is used as the price. If this individual demand function is linear, then one convenient way of expressing it is

$$(2.14) \qquad \mathrm{m}(\theta) = \; \mathrm{m}_0 - \theta(\mathrm{m}_0 - \mathrm{m}_1) = \mathrm{m}_0 - \theta(\mathrm{m}_0 - \beta \mathrm{m}_0) = \mathrm{m}_0 \; \left(1 - \theta(1 - \beta)\right),$$

where $\beta \equiv \frac{m_1}{m_0}$ is the level of health care spending at a full price as a fraction of spending when care is free.

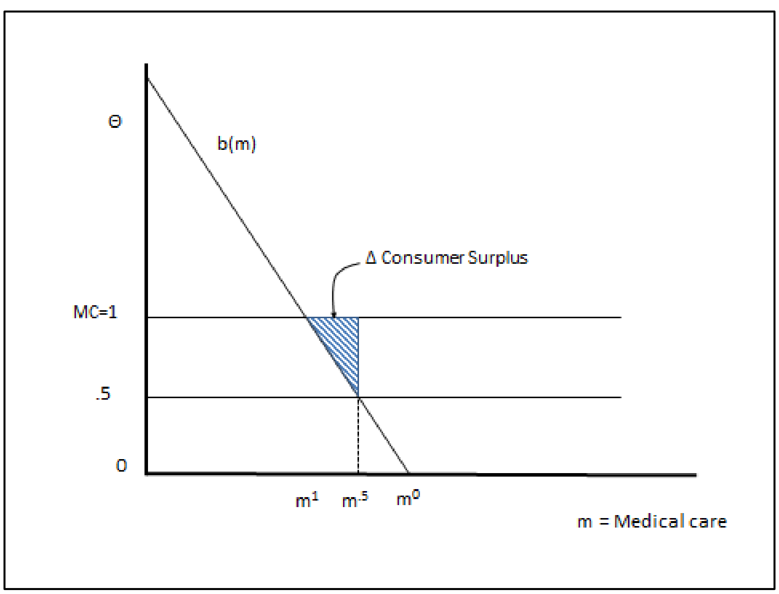


Figure 2.4 Consumer surplus loss from lowering price to half of MC

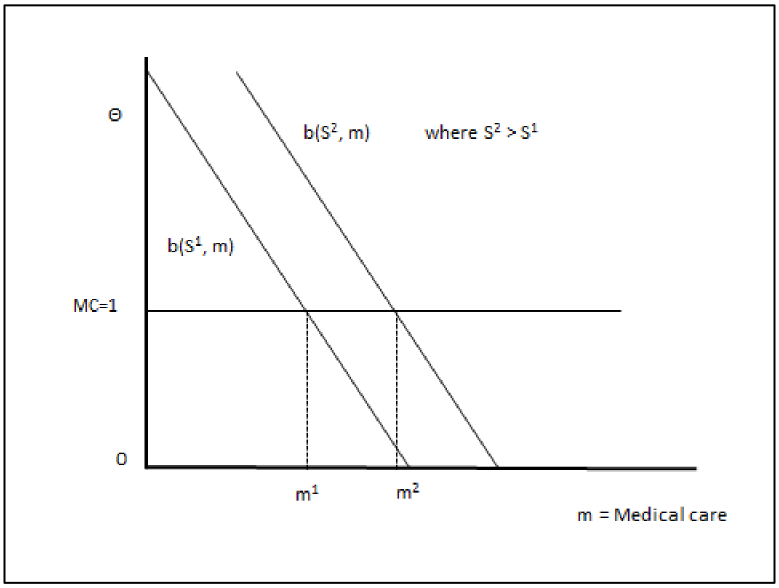


Figure 2.5 Effect of an exogenous shift in medical care demand curve on quantity demanded

Note that S can be any of the large number of shift factors identified at the start of this article that influence health and hence health demand, and the interpretation of consumer surplus change using a graphical approach may be misleading. If the increase in demand resulted from higher income, then it is plausible to infer that the increase in total surplus reflects an improvement in patient welfare. However if the shift upward in the curves reflects a worsening in the patient's health, or a worsening in the environment (air, water, etc) then the increase demand is reflecting a worsening in the patient's welfare. A partial equilibrium approach just using demand curves will be misleading.

