

HEALTH INSURANCE THEORY:
THE CASE OF THE VANISHING WELFARE GAIN

by

John Nyman

Discussion Paper No. 319, January 2003

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, MN 55455

**Health Insurance Theory:
The Case of the Vanishing Welfare Gain**

John A. Nyman*
University of Minnesota

January 7, 2003

Abstract

This paper presents theory that an important source of value is missing from conventional theory of the demand for health insurance, namely, the effect of the transfer of income (from those who purchase insurance and remain healthy to those who purchase insurance and become ill) on purchases of medical care. Because the portion of moral hazard that is attributable to income is welfare increasing and would replace some of moral hazard that is spuriously deemed to be welfare decreasing, the new theory suggests that the value of health insurance has been dramatically undervalued. Implications for policy are outlined.

*John A. Nyman
Division of Health Services Research and Policy
University of Minnesota
420 Delaware St. SE, Box 729
Minneapolis, MN 55455-0392
phone: 612.626.4425
fax: 612.624.2196
email: nyman001@umn.edu

Introduction

Consider Elizabeth. This year, Elizabeth becomes one of the 12 percent of women (lifetime) in the U.S. who is diagnosed with breast cancer (American Cancer Society, 2003). Without health insurance, Elizabeth would purchase a \$20,000 mastectomy, including chemotherapy, which is the care that she needs to rid her body of the cancer. Because a mastectomy is a disfiguring procedure, she would consider purchasing a breast reconstruction procedure for an additional \$20,000, but without insurance she would choose not to do so because, even though she could raise the funds privately, the competing claims on her income and resources make it too expensive. Fortunately, Elizabeth had purchased a contingent-claims insurance policy this year for \$3,000 that paid her a \$40,000 cashier's check when she was diagnosed with breast cancer, which is exactly enough to cover the cost of both procedures. With the extra ($\$40,000 - \$3,000 =$) \$37,000 in income, Elizabeth chooses to purchase the additional \$20,000 breast reconstruction. She is happy with her purchase because, while she could have spent the extra income on anything she wanted, she obtained the highest utility from the additional medical care.

The standard translation of this scenario into economic theory is that Elizabeth's demand for medical care has shifted out because of the additional income. In Figure 1, without insurance, the ill consumer's demand curve is D^u , and she consumes M^u of medical care because at M^u , her willingness to pay equals the price, P . With the additional income from insurance, Elizabeth's demand shifts out to D^i and she consumes M^i . Because the price of medical care is assumed to be constant and equal to the marginal cost of care, Elizabeth (and society) experiences a welfare

gain from the additional medical care, that is, from the moral hazard. The moral hazard welfare gain is equal to the consumer surplus from $(M^i - M^u)$: area ABC. In addition, because her income has increased by \$37,000 compared to without insurance, she is willing to pay more for the original mastectomy and therefore the value that she places on the original procedure increases as well. Area DEBC would represent the total welfare gain from the income transfer, including the gain from the additional medical care that she consumes because of insurance.

Now, consider what would have happened if the market had not offered contingent-claims health insurance contracts, but instead only offered insurance that paid off by covering the cost of any medical care that Elizabeth chose to purchase. That is, what if the checks were written to the providers of the mastectomy, chemotherapy, and reconstruction surgery, instead of to Elizabeth? Furthermore, to hold things constant and provide the starkest contrast, what if Elizabeth's premium remained \$3,000 and her voluntary behavior with this policy was the same as her behavior under the contingent-claims policy, and \$40,000 in checks were simply written to the providers for the exact same procedures?

The conventional translation of this payoff mechanism and behavior is that it represents a reduction to zero of the price of medical care that Elizabeth faces (Pauly, 1968).

Diagrammatically, as a result of this price decrease, Elizabeth again decides to purchase M^i on Figure 1, which this time is conventionally translated into a movement along Elizabeth's original uninsured demand curve, D^u . By this simple change in the payoff mechanism, the original welfare gain from insurance vanishes and is replaced by a welfare loss equal to BCM^i . This apparent loss is generated because the value of the additional care, area BM^iM^u , is assumed to be less than the cost of the additional care, area BCM^iM^u , by area BCM^i .

Yet, Elizabeth's behavior has not changed at all. She would voluntarily pay the same \$3,000 premium for both insurance contracts, she receives the same \$40,000 payoff when ill, and she consumes the same \$20,000 of additional care because of being insured, that is, she exhibits the same moral hazard. Most importantly, she is equally happy with her purchases under either insurance contract, that is, they generate the same level of utility. Yet, under the economic analysis of the contingent-claims insurance, society experiences a welfare gain because the income payoff increases her willingness to pay at all prices, but under conventional analysis of the price payoff mechanism, society experiences a welfare loss because the care she receives is not worth the cost of producing it. That so much value can be lost simply by writing the check to the provider, instead of the beneficiary, for the same services does not make sense, and implies a fundamental problem with conventional theory.

This paper analyzes the case of the vanishing welfare gain from health insurance. In the next section, the underlying indifference curve analysis is presented, showing that in actuality the welfare gain does not vanish. Instead, this paper shows that the conventional analysis is wrong, and at best, represents only a relatively unimportant special case. In the third section, the ex post consumption decision described in the previous section is incorporated into an expected utility model of the decision to purchase health insurance. In the fourth section, the analysis of the price reduction in price-payoff insurance is compared with the price reduction of a government subsidy. The paper concludes by addressing the theory's implications for public policy.

The Ex Post Consumption Decision

Basic model. The indifference curve diagram representing Elizabeth's behavior is illustrated in Figure 2. Units of medical care consumed are represented by M , and units of other goods and services consumed are represented by Y . For simplicity, assume that Y is the numeraire, denominated in dollars, and that the price of each unit of M is \$1. The consumer's uninsured budget constraint is represented by the curve labeled "P=1, no income transfer." Its intercept, Y^o , represents the consumer's endowed income. Assume that the insurance contract is for a specific disease, that every consumer has the same probability, p , of becoming ill with this disease during the contract period, and that all purchasers of insurance have the same preferences and endowments. The representative consumer's choice over Y and M is an ex post decision, that is, it is conditional on illness having occurred.

Mathematically, without health insurance, the ill consumer faces the problem,

$$\max U^s(M, Y) \quad (1)$$

$$\text{s.t. } Y^o = M + Y, \quad (2)$$

where U^s indicates utility that is dependent on being ill. The consumer solves this problem by consuming bundle (M^u, Y^u) , consistent with the first order conditions,

$$U^s_M / U^s_Y = -1 \text{ and} \quad (3)$$

$$Y^o = M + Y. \quad (4)$$

With price-payoff insurance that reduces the price from 1 to c , such that $0 < c < 1$, the consumer solves,

$$\max U^s(M, Y) \quad (5)$$

$$\text{s.t. } Y^o - R = cM + Y, \quad (6)$$

where R is the fair insurance premium, taken as a given. The consumer maximizes utility at (M^{ppi}, Y^{ppi}) consistent with the first order conditions,

$$U_M^s / U_Y^s = -c \text{ and} \quad (7)$$

$$Y^o - R = cM + Y. \quad (8)$$

Before setting R , the insurer conducts an actuarial study to determine demand for medical care, M^{ppi} , given R , c , and Y^o , and sets the fair premium at $R = p(1-c)M^{ppi}$. Substituting this back into equation (8) yields

$$Y^o - p(1-c)M^{ppi} = cM^{ppi} + Y^{ppi}. \quad (9)$$

Adding $(1-c)M^{ppi}$ to both sides of equation (9) and rearranging terms shows that the consumer's budget constraint with price-payoff insurance is

$$Y^o + (1-p)(1-c)M^{ppi} = M^{ppi} + Y^{ppi}, \quad (10)$$

where $(1-p)(1-c)M^{ppi}$ are transfers of income from those $(1-p)/p$ who purchase insurance but remain healthy.

Comparing equation (10) with equation (4) shows the effect of becoming insured on the income of the ill consumer. Without insurance, spending $(M^u + M^i)$ by the ill consumer is constrained by Y^o , her endowed income. With insurance, her spending $(M^{ppi} + Y^{ppi})$ is constrained by $Y^o + (1-p)(1-c)M^{ppi}$, that is, by endowed income plus the income transfers from the healthy. These transfers exist because not all consumers become ill during the contract period (that is, because $p < 1$). Equation (10) shows that if everyone became ill during the contract period (that is, if $p = 1$) and the insurer charged the fair premium, a reduction in the price of medical care to c would generate no income transfers at all because each consumer would bear

the cost of the insurer's payoff as part of her insurance premium. Thus, despite the price decrease, the consumer would be constrained to consume within her original budget constraint and endowed income.

If the consumer had purchased a contingent-claims contract that cost the exact same premium R , [that is, $R = p(1-c)M^{ppi}$ with R is taken as a given] as under the price-payoff policy, and paid off with the same expenditure, only this time as a lump sum income payoff, I [that is, $I = (1-c)M^{ppi}$], the consumer would face the problem

$$\max U^s(M, Y) \tag{11}$$

$$\text{s.t. } Y^o - R + I = M + Y. \tag{12}$$

and solve it at (M^{cci}, Y^{cci}) , consistent with first order conditions

$$U_M^s / U_Y^s = -1 \text{ and} \tag{13}$$

$$Y^o - R + I = M + Y. \tag{14}$$

Because the premiums and the payoffs are the same, the income transfers under the contingent-claims contract equal those under the price-payoff contract. That is, equation (14) can be written

$$Y^o - p(1-c)M^{ppi} + (1-c)M^{ppi} = M^{cci} + Y^{cci}, \text{ or} \tag{15}$$

$$Y^o + (1 - p)(1-c)M^{ppi} = M^{cci} + Y^{cci}, \tag{16}$$

and the left hand side, representing available income, is identical to the left hand side of equation (10). If the endowments, premiums, payoffs, and income transfers are the same for these two contracts, then whether M^{cci} equals M^{ppi} or not depends on preferences alone.

No substitutability case. How to represent consumer preferences for medical care, however, presents a challenge. In the introductory example, Elizabeth responds to the price reduction in the exact same way that she responds to the income transfer. Thus, her preferences

must reflect the fact that no substitution occurs at the margin (between medical care and other goods and services) when the price of medical care drops to zero. Economists typically represent a lack of substitutability by drawing angular indifference curves, reflecting the idea that while still nominally convex, the level of convexity does not result in a noticeable change in behavior as relative prices change. Such preferences appear in Figure 2 and show that for large changes in price, holding utility constant, the consumer would not change her behavior.

This appears to be a reasonable assumption for many illnesses, especially those illnesses that require major medical procedures like mastectomies, breast reconstructions, coronary bypass operations, organ transplants, and so on. Angular preferences represent the lumpiness of these procedures. That is, the patient in need of a coronary bypass procedure is willing to give up a very large amount of other goods and services in exchange for one medical procedure, but willing to give up very little for a second procedure. Moreover, the ill consumer does not know enough to evaluate each component of the procedure to determine whether it is worth the value of other goods and services forgone to obtain it. As a result, for many expensive medical procedures, little substitution would occur at the margin as prices changed.

Even though many, perhaps most, of the more expensive components of major medical procedures are not subject to discretion at the margin, there may be some components where the consumer is responsive to price. For example, the consumer who pays for care out of her income plus lump sum insurance payoff may not choose to purchase any extra days in the hospital, while the consumer who faces a zero price may purchase a longer stay. The physician, however, may act to limit this substitutability. That is, she may permit some additional hospital days, but would probably limit the stay of the patient who simply likes the comfort and lifestyle associated with

living in the hospital.