2.2 Empirical Estimates of Market Demand Curves

Corresponding to the analytical estimates of demand curves shown in the preceding figures, empirical estimates of demand are abundant. The most influential insights come from a large controlled experiment done in the 1970's, the RAND Health Insurance Experiment, in which 5,800 individuals were enrolled over up to five years in a wide range of health plans. Although the plans examined differed in several dimensions, the most important variation across plans was in the coinsurance rate, i.e., the share of costs that the consumer had to pay up to a moderate deductible. Table 2.1 presents results from an early and highly influential paper by Will Manning et al that appeared in the American Economic Review in 1987. Results in the top half of the table show results in their natural units, while in the bottom half all results have been normalized by dividing by the level of spending when care was free. Figure 2.6 summarizes the results from the bottom half of Table 2.1, better highlighting the essential nonlinearity of observed demand. (Largely because hospitalizations are rare, the estimated demand for inpatient services is not precisely estimated, and which may explain the non-monotonic pattern for inpatient care.) Table 2.2 and Figure 2.2 present results for various subsets of outpatient of interest.

Price Responsiveness from the Rand HIE
Manning et al, 1987
Dollars (1991$)

Cost sharing	**Outpatient Expenses**	**Inpatient Expenses**	**Probability of any use**	**Total Health Spending**
free	488	531	86.7	1019
25%	379	447	78.8	826
50%	308	456	74.3	764
95%	282	418	68	700

Ratios of observed spending to free care levels

Cost sharing	**Outpatient care**	**Inpatient Care**	**Probability of any use**	**Total Health Spending**
0	1.00	1.00	1.00	1.00
25%	0.78	0.84	0.91	0.81
50%	0.63	0.86	0.86	0.75
95%	0.58	0.79	0.78	0.69

Table 2.1 Estimates of demand response from a controlled experiment in which insurance copayments were varied (Manning et al, 1986b, 1988, and Newhouse 1991, Tables 3.3 and 3.24)

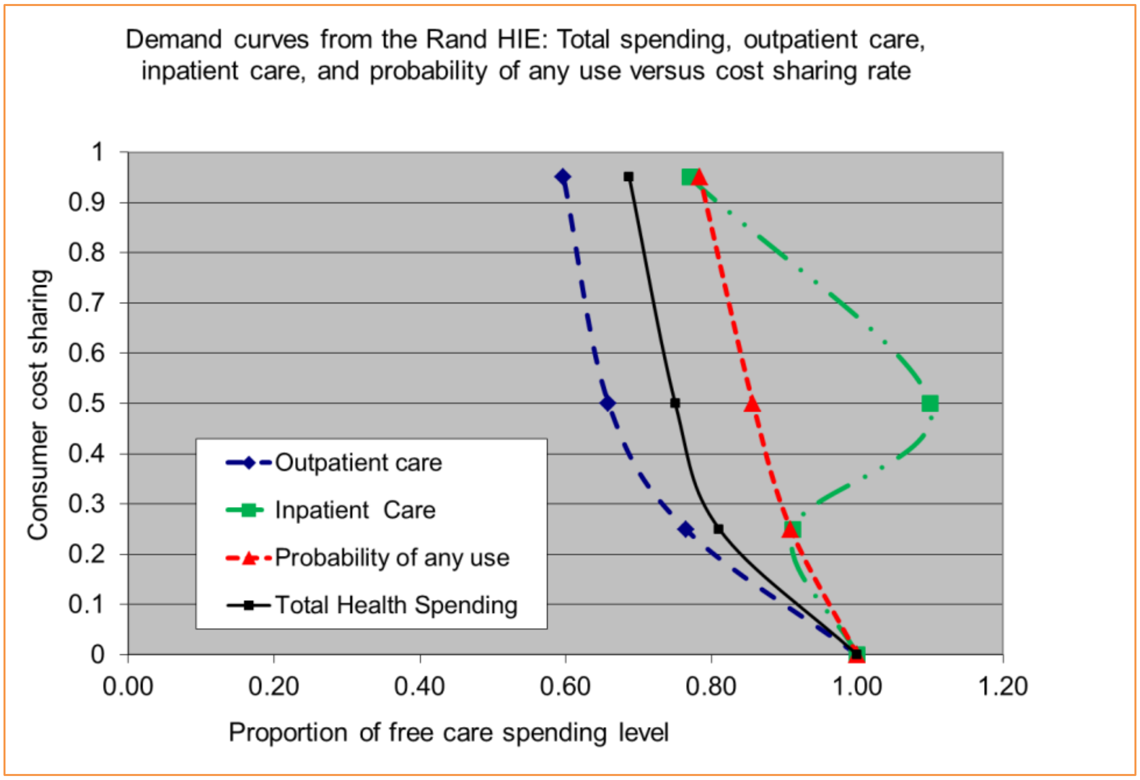


Figure 2.6 Demand Curves from Rand Health Insurance Experiment from Table 2.1

Price Responsiveness from the Rand HIE
Newhouse, 1993
Dollars (1991$)