This preference-constraining function of the physician is not unusual. The physician generally acts to limit medical expenditures to those that can be justified medically, and thus to limit the substitutability of medical care for other goods and services. For example, the physician limits hospital admissions to those who are ill, so that people who travel cannot stay in a hospital instead of a hotel. Or, the physician limits the use of steroids to those who have legitimate medical needs, and does not permit, say, body-builders to obtain them at the low insurance prices. Thus, the physician acts to constrain the preferences of the consumer, essentially making them exhibit a lower degree of substitutability, that is, making the indifference curves more angular.

Most importantly, there is a self-limiting aspect to much of medical care. That is, most medical procedures are associated with pain or time costs, or risk of mortality or morbidity, and therefore, are not desired by those who are healthy. For example, a person with cancer who undergoes a standard course of chemotherapy would not want to undergo another one unnecessarily because of the pain and quality of life reductions associated with this form of medical care. If the additional chemotherapy is painful and has no medical value, the indifference curve for the consumption of additional chemotherapy would need to be upward sloping, indicating that the consumer would need to be paid in other goods and services to endure more of this treatment. Indifference curves that are u-shaped, perhaps approaching a v-shape, would describe the preferences for this medical care. A v-shaped curve would indicate that the consumer has a high willingness to pay for a certain effective level of medical care, but additional care that is medically unnecessary would be so painful that additional payments of

other goods and services would be required to keep the consumer indifferent. Indifference curves that approach a v-shape would also imply that the consumer is largely unresponsive to differences in relative prices.

In general, different consumers can have differing preferences regarding different types of medical care. For example, while some consumers may prefer the comfort of a few extra days in the hospital, others may find them tedious and require additional income payments in order keep utility constant. Thus, for each consumer and each illness, medical care has its own unique set of indifference curves, based on the degree of substitutability between medical care and other goods and services, filtered through and constrained at times by the physician, who functions mainly, but not entirely, as the knowledgeable agent of the consumer.¹

The indifference curves in Figure 2, therefore, implicitly assume that the degree of substitutability of medical care for other goods and services is negligible. If so, then it does not matter to the consumer whether the health insurance contracts pays off by writing a cashier's check to the consumer for a lump sum amount I , equal to $(1-c)M^{ppi}$, or whether the contract pays off by reducing the price of medical care to c , so that the insurer spends $(1-c)M^{ppi}$ worth of premiums in purchasing the same procedures on the consumer's behalf. In either case, the consumer receives an income transfer equal to the payoff minus the premium, or $(1-p)(1-c)M^{ppi}$, and in either case, the consumer's ex post consumption bundle is exactly the same. More importantly, the two contracts generate the same increase in utility, from U_u^s to U_s^s , so that the

¹It is well known that the agency function of the physician may also create a response to insurance on behalf of the interests of the patient, or on behalf of the interests of the physician. Suffice it to say that the behavior of the physician-patient decision unit can reflect any of the responses to the price reduction and income transfers discussed in this paper.

consumer is equally well-off.

Because this consumer lacks the desire (or ability) to substitute medical care for other goods and services at the margin, but is responsive to income transfers, the demand curve diagrams for these two contracts must show the same welfare gain. Figure 3a shows the response to the income transfer under the contingent claims contract: additional medical care is demanded at every price.² The consumer purchases M^u without insurance and M^{cci} with insurance, because of the transfer of $(I - R) = (1-p)(1-c)M^{ppi}$ worth of income to the ill consumer. The consumer experiences a welfare gain from moral hazard equal to ABC.

Similarly, Figure 3b shows the response to the income transfer contained in the price-payoff contract. Because the consumer must pay only cM^u for care that, if uninsured, would have cost her M^u , her residual income after consuming M^u of care has increased by $(1-c)M^u$ with this insurance.³ Therefore, her willingness to pay for the original M^u units of care has increased, as has her willingness to pay for the next $(M^{ppi} - M^u)$ units of medical care.⁴ Once the effect of the income transfers has been accounted for, the consumer's demand curve is drawn so that reductions in price or more precisely, the percentage decrease in the original price due to c , where c can only represent a price decrease do not elicit any substitution of medical care for other goods and services. Because M^{ppi} and M^{cci} are the same, moral hazard is also the same

²The demand curves, as illustrated in Figures 3a and 3b, are drawn as linear and represent a stylized response to income. That is, they are not drawn to be consistent with the high degree of convexity assumed in Figure 2.

³This assumes that p approaches 0, so that there is a negligible premium effect on consumption.

⁴This effect is better shown in the set of diagrams reflecting limited substitutability because the convexity of the indifference curves is clearer.

under both contracts, and the consumer experiences the same welfare gain from both, area ABC. That is, in both cases, the willingness to pay for the additional care, after the effect of the income transfer is accounted for, exceeds the cost of producing the additional care.

It should be noted that the consumer is also consuming the same amount of other goods and services, $Y^{cci} = Y^{ppi}$, with either insurance contract. With the contingent-claims contract, the consumer spends $M^{cci} = M^{ppi}$ on medical care with $Y^o + (1 - p)(1 - c)M^{ppi}$ in income, so that,

$$Y^{cci} = Y^o + (1 - p)(1 - c)M^{ppi} - M^{ppi}. \quad (17)$$

Similarly, with the price-payoff contract, the consumer spends cM^{ppi} of his endowed income, Y^o , after she has purchased insurance for a premium of $p(1 - c)M^{ppi}$, so has

$$Y^{ppi} = Y^o - p(1 - c)M^{ppi} - cM^{ppi}. \quad (18)$$

Simultaneously adding and subtracting $(1 - c)M^{ppi}$ from the right-hand-side of equation (18) yields

$$Y^{ppi} = Y^o + (1 - p)(1 - c)M^{ppi} - M^{ppi}, \quad (19)$$

which is identical to the right-hand-side of equation (17). Thus, the entire consumption bundle is the same under both insurance contracts, with these preferences.

Unaffordable care. One case that merits special attention is that of the consumer who becomes ill and does not have sufficient resources to purchase the current treatment of choice. For example, the current treatment of choice for many liver failure cases is a liver transplant, currently costing about \$300,000. A consumer without insurance and only \$100,000 in assets would need to either save or borrow the difference. Because of the urgency of the need for treatment, saving \$200,000 is not feasible, and few lending institutions would be willing to make an uncollateralized \$200,000 loan for this somewhat risky procedure. Thus, without insurance

(and without charity or government safety nets) the consumer would not be able to purchase the liver transplant and, perhaps, receive only palliative care for the short period before death. With insurance, the consumer can take advantage of the approximately 1/75,000 probability of receiving a liver transplant in a given year, by purchasing insurance coverage for an actuarially fair premium of $(\$300,000/75,000 =) \4 . Regardless of whether the insurer pays off by writing a check for \$300,000 to the beneficiary, or to the surgeon and hospital, there is still a transfer of $(\$300,000 - \$4 =) \$299,996$ in income from those who purchase this coverage and remain healthy to the 1 in 75,000 who becomes ill. With this additional income, the consumer purchases the same liver transplant procedure under either type of policy and experiences the same increase in utility.

This case is illustrated in Figure 4, which again assumes negligible substitutability, consistent with the stylized preferences for major procedures. Figure 4 shows the gain from moral hazard when the consumer faces a binding liquidity constraint in purchasing M^i , compared with purchasing M^u , the often palliative care that is affordable without insurance. In this case, $(M^i - M^u)$ again represents moral hazard and the demand analysis would correspond to Figures 3a and 3b, except that the consumer surplus would be larger.

This is an important special case because the health effects for such procedures are often very valuable because they save lives and reduce morbidity. Thus, for society, this case may have large impacts on welfare through its effect on health (see the studies of the effect of being insured on health, such as, Franks, Clancy and Gold, 1993; Currie and Gruber, 1996; Hanratty, 1996; and Lichtenberg, 2001). It is also important because this case represents a large portion of total health care spending. It is estimated that well over 30 percent of the health insurance

premium of the typical purchaser of insurance in the U.S. is devoted to paying for care that the purchaser would not otherwise be able to afford (Nyman, 1999a).

Limited substitutability case. Different diseases (or more generally, different health states) and different treatments may generate different preferences. For some diseases, the consumer may exhibit a limited level of substitutability between medical care and other goods and services. For example, in the introductory example of the breast cancer, Elizabeth may exhibit some willingness to trade off other goods and services for medical care. For example, with her price-payoff policy, she might spend an extra day or two in the hospital to recover, compared to her hospital stays under a contingent-claims policy.

Figure 5 illustrates this case. Without insurance, the consumer purchases (M^u, Y^u) . With insurance that pays off by reducing the proportion of the price paid from 1 to c , she purchases (M^{ppi}, Y^{ppi}) . Again the consumer is assumed to pay a fair premium equal to $p(1-c)M^{ppi}$, to receive a payoff equal to $(1-c)M^{ppi}$ spent on medical care by the insurer on her behalf, implying a net income transfer of $(1-p)(1-c)M^{ppi}$, financed by those who remain healthy. If the consumer had instead purchased a contingent-claims insurance contract for the same fair premium, payoff level, and income transfer, only in this case, the payoff were made as a lump sum income payment, the consumer would consume at (M^{cci}, Y^{cci}) . In both cases, the consumer is paying the same fair premium for the same income transfer from the insurer, thus holding constant the exchange that represents the insurance contract.

The welfare implications of moral hazard are clearer with these more realistic preferences. The income transfer from those who are healthy results in increased consumption of medical care from M^u to M^{cci} . This increase in medical care, along with the increased

consumption of other goods and services (from Y^u to Y^{cci}), increases utility from U_u^s to U_{cci}^s . Holding constant the premium, payoff and income transfer but paying off by reducing price to c , some additional medical care would be purchased, $(M^{ppi} - M^{cci})$. This is because, under a price-payoff contract, the consumer's preferences are such that they respond to differences in relative prices. These additional purchases of medical care actually cost society $(Y^{cci} - Y^{ppi})$ to produce, and as a result reduce the gain in utility from $(U_{cci}^s - U_u^s)$ to $(U_{ppi}^s - U_u^s)$ when the entire bundle is considered.⁵

Figure 6a shows the corresponding demand curve analysis for the contingent-claims case, indicating that the insurance payoff generates a welfare gain from moral hazard equal to area ABC. This reflects the fact that the consumer's willingness-to-pay has increased with the additional income. For example, returning to Figure 5, the willingness to pay for unit M^u of medical care without the income transfer from insurance is represented by the slope of the indifference curve at point a. With the income transfer from insurance, the willingness to pay for M^u has increased because the slope of the consumer's indifference curve at point b is greater, owing to the assumed (quasi-)convexity of preferences. Returning to price-quantity space (Figure 6a), this translates into a shifting out of the demand curve so that with insurance, willingness to pay exceeds the price by some margin at M^u . Indeed, with the additional income, willingness to pay exceeds the price for all medical care purchased until M^{cci} .

Figure 6b shows the corresponding demand curve diagram for price-payoff insurance, assuming that $c = 0$, compared with uninsured demand. The same increase in willingness to pay

⁵Note that both the income transfer gain and the price effect loss are expected, that is, they occur ex ante with a probability of p .

(as illustrated in Figure 5) occurs with price-payoff insurance between M^u and M^{cci} and represents the welfare gain from purchasing additional health care with the additional income that is transferred from the healthy.⁶ The moral hazard that is generated by the income transfers in the price-payoff contract is exactly equal to the moral hazard that would occur with a contingent claims contract that transferred the same amount of income. This portion of moral hazard generates the same welfare gain, equal to area ABC.

In addition, however, using a percentage, c , of the original price to payoff the contract generates additional moral hazard because of the consumer's willingness to substitute some medical care for other goods and services at the lower effective insurance price. The portion of moral hazard attributed to the pure price effect is $(M^{ppi} - M^{cci})$, and the welfare loss is the difference between the marginal cost of producing this additional care and the willingness to pay for this care where the ill consumer must pay for any price reductions herself (for a formal derivation of this portion of the demand function, see Nyman, 1999b, 2003; Nyman and Maude-Griffin, 2001). The cost of this additional care in Figure 6b is area $CFM^{ppi}M^{cci}$, the willingness to pay for this care is area $CM^{ppi}M^{cci}$, and the welfare loss is equal to area CFM^{ppi} . The net moral hazard welfare effect is the gain, area ABC, minus the loss, area CFM^{ppi} . As drawn, the gain exceeds the loss and this moral hazard is on net welfare increasing. To find the total utility gain from medical care with insurance, the gain in willingness to pay for the original uninsured M^u must also be added.

⁶See Mishan (1981, pp. 175-183), for a discussion of the demand curve measure of the Hicksian consumer surplus that is generated by an exogenous reduction in the market price. The income effect from the exogenous price reduction causes an increase in the marginal value of medical care for every level of medical care consumption, given convexity of preferences (or, in Mishan's framework, given concavity of the utility function).

Total substitutability. Some health states may be such that the entire moral hazard is due to the willingness to substitute medical care for other goods and services, and nothing is due to income transfers. Cosmetic surgery, drugs to enhance sexual performance, and designer prescription sunglasses may be representative of this case. This case is illustrated in Figure 7. The consumer purchases (M^u, Y^u) without insurance, and with price payoff insurance, purchases (M^{ppi}, Y^{ppi}) . With a contingent claims policy that contains the same income transfer, the consumer would consume at (M^{cci}, Y^{cci}) such that $M^{cci} = M^u$. All the moral hazard would be due to the willingness to substitute medical care for other goods and services at the margin.

Figure 8a shows the demand curve diagram for the contingent-claims case. Even though income has increased, the demand with insurance does not respond to it, and there is no welfare gain due to income transfers. Figure 8b shows the demand curve diagram for the price-payoff insurance compared with the Marshallian demand, representing the observed relationship between the exogenously determined market price and quantity. These demand curves are drawn to diverge, so that at $P=0$, the quantity of medical care demanded with an exogenous market price decrease, $M^{p=0}$, exceeds the amount demanded if the consumer were required to pay a premium out of endowed income to purchase a policy that reduced the price of medical care to 0 in the event of illness. That is, while the indifference curve analysis in Figure 7 shows no income effect at $P=1$, it is possible that at lower prices, income effects may occur. As c drops, the premium increases, and remaining income diminishes, so Figure 8b is drawn to show that at these lower price levels, an income effect may exist.⁷ If no income effect exists throughout the

⁷For example, while the indirect costs of medical care (that is, the transportation and time costs) may not enter the decision when M^u is purchased, they may have a greater influence on consumption as the price of medical care itself is reduced.

entire range, then the Marshallian demand and the insured demand would be the same. In either case, the entire moral hazard is associated with a welfare loss. This is the case that represents conventional theory.

Price-reductions with no income payoffs. While the total substitutability case assumes that no income transfer effect occurs because of the nature of preferences, a final case assumes that no income transfer effect occurs because there is no income transfer. That is, what if insurance were purchased for coverage of services, such as, annual checkups, annual screening, flu and other vaccinations, that were purchased annually by virtually everyone who is insured. If so, the probability of such expenditure would approach 1, and there would be no income transfer from insurance because insurer's cost is paid for entirely in each person's premium. Thus, the insurance payoff simply cancels the premium payment, and income remains constant.

Figure 9 shows the indifference curve analysis of this case, assuming preferences that reflect limited substitutability like those in Figure 5. With contingent-claims insurance, no shifting of the budget constraint occurs in the contract period because the income payoff equals the insurance premium. With price-payoff insurance, the consumer is constrained to simply purchase a lower price for medical care. The new optimum would occur at the intersection of the original budget constraint (reflecting no income transfers) and the consumer's income expansion path for $P=c$. Thus, insurance would either leave the consumer on her original indifference curve, or on a lower one, unambiguously reducing welfare. Figure 10a shows that with contingent-claims insurance, the demand for medical care would stay constant without an income transfer. Figure 10b shows that if the consumer alone must pay for the additional care received

in response to the price change, price payoff insurance could only represent a welfare loss.⁸

Figure 10b also illustrates the response to price-payoff insurance in comparison to the observed Marshallian demand to an exogenous market price reduction. This comparison shows that, if the consumer must pay the entire cost of the price reduction, the consumer's behavior would exhibit a smaller moral hazard effect than conventional theory would lead one to believe, and result in a smaller welfare loss.

This case suggests that without an income transfer, health insurance would either do no good or would harm society. This is compelling evidence that the income transfer is central to understanding the demand for health insurance. It also suggests that the *level* of income transfers is important in understanding the demand for health insurance. That is, conventional theory often assumed that the effect of the premium on demand was small enough to be negligible, that we could ignore the effect of the premium on demand, and that the Marshallian demand curve would reflect behavior if insured (Pauly, 1968). This analysis shows that the premium is only one side of the relationship. The more important side is the income transfer that must be explicitly accounted for in determining the welfare implications of the demand response. If this income transfer does not exist, then there is no gain from purchasing insurance.

Summary. Conventional insurance theory suggests that all of moral hazard generates a welfare loss, and that the magnitude of this loss can be represented by Marshallian demand (Pauly, 1968). The theory presented here suggests that this is a very special case of a more general response. Most moral hazard either generates a welfare gain in its entirety or a combination of a gain and loss, and that the magnitude of most of these losses is smaller than the

⁸A tax subsidy would complicate this analysis considerably.

loss that would be represented by a Marshallian demand response. Thus, conventional theory has dramatically overstated the loss and understated that gain from health insurance.

The Demand for Health Insurance

The conventional theory of the demand for health insurance is based on a model that does not allow income increases to generate additional medical spending. The conventional model suggests that medical care spending is unaffected by any income payoffs and that the consumer simply purchases a payoff that covers the spending on medical care that would occur without insurance (Friedman and Savage, 1948). There is no recognition that the payoff net of the premium represents an increase in income, and that the consumer who becomes ill is likely to purchase additional medical care with this additional income. As a result, moral hazard is conventionally only attributable to a price effect and any additional consumption that is due to becoming insured can only be associated with a welfare loss.

In contrast, the new theory (see Nyman, 2003, for the complete theory) suggests that the missing income effect on medical care spending is central to understanding the demand for health insurance. Ex ante, the consumer considers expected utility without insurance,

$$EU_u = (1-p)U^h(0, Y^o) + pU^s(M^u, Y^o - M^u), \quad (20)$$

and expected utility with price-payoff insurance,

$$EU_i = (1-p)U^h[0, Y^o - p(1-c)M^{ppi}] + pU^s[M^{ppi}, Y^o + (1-p)(1-c)M^{ppi} - M^{ppi}]. \quad (21)$$

U^h is utility when healthy, where M is assumed to be 0 and where consumption of other goods

and services is equal to endowed income after the premium is subtracted. Note that although equation (21) shows the components $[p(1-c)M^{ppi}]$ of the actuarially fair premium, the actual premium (R) is taken as a given by the consumer. It is only because of the actuarial study by the insurer that we know that $R = p(1-c)M^{ppi}$. Insurance is purchased if $(EU_i - EU_u) > 0$, or if

$$EU_i - EU_u = (1-p)U^h[0, Y^o - p(1-c)M^{ppi}] + pU^s[M^{ppi}, Y^o + (1-p)(1-c)M^{ppi} - M^{ppi}] - (1-p)U^h(0, Y^o) - pU^s(M^u, Y^o - M^u) \quad (22)$$

$$= (1-p)\{U^h[0, Y^o - p(1-c)M^{ppi}] - U^h(0, Y^o)\} + p\{U^s[M^{ppi}, Y^o + (1-p)(1-c)M^{ppi} - M^{ppi}] - U^s(M^u, Y^o - M^u)\} > 0. \quad (23)$$

Thus, in considering the purchase of insurance, the consumer weighs the effect of paying the premium when healthy on consumption of other goods and services, against the effect of the income transfer when ill on consumption of both medical care and other goods and services.⁹ Some of the moral hazard, however, may be generated by a price effect and this moral hazard causes a reduction of utility, compared with a fair contingent claims contract resulting in the same income transfer. The welfare gain from the additional medical care that is generated by the income transfers, however, is likely to far outweigh this welfare loss (see the estimate in Nyman, 2003). Because of the important welfare benefit associated with gaining access to those expensive and effective medical services that would otherwise be beyond the consumers liquidity constraint, the net welfare effect of moral hazard is likely to be positive. Indeed, value of health insurance under the new theory is dominated by the one component that is totally missing from conventional theory: the welfare gain from consumption of additional medical care that is

⁹Note that the expected premium, $-(1-p)p(1-c)M^{ppi}$, equals the expected income transfer, $p(1-p)(1-c)M^{ppi}$ and, as a result, the insurance premium is fair.

generated by the income transfer.

Health Insurance and the Theory of Subsidies

The conventional theory of the moral hazard welfare loss is derived from the theory of excise subsidies. A government price subsidy results in a welfare loss to the extent that it causes consumption to increase, but the value of the additional goods and services are less than the cost of producing them. The conventional analysis of a subsidy is that it is simply a negative tax, and as such, represents a price decrease that contains the same Hicksian price and income effects that would characterize the decomposition of an exogenous price decrease in the market. Thus, the conventional analysis of a government subsidy would eliminate the income effect before determining the welfare loss from the compensated demand alone. The income effect would simply be missing from the analysis.

The income effect is missing from the analysis because typically no further assumptions are made regarding the income. For a tax analysis, the income is implicitly retained by the government and represents potential, but not actual, spending on projects that would benefit citizens. Similarly, for the subsidy, the income used to finance the subsidy comes from the government, with no assumptions about which taxpayers actually contributed this income. Therefore, it is not clear whether the beneficiaries of the subsidy financed their own price reduction, or whether it was financed by taxpayers who do not consume the commodity that is being subsidized. Indeed, for most consumer goods for which a subsidy has historically been applied food, housing, clothing it is difficult to limit consumption of the commodity to only one

segment of the population.¹⁰ Therefore, the income effect is simply missing from the analysis of subsidies, and the analyst conventionally focuses on the substitution effect and the welfare loss that it causes.

With additional assumptions about the financing of the subsidy, the analyst could perform a more complete analysis. For example, if the price reduction were intended to transfer income from rich to the poor, and if it were known that only the rich paid taxes and only the poor would take advantage of the lower price of the subsidized commodity, then the welfare effect of the subsidy would be determined by how the poor behaved with the price subsidy compared with a direct income transfer of income equal to the tax spending from the subsidy.

For example, if an excise subsidy were to reduce the price of bread from \$2 a loaf, the true marginal cost of producing bread, to \$1 a loaf, and the poor consumer purchased 3 loaves a week without the subsidy but 6 loaves with the subsidy, the conventional analysis, assuming linear demand curves, would hold that a welfare loss of $[\$6 - (\$3 + 0.5 \cdot 3 \cdot \$1)] = \1.50 occurred. However, if the price subsidy is the mechanism for making a \$6 income transfer, then to determine its true welfare effect, we would first need to determine how much bread the poor consumer would have purchased if he were transferred $(6 \cdot \$1 =)$ \$6 directly in income, the cost of the subsidy, as a lump sum every week. If he would have purchased the same 3 additional loaves at the original \$2 price, it would mean that the poor consumer benefitted more by spending those additional \$6 on bread than anything else. The standard analysis would shift demand out, and

¹⁰Vouchers are one way to limit the subsidy to certain consumers, but vouchers are not the same as subsidies, which lower prices to all.