Cost sharing	Dental Care	Preventive Care	Office visits	Drugs	Mental Health
free	380	61.3	4.55	82	42.2
25%	224	48.3	3.33	63	28.4
50%	219	44.5	3.03	49	13.1
95%	147	46.6	2.73	46	18.1

Ratios

Cost sharing	Dental Care	Preventive Care	Office visits	Drugs	Mental Health
0	1.00	1.00	1.00	1.00	1.00
25%	0.59	0.79	0.73	0.77	0.67
50%	0.58	0.73	0.67	0.60	0.31
95%	0.39	0.76	0.60	0.56	0.43

Table 2.2 Demand responsiveness of various outpatient measures from the RAND health insurance experiment (Newhouse, 1993)

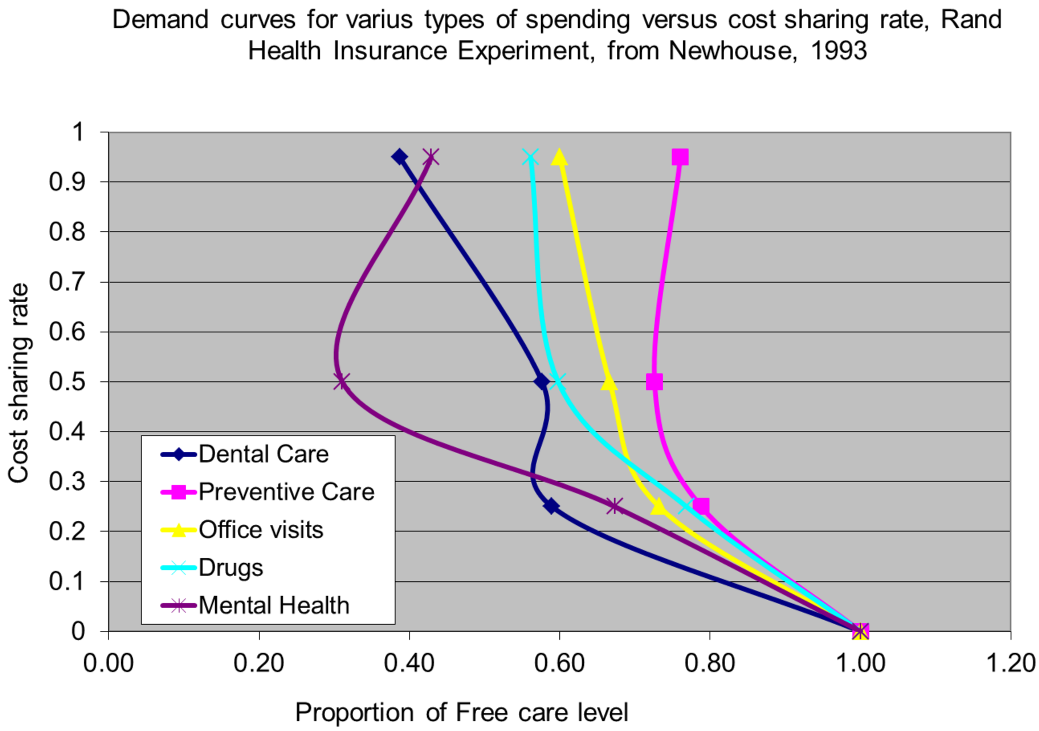


Figure 2.6 Demand Curves from Rand Health Insurance Experiment from Table 2.2

2.3 Optimizing versus non-optimizing decision rules.

A central concept to most of economics is the idea that agents optimize some objective function. Often, consumers are said to maximize utility, or doctors maximize health benefits to their patients, or health plans maximize profits. Frequently, the issues that make this optimization interest are the constraints operating on the decision-makers, which interact with the five main themes identified in Chapter 1. We will often use economic models of decision-making that assume agents are rational and make decisions that are optimal.

A growing literature suggests that although rational optimizing behavior is very convenient to assume, consumers often are not as rationale as the models assume them to be. Examples include that consumers tend to discount the future more heavily than conventional discount formulas suggest is optimal (FXXX = add cites), consumers place more weight on low probability events than they should, consumers are heavily influenced by framing and the context in which choices are made, and when confronted with complex choice options, consumer decisions are not always internally consistent. These apparently irrational behaviors are the focus of a whole subfield of economic research called behavioral economics. Books by Daniel Kahneman (2011) and Daniel Ariely (2010), give particularly good overviews of the issues involved. Because non-price transactions are often involved, behavioral economics plays an important role in health care markets. We review their key concepts and findings for health economics here and attempt through the book to integrate this alternative approach and contrast its implications with those from traditional economics.

What would the experimental literature have to say about the Rand Health Insurance Experiment results?