Food and clothing is subsidized in many states by an absence of a sales tax. Housing is subsidized in a variety of ways.

society would experience a welfare gain associated with these 3 additional loaves, rather than a welfare loss. The net effect of the subsidy would then weigh the gain of utility from the additional income received by the poor against the loss of utility from the taxes paid by the rich. The consumer surplus for the additional bread purchased with the subsidy income would enter the analysis as a benefit, rather than a welfare loss. If an external benefit from taxpayer altruism existed, the subsidy would generate an even greater societal surplus.

Such behavior, however, is unlikely. First, it is unlikely that the benefits of an excise subsidy will be conferred on only one segment of the population. To do so, it would be necessary to subsidize a commodity that only the poor would consume and to finance the subsidy by taxing only the rich. However, everyone consumes some level of food, shelter and clothing, and would all benefit from a general subsidy of these commodities. Moreover, not all taxes are paid by the rich. Thus, one reason why income effects are missing from a subsidy is that it is not clear who finances the subsidy and to whom it would go.

In contrast, with health insurance, because medical care is primarily consumed by those who have fallen ill, prices of medical care can fall and most of the added consumption of medical care will be consumed by the ill, the intended beneficiaries of insurance. Moreover, it is perfectly clear who finances these price reductions: those who purchase insurance, but remain healthy. For private insurance, this is a voluntary arrangement where this transfer is understood quid pro quo at the point at which insurance is purchased.

Subsidies that represent only income effects are also unlikely because, if the same tax expenditures that financed the excise subsidy were simply given to the poor as a lump sum transfer, the poor would likely spend it on things other than more of the subsidized commodity.

These alternative commodities may be unworthy (in the minds of taxpayers) of the taxpayers' income transfers. Thus, another reason why income effects are missing from the analysis of a subsidy is that it is not clear what the beneficiaries would consume with an income transfer of the same amount.

In contrast, with medical care, there is the reasonable presumption that if an income payoff were received by someone who had just been diagnosed with a illness, a large portion of that payoff would be used to purchase medical care. This is especially true for illnesses that are serious with regard to their morbidity and mortality implications, but for which the standard treatments would otherwise be unaffordable.

It should be noted that this analysis suggests that consumers are better off with contingent claims insurance than price payoff insurance, *ceteris paribus*. Contingent claims insurance is rarely sold on the market, however, because it is difficult to observe illness. As a result, it is costly to monitor for fraud and even more so to write complex legal contracts that would accurately specify the conditions under which different payoffs would be made. Because of these considerations, most health insurance contracts pay off by reducing price.

Conclusions

This paper has argued that a large source of value is completely missing from the conventional theory of the demand for health insurance. Conventional theory has focused on the effect of the income payoff on increasing purchases of other goods and services, and on the effect of price on purchases of medical care. The effect of the income payoff on increasing purchases

of medical care is real, but has completely vanished from conventional theory.

This is an important omission and one that changes the value of insurance dramatically. Without recognition of this income effect, the voluntary purchase of health insurance at traditional coverage parameters appears to reduce welfare (Feldstein, 1973; Feldman and Dowd, 1991; Manning and Marquis, 1996) because all the additional medical care purchased is assumed to be in response to the willingness to substitute medical care for other goods and services as relative prices change. As such, it could only mean that the additional care purchased with insurance is not worth the cost of producing it. With recognition of the income effect, however, not only is much of the welfare loss under conventional theory eliminated, but furthermore, it is *replaced* by a welfare gain. This welfare gain is due to the fact that with the additional income from insurance, the consumer would be willing to pay an amount that exceeds the cost of much of moral hazard. This gain is very important because it is often derived from those major medical procedures that would otherwise be beyond the consumer's liquidity constraint. There is evidence that, when the spurious losses are eliminated and true gains are accounted for, (1) the value of moral hazard far exceeds the costs of producing it, (2) the gain from moral hazard represents an important reason for purchasing health insurance, and (3) the voluntary purchase of health insurance generally increases society's welfare (Nyman, 2003).

Recognition of the income effect also dramatically changes the focus of public policy. Without this income effect, public policy has been directed at reducing consumption of medical care at the margin in order to reduce costs. Over the last 30 years or so, economists have promoted cost-sharing and managed care as efficient cost-containment policies. With recognition of an income effect, policy becomes more complicated because it is necessary use

cost-containment policies that distinguish efficient from inefficient moral hazard. Copayments and managed care should only be directed at the inefficient portion of moral hazard that is generated by a pure price effect. Moreover, policies to reduce monopoly pricing of medical care, once thought of as being counterproductive because high prices counteracted some of the moral hazard welfare loss of insurance (Pauly, 1995), must now be considered legitimate ways to limit medical care spending (see Nyman, 2003).

The ideal insurance policy is also different with the new theory. This theory suggests that once the consumer becomes ill, it makes little sense to impose copayments to limit consumption. Office visits for chronic diseases (such as, diabetes, asthma, chronic obstructive pulmonary disease, and others) and expensive procedures that are deemed to be standard treatments for common ailments (such as, hip replacements, treatment of trauma, and others) should have “first dollar” coverage. If cost sharing is imposed, it should only be with regard to any care that responds to relative prices, such as, days in the hospital in excess of the standard stay. This theory also suggests the current “managed care backlash” is due to managed care’s denying coverage for procedures that the consumer would gladly be willing to pay for with the income transfer from insurance. It may also explain (at least in part) why so many elderly in the U.S. (about 90 percent) have supplemental insurance designed to reduce Medicare cost-sharing to zero.

Finally, the new theory provides a solid theoretical justification for insuring the uninsured and for implementing national health insurance. Under the conventional theory, the sole source of value of health insurance is derived from satisfying the consumer’s preference for certain (as opposed to uncertain) financial losses. Under the new theory, the benefits of insurance are

largely derived from the access that it provides to care that would otherwise be beyond the ill consumer's liquidity constraint. Providing access to care for those who are ill would be a source of external benefit for many in society and would justify either tax subsidies of health insurance, or the implementation of a national health insurance scheme.

To summarize, conventional theory analyzes the consequences of health insurance on other goods and services separately from its consequences on medical care. The gains from health insurance are determined by the effect on utility of spending on other goods and services only, and are seemingly derived (because of the specification conventionally used to model the insurance decision) from exchanging the payment of a small amount of income (that is, other goods and services) with certainty (that is, the premium) for a payment of a larger amount of income with uncertainty (that is, the income reduction from medical expenditures if ill). Because certainty could not be achieved if the income payoff resulted in the purchase of additional medical care, income effects on medical care are totally excluded from the analysis.¹¹ The implication is that any additional purchases of health care can only be due to a price (substitution) effect, but this implies that health insurance can generate only welfare losses from its impact on medical care consumption. Thus, an important welfare-increasing component of health insurance is completely excluded from conventional theory: the impact of income transfers on medical care purchases. This missing income transfer effect may represent the most

¹¹The exception is de Meza's (1981) model, but he addresses the choice to purchase health insurance compared with saving and borrowing the additional income. Although this is a very important article, it does not compare the purchase of insurance to remaining "self-insured," nor does it recognize the difficulty with saving and borrowing for expensive medical care. Finally, because it assumes contingent-claims rather than price-payoff insurance, it does not recognize that within a price payoff, there is both an income effect and a pure price (substitution) effect, and that they have different implications for welfare.

important factor in explaining the demand for health insurance.

References

- American Cancer Society website, 2003. "What Are the Key Statistics for Breast Cancer?"
www.cancer.org/docroot/CRI_2_4_1X_What_are_the_key_statistics_for_breast_cancer.
- Chernew, Michael E., William E. Encinosa, and Richard A. Hirth. "Optimal Health Insurance: The Case of Observable, Severe Illness," Journal of Health Economics vol. 19, no. 5, 2001, pp. 585-610.
- Currie, Janet and Jonathan Gruber. "Health Insurance Eligibility, Utilization of Medical Care and Child Health." Quarterly Journal of Economics, vol. CXI, 1996, pp. 431-466.
- de Meza, David. "Health Insurance and the Demand for Medical Care." Journal of Health Economics, vol. 2, no. 1, 1983, pp. 47-54.
- Feldman, Roger, and Bryan Dowd. "A New Estimate of the Welfare Loss of Excess Health Insurance." American Economic Review, vol. 81, no. 1, 1991, pp. 297-301.
- Feldstein, Martin S. "The Welfare Loss of Excess Health Insurance," Journal of Political Economy, vol. 81, 1972, pp. 251-80.
- Franks, Peter, Carolyn M. Clancy, and Marsha R. Gold. "Health Insurance and Mortality:

Evidence from a National Cohort.” Journal of the American Medical Association, vol. 279, 1993, pp. 737-741.

Friedman, Milton, and L. J. Savage. “The Utility Analysis of Choices Involving Risk.” Journal of Political Economy, vol. 66, no. 4, 1948, pp. 279-304.

Hanratty, Maria. “Canadian National Health Insurance and Infant Health.” American Economic Review vol. 86, no. 1, 1996, pp. 276-284.

Lichtenberg, Frank. The Effects of Medicare on Health Care Utilization and Outcomes.

Prepared for presentation at the Frontiers in Health Policy Research Conference, National Bureau of Economic Research. Washington, DC, 7 June 2001. Unpublished manuscript.

Manning, Willard G., and M. Susan Marquis. “Health Insurance: The Tradeoff Between Risk Pooling and Moral Hazard.” Journal of Health Economics vol 15, no. 5, 1996, pp. 609-40.

Mishan, E. J. Introduction to Normative Economics. New York: Oxford, 1981.

Nyman, John A. “The Value of Health Insurance: The Access Motive.” Journal of Health Economics vol. 18, no. 2, 1999a, pp. 141-52.

Nyman, John A. "The Welfare Economics of Insurance Contracts that Pay Off by Reducing Price." Discussion Paper No. 308. Center for Economic Research, Department of Economics, Minneapolis: University of Minnesota, 1999b.

Nyman, John A. "The Economics of Moral Hazard Revisited." Journal of Health Economics vol. 18, no. 6, 1999c, pp. 811-24.

Nyman, John A. "The Income Transfer Effect, the Access Value of Insurance and the Rand Health Insurance Experiment." Journal of Health Economics, 20, No. 2 (2001a), 295-98.

Nyman, John A. "The Theory of the Demand for Health Insurance," Discussion Paper No. 311, March, Center for Economic Research, Department of Economics, University of Minnesota, (2001b).

Nyman, John A. "The Demand for Insurance: Expected Utility Theory from a Gain Perspective." Unpublished manuscript; Minneapolis: Division of Health Services Research and Policy, University of Minnesota, (2001c).

Nyman, John A. The Theory of Demand for Health Insurance. Stanford, CA: Stanford University Press, 2003.

Nyman, John A. and Roland Maude-Griffin. "The Welfare Economics of Moral Hazard."

International Journal of Health Care Finance and Economics, 1, No. 1 (2001), 23-42.

Pauly, Mark V. "The Economics of Moral Hazard: Comment," American Economic Review, 58, No. 3 (1968), 531-37.

Pauly, Mark V. "More on Moral Hazard." Journal of Health Economics, vol 2, no. 1, (1981), 81-86.

Pauly, Mark V. "When Does Curbing Health Care Costs Really Help the Economy?" Health Affairs vol. 14, no. 2, 1995, pp. 68-82.

Zeckhauser, Richard. "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives." Journal of Economic Theory, 2 (1970), 10-26.

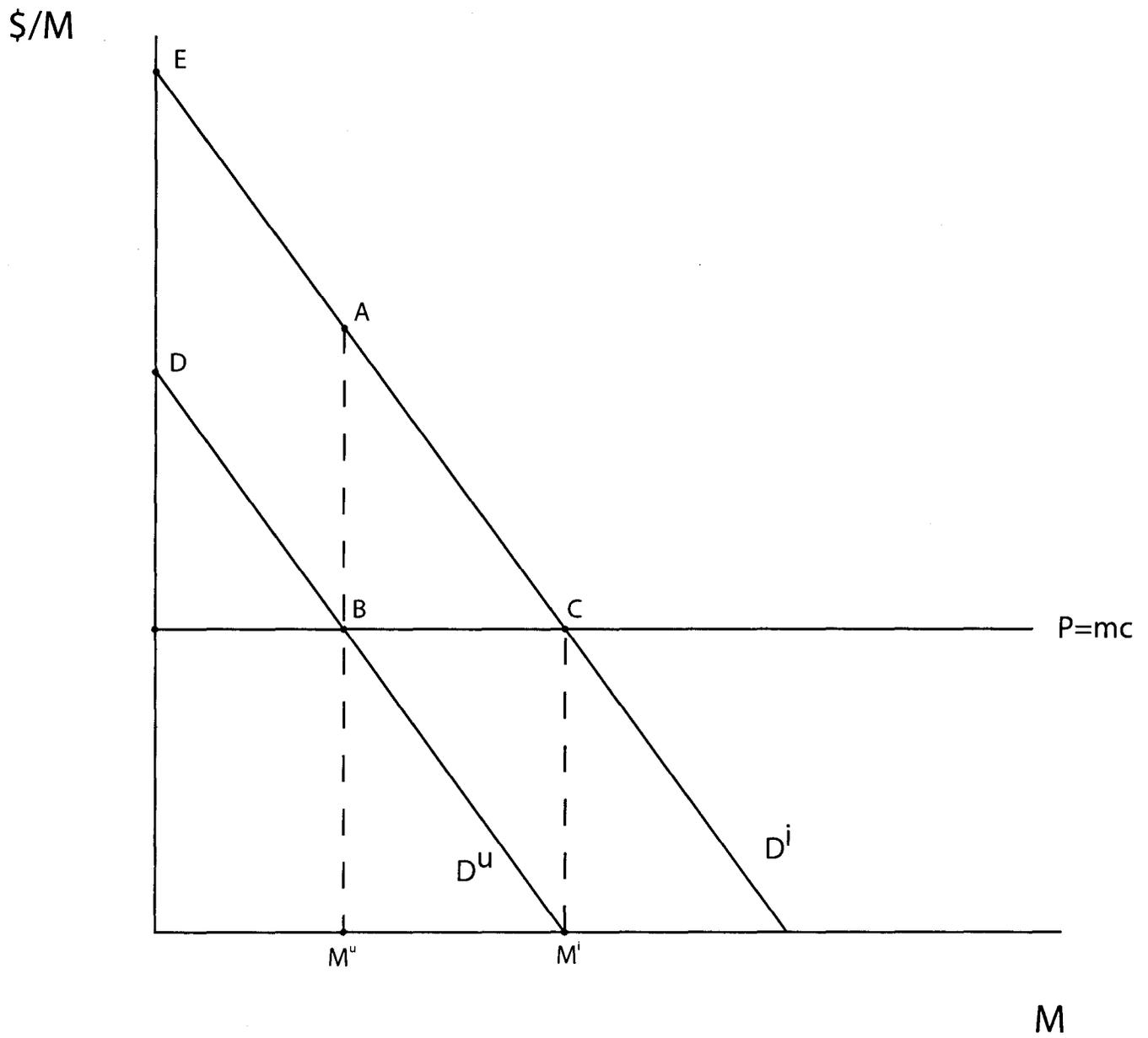


Figure 1
Demand Analysis of Elizabeth's Decision

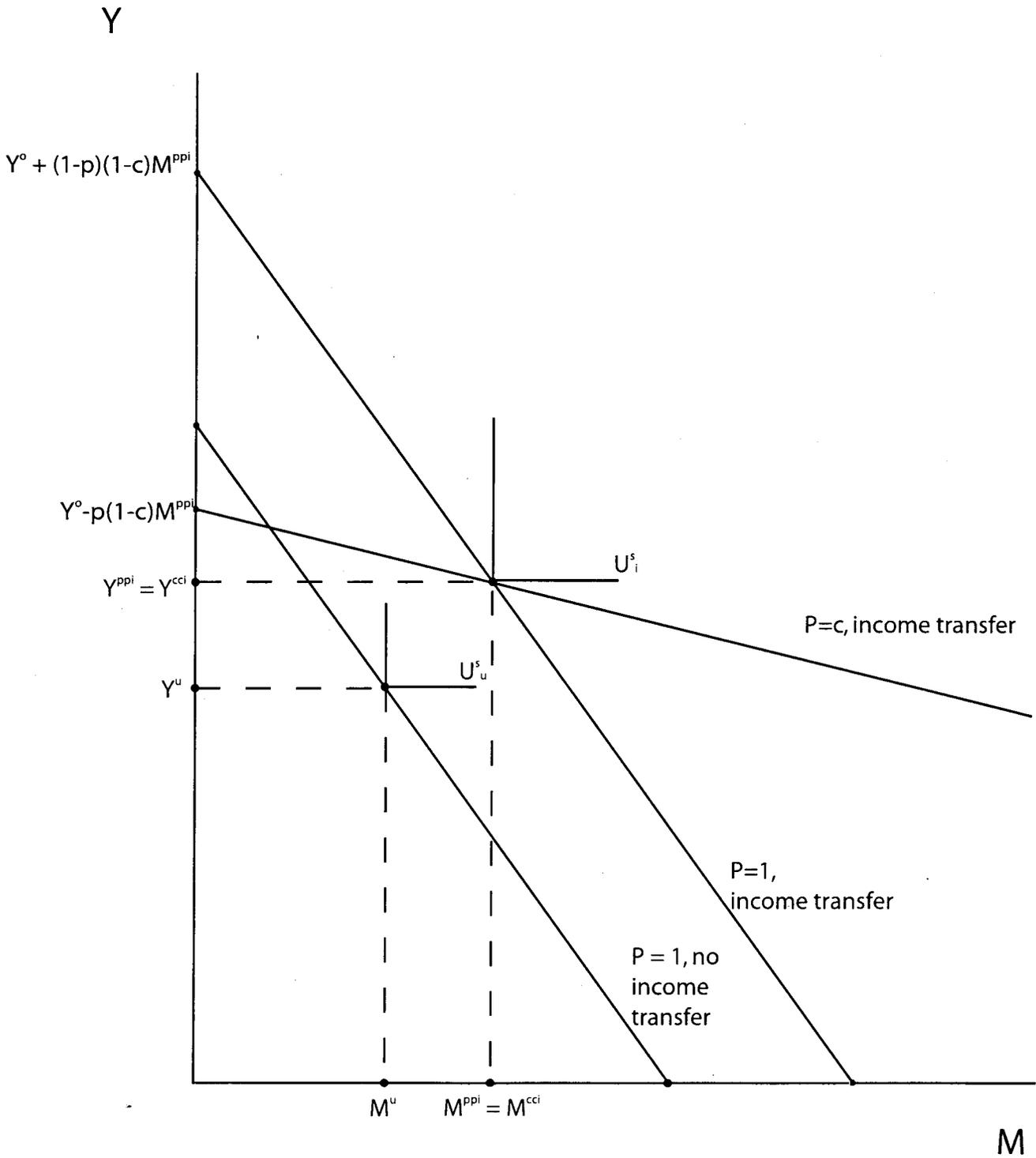
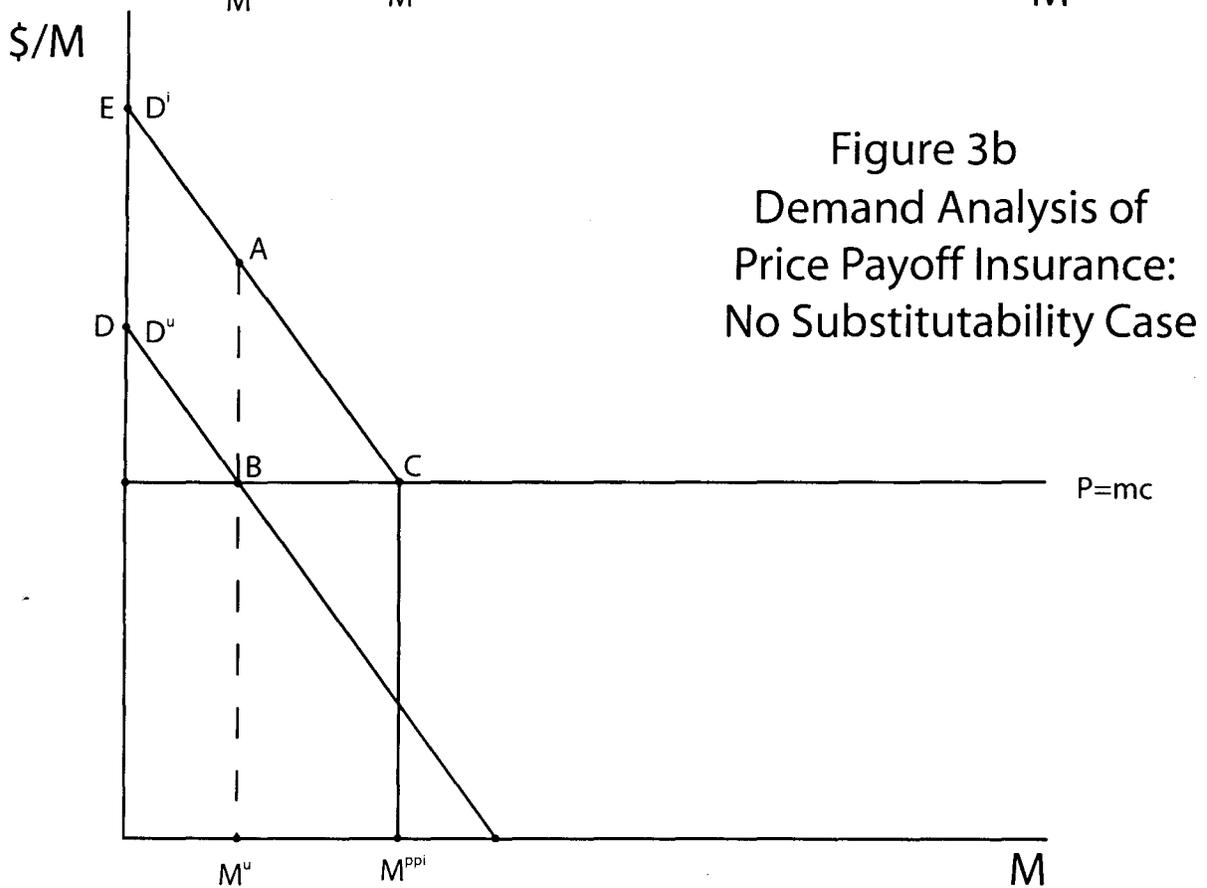
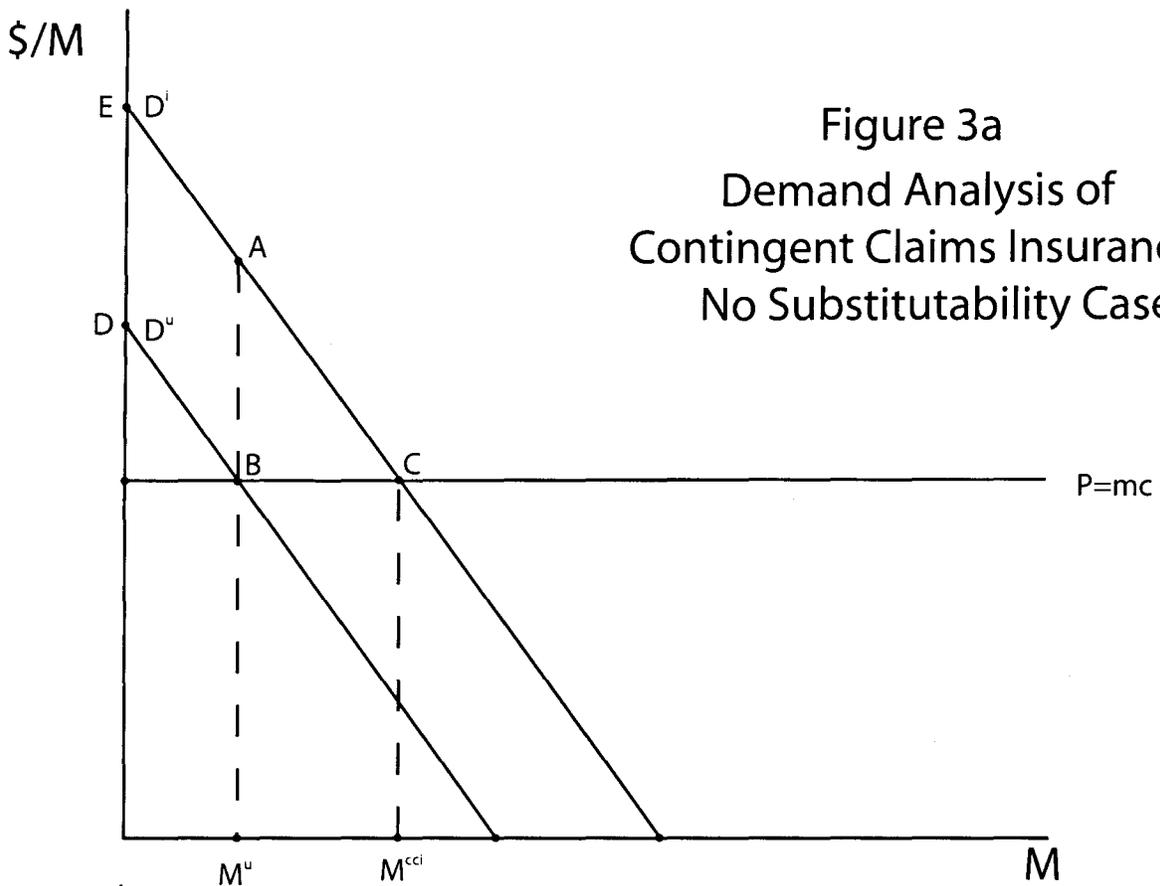