Arielly (2010) highlights the strong evidence that there is a discontinuity in demand behavior at a zero price, so that the “free care” health plan is fundamentally different from plans with positive cost shares. It could be that the strong increase for the free care plans reflects this feature, however since the HIE did not include any “low but not zero” copayment plans, this cannot be determined from that data.

Given the highly imperfect information that consumers have about the benefits of treatment, it may seem hard to justify using the demand curve as equivalent to the marginal benefit function.

Further notes

Most of the reduction in demand is explained by reductions in patients seeking care when in the higher copayment plans, not in changes in intensity of treatment conditional on making a visit. This is consistent with doctors largely determining the intensity of treatment when ill.

Further work in Newhouse (1993) and in O’Grady et al (19XX) about emergency room use suggests that demand side cost sharing reduces appropriate and inappropriate care about equally; suggesting that consumers (or their physicians) do not know which care is appropriate or inappropriate.

Cost sharing does seem to have a greater impact on less urgent types of care relative to more urgent care. This shows up in Figure 2.2 in which mental health, drugs, and dental care are more responsive than overall office visits.

Later, when we study preventive care, the fact that prevention is less responsive to cost sharing will make sense: greater insurance coverage may make spending on prevention less important than weak insurance coverage.

Need to present results showing variation in demand with other covariates.

Need to talk about placebo effects

Need to summarize results about income responsiveness and education.

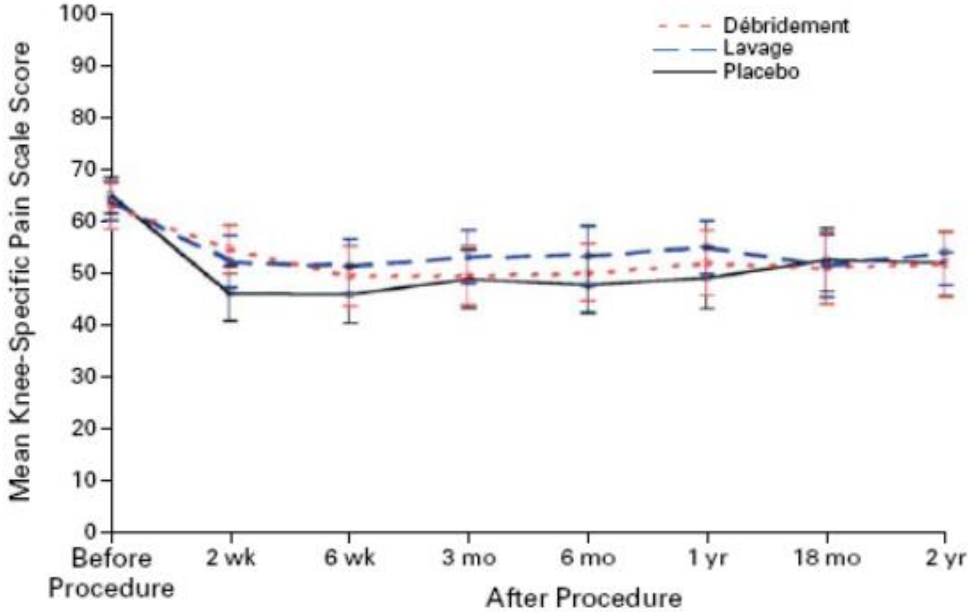
NEJM, July 11, 2002

Results from Moseley et al, "A controlled Trial of Ahrthroscopic Surgery for Ostoarthritis of the Knee

N=180 Mean Knee Specific Pain Scale 0 to 100 (higher is worse)

	12 months after surgery			**24 months after surgery**		
	Mean	Standard Dev	Approximate Standard error	Mean	Standard Dev.	Approximate Standard error
Placebo	48.9	21.9	2.83	51.6	23.7	3.06
Lavage	54.8	19.8	2.56	53.7	23.7	3.06
Debridememt	51.7	22.4	2.89	51.4	23.2	3.00

Tests	test statistic	p value	reported p value	test statistic	p value	reported p value
lavage vs placebo	1.55	0.12	0.14	0.49	0.63	0.64
debridement vs placebo	0.69	0.51	0.51	-0.05	0.96	0.96



No. at Risk								
Placebo	60	59	57	56	57	53	52	55
Lavage	61	59	57	59	59	57	56	55
Débridement	58	59	59	58	56	50	51	53

Figure 1. Mean Values (and 95 Percent Confidence Intervals) on the Knee-Specific Pain Scale.
Assessments were made before the procedure and 2 weeks, 6 weeks, 3 months, 6 months, 12 months, 18 months, and 24 months after the procedure. Higher scores indicate more severe pain.

2002: 650,000 procedures $5000 apiece = $3.2 billion/year