Figure 2

The Consumer's Decision Under Contingent-Claims and Price Payoff Insurance: No Substitutability Case



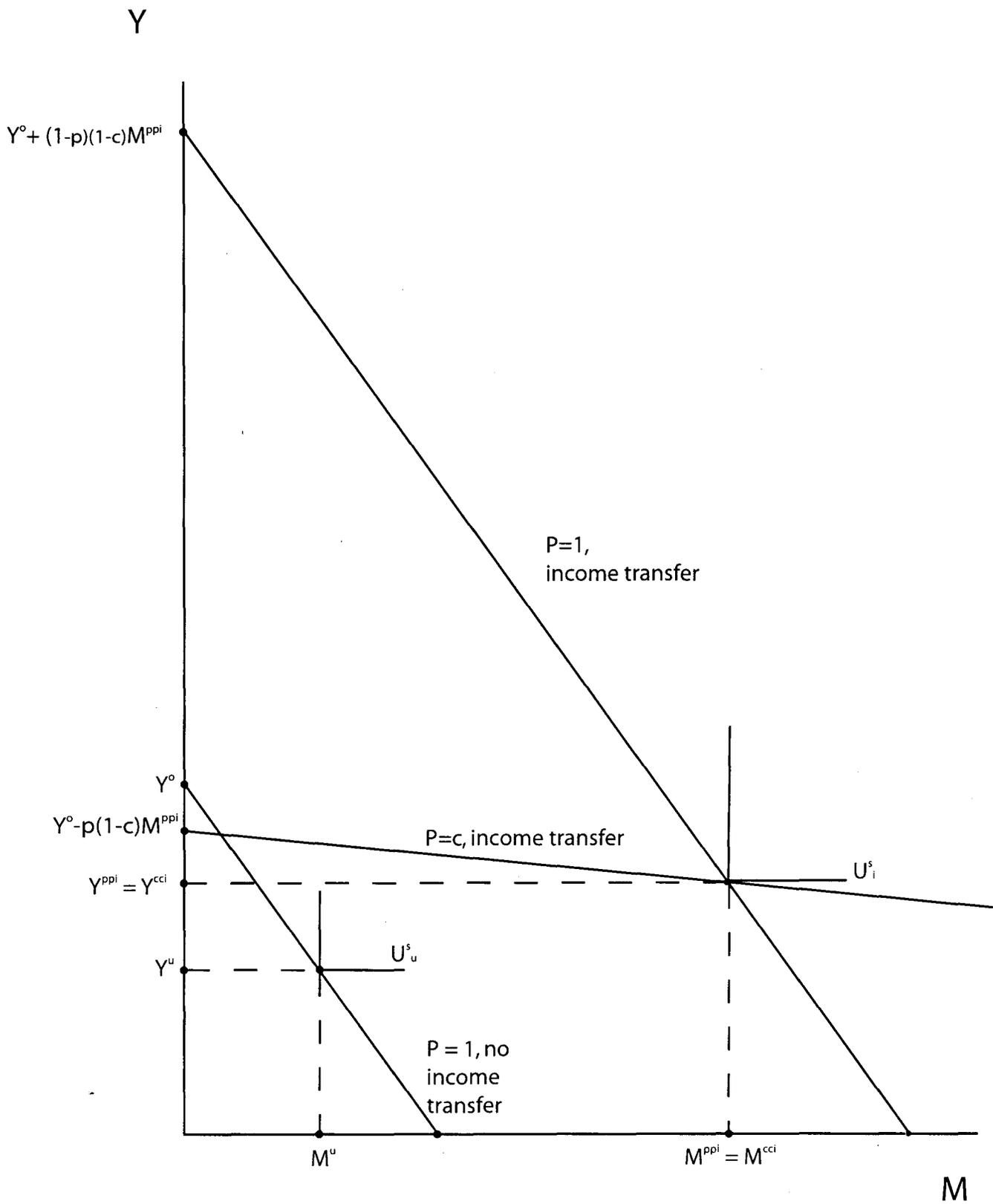


Figure 4

Insurance for Otherwise Unaffordable Medical Care

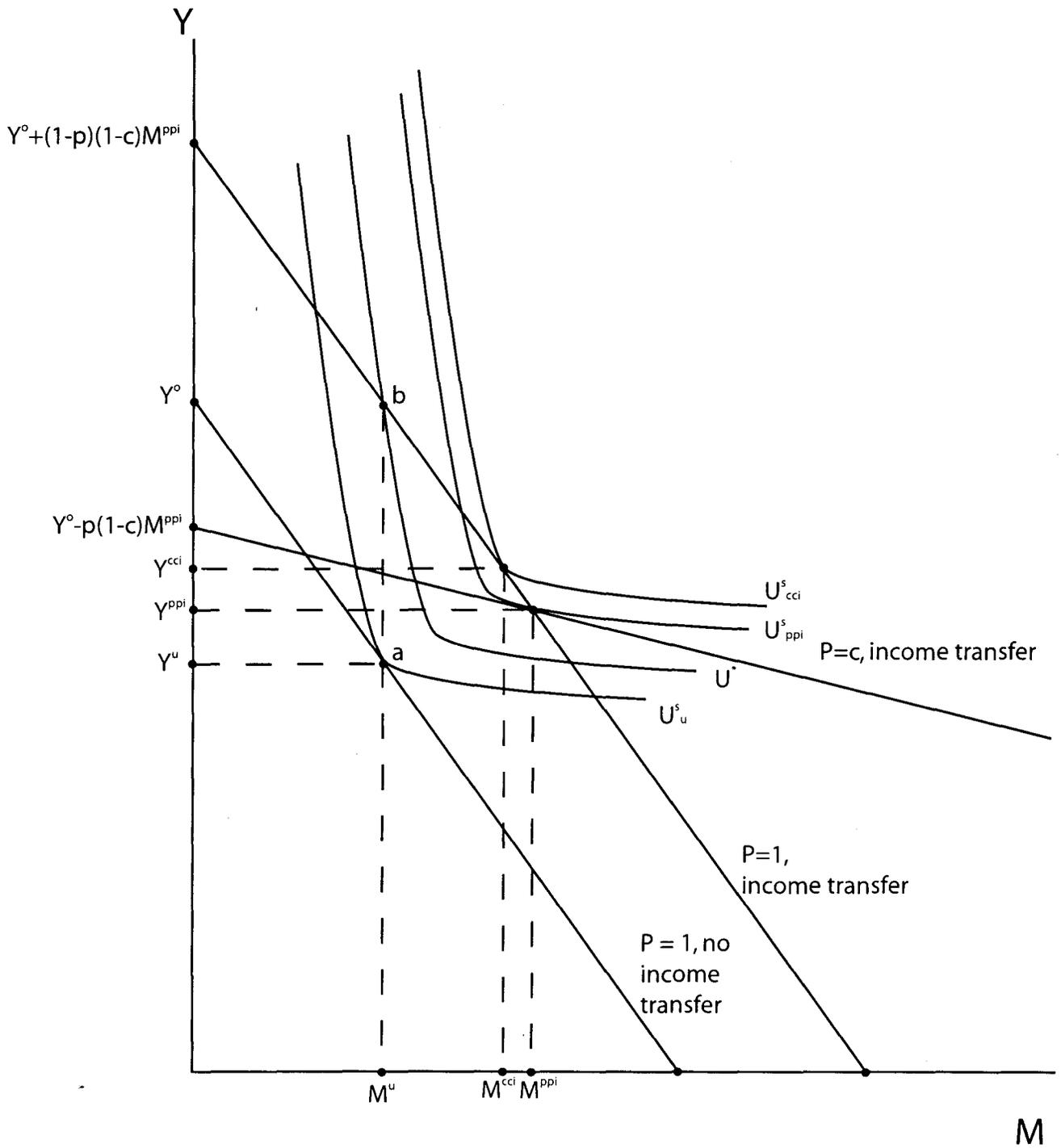
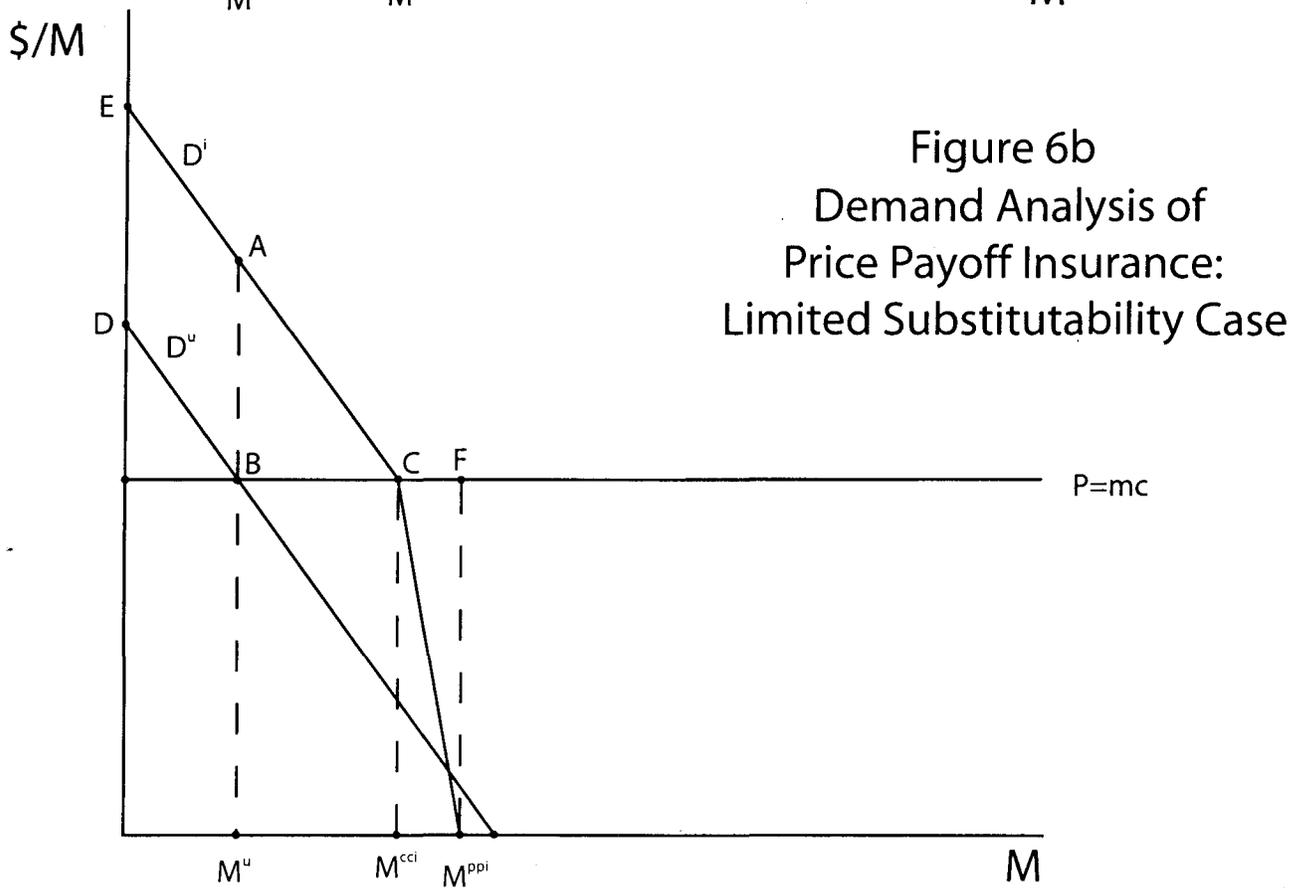
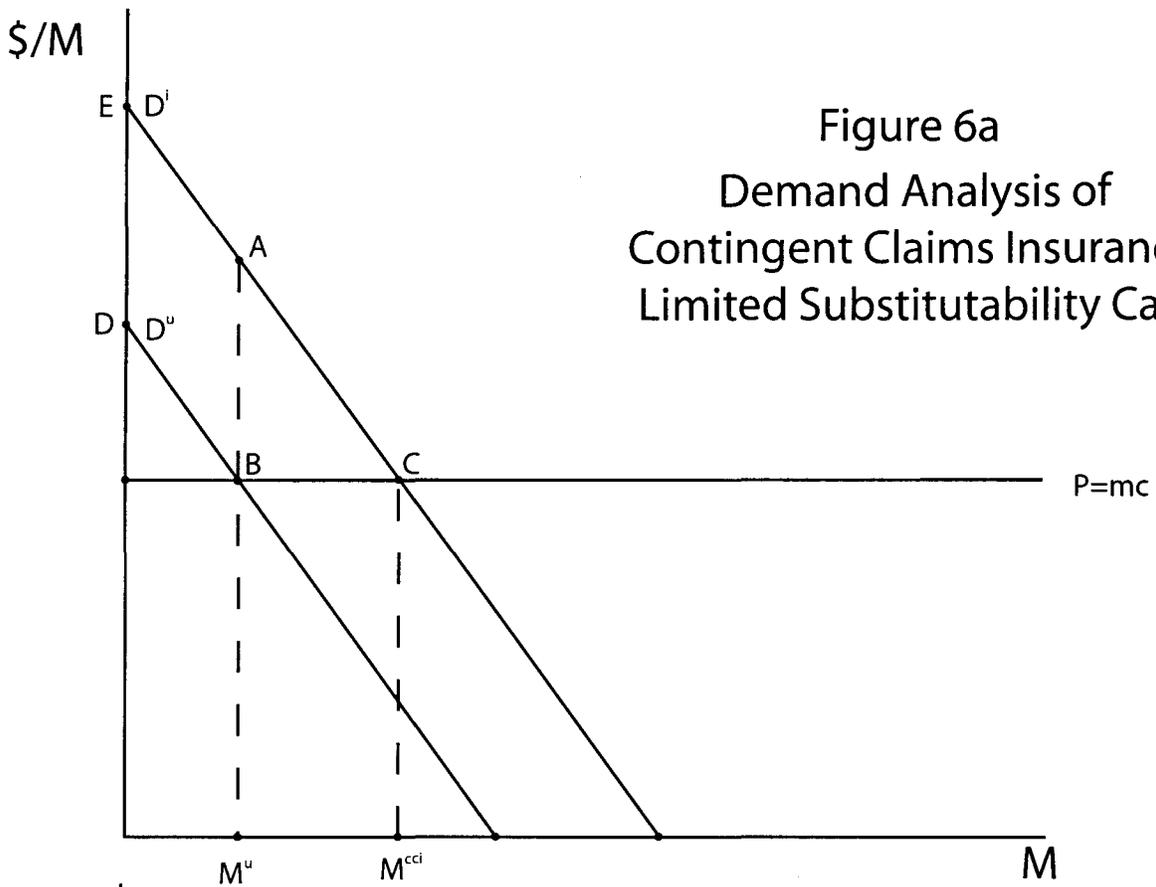


Figure 5

The Consumer's Decision Under Contingent-Claims and Price-Payoff Insurance: Limited Substitutability Case



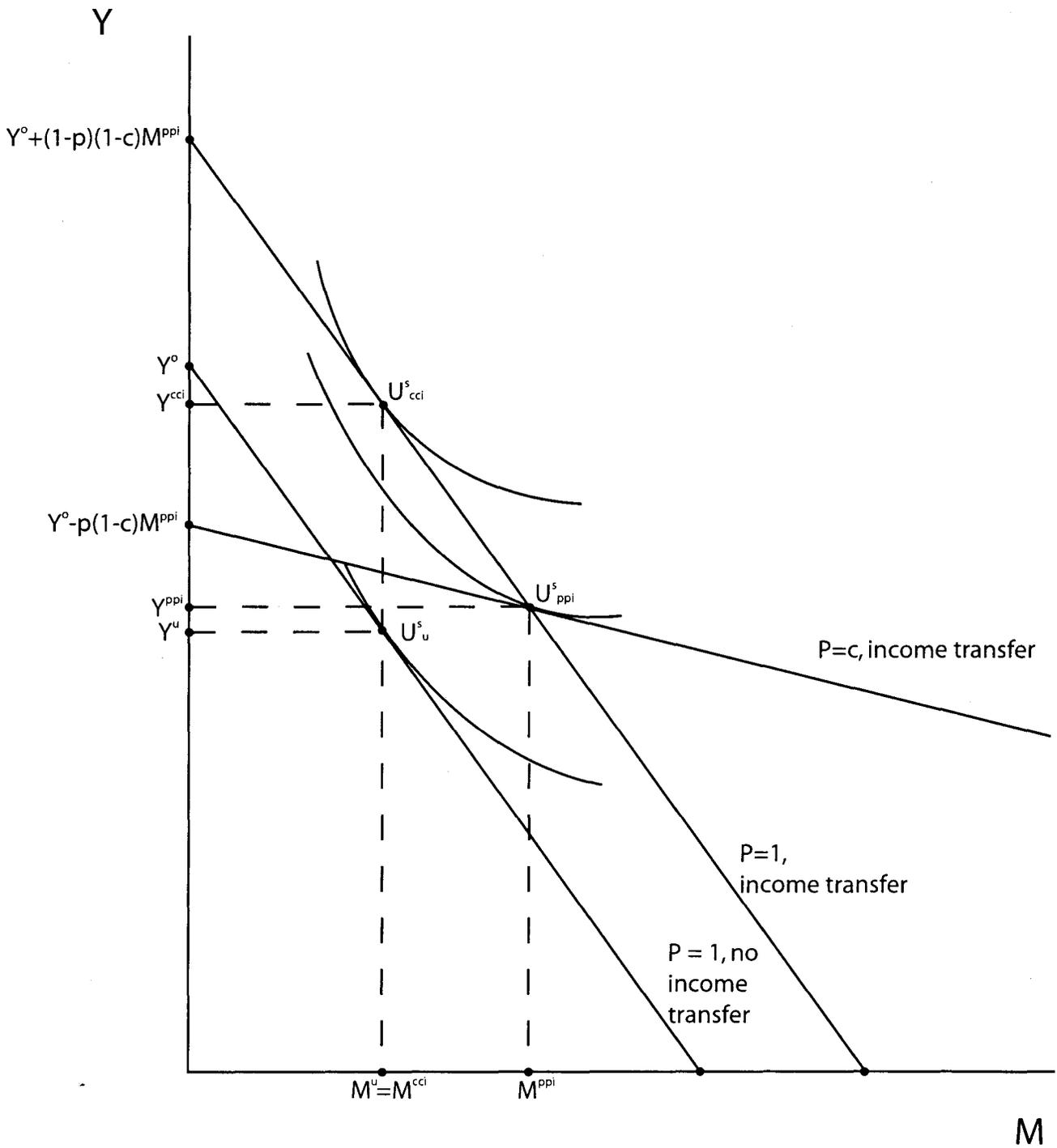


Figure 7

The Consumer's Decision Under Conventional Assumptions of Total Substitutability and No Income Effect

Figure 8a
 Demand Analysis of
 Contingent Claims Insurance
 Under Conventional Assumptions

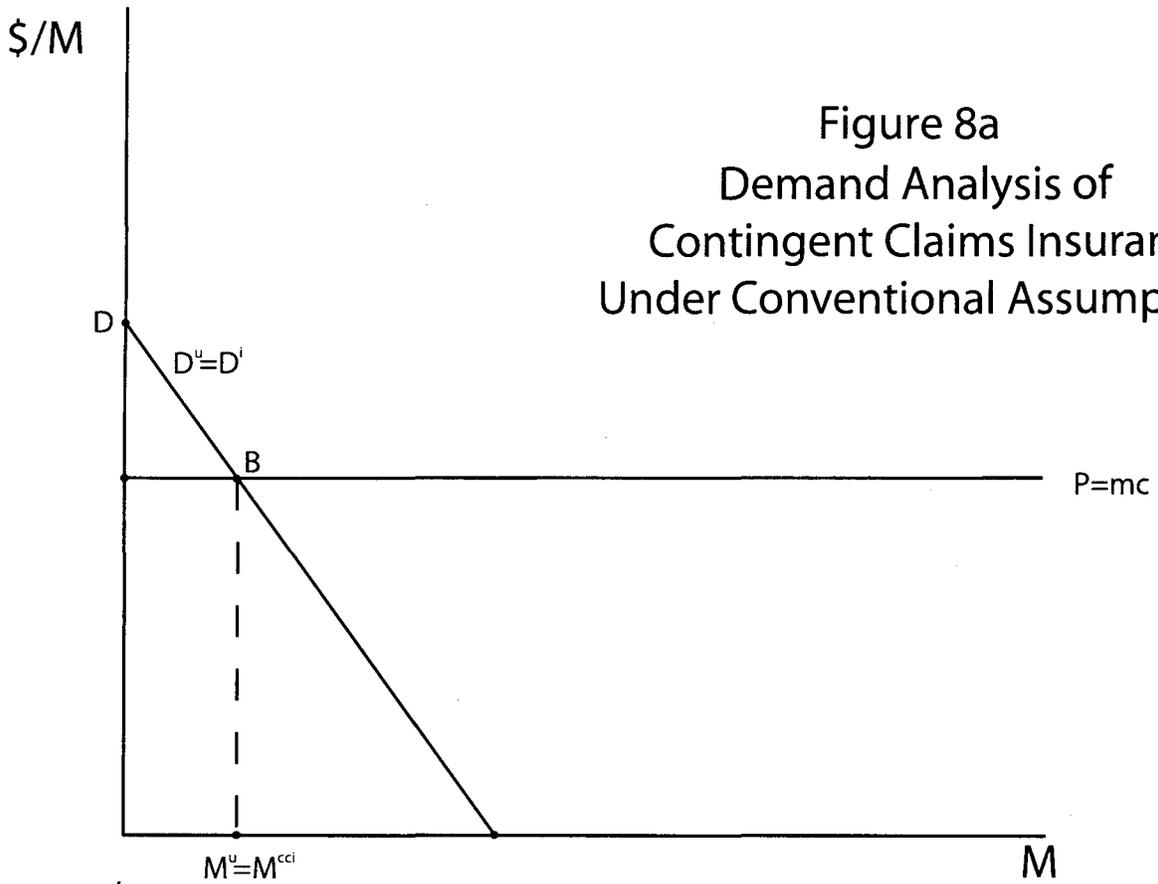
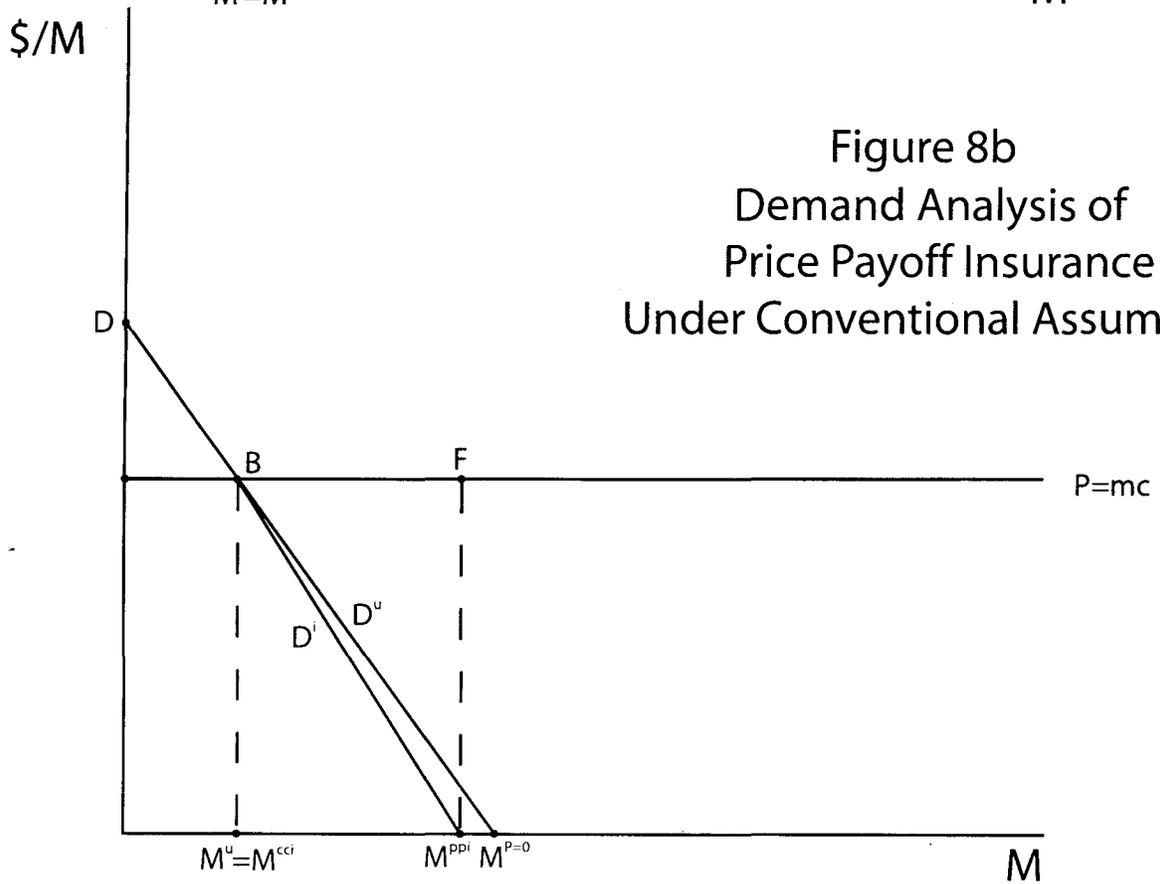


Figure 8b
 Demand Analysis of
 Price Payoff Insurance
 Under Conventional Assumptions



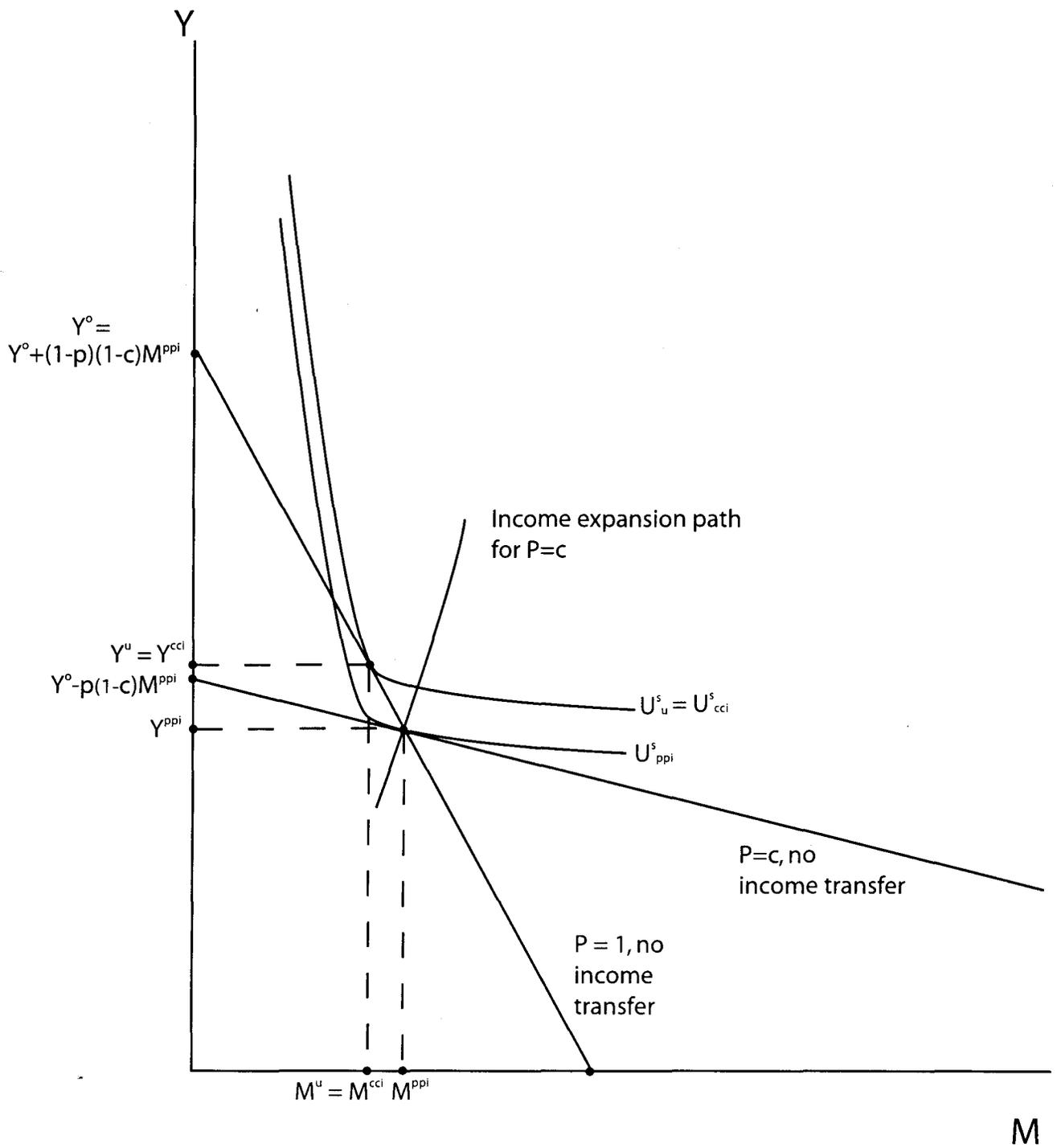


Figure 9
 The Consumer's Decision Under Insurance
 When No Income Transfers Occur

Figure 10a
 Demand Analysis of
 Contingent Claims Insurance
 With No Income Transfer

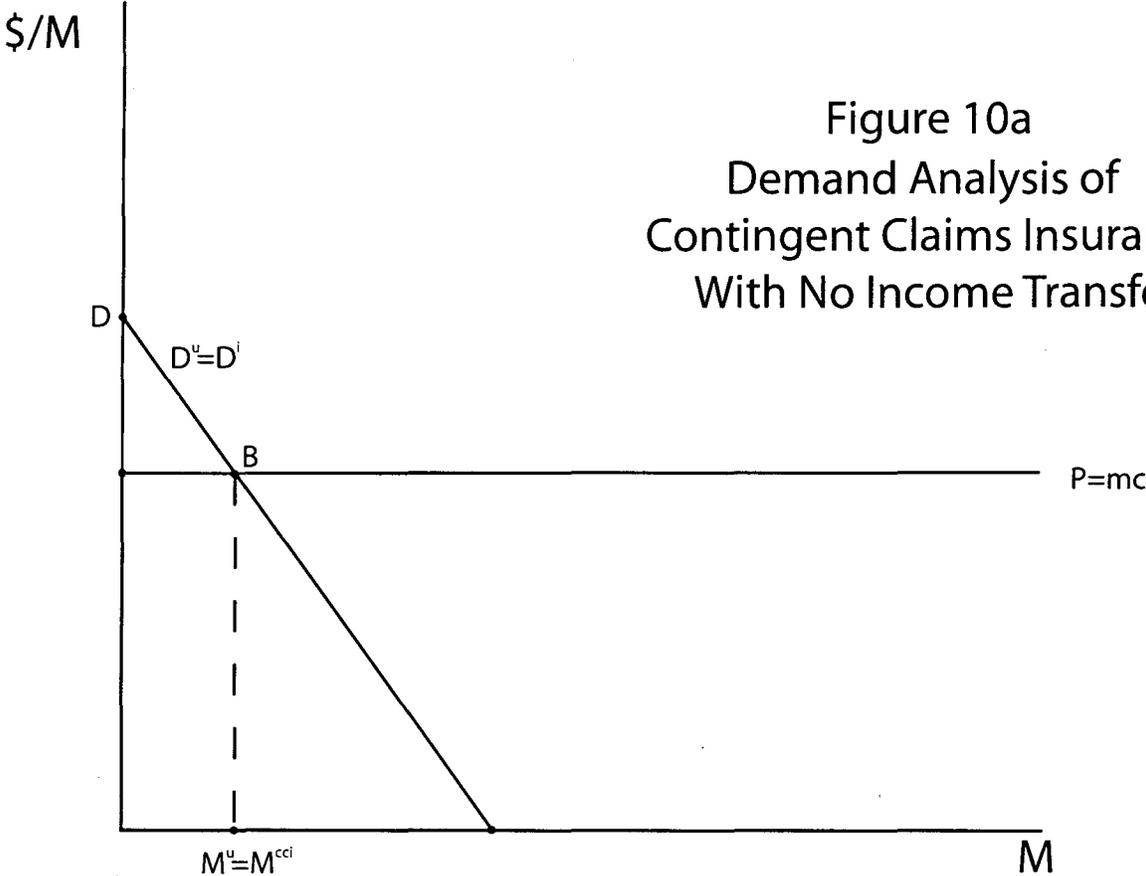


Figure 10b
 Demand Analysis of
 Price Payoff Insurance
 With No Income Transfer

