The Theory of Demand for Health Insurance: A Review Essay

Joseph G. Eisenhauer*

Abstract: This paper critically evaluates a recently introduced theory of the demand for health insurance. The new model makes a valuable contribution to the theoretical literature on moral hazard and the debate over national health insurance, but as a general explanation of the demand for health insurance it fails to provide a robust alternative to conventional theory. [Key words: Moral hazard, medical care, risk aversion]

INTRODUCTION

In a justly famous article published more than four decades ago, Kenneth J. Arrow (1963) argued that where private markets for insurance—particularly health insurance—were absent, a strong case could be made for governmental provision of insurance. In an almost equally well known comment five years later, Mark V. Pauly (1968) observed that health insurance often induces moral hazard, resulting in an inefficient reallocation of resources, and that institutionalizing such inefficiency through government regulation could potentially be welfare-reducing. Thus, moral hazard weakened the case for national health insurance.[1]

John A. Nyman's (2003) book, *The Theory of Demand for Health Insurance*, reconsiders moral hazard and offers a new perspective on the reason why consumers buy medical insurance in the first place. His explanation is a departure from conventional views, and Professor Nyman acknowledges at the outset that his position is controversial in several respects. Though

*Joseph G. Eisenhauer (eisenhauer@canisius.edu) is Professor of Economics and Finance at Canisius College in Buffalo. His research on risk and insurance has appeared in *Insurance: Mathematics and Economics*, the *Journal of Risk and Insurance*, the *Journal of Insurance Issues*, *Risk Management and Insurance Review, Applied Economics*, and other professional journals.

Journal of Insurance Issues, 2006, **29**, 1, pp. 71–87.

it is framed largely as a refutation of Pauly's analysis, the book is by no means an endorsement of Arrow's work; indeed, it rejects the fundamental notion that risk aversion motivates insurance purchases. A treatise that seeks to overturn established theory so thoroughly demands scrutiny, and to that end, the present paper briefly reviews the literature leading up to Nyman's book, evaluates the author's critique of standard theory, and then examines his proposed alternative.

THE MORAL HAZARD DEBATE

Pauly's (1968) essay assumed a fixed individual demand curve for health care and a constant marginal cost of production. Together, these determined an efficient optimum for an uninsured patient: the marginal willingness to pay for care (as represented by the demand curve) was equal to the marginal cost of care. If the same individual were insured, however, she would perceive a lower out-of-pocket price for care (zero, if there was no coinsurance), and move down the demand curve; unless demand had no price-elasticity, the insured would then consume more units of medical treatment. The marginal cost of health care would exceed the consumer's willingness to pay for the extra units, and inefficiency would thereby be introduced. Moreover, forcing such individuals to pay for access to this care through taxes could potentially make them worse-off than they would be without insurance.

Fifteen years later, David de Meza (1983) argued that an ill consumer's demand curve is not the same when insured as when uninsured. Rather, the reimbursement of medical expenses provided by insurance shifts the demand curve outward just as a cash transfer would. Thus, the consumer's willingness to pay increases with insurance coverage, and Pauly's (1968) model therefore overstates the inefficiency induced by moral hazard. In response, Pauly (1983) acknowledged that income effects might indeed matter for critically ill patients, but asserted that moral hazard among healthier consumers was still largely inefficient.

In subsequent articles, Nyman (1999) and Nyman and Roland Maude-Griffin (2001) elaborated on de Meza's basic insight, using indifference curves and budget constraints to illustrate the difference between efficient and inefficient moral hazard.[2] Nyman's (2003) book pursues this idea even further—indeed, it seeks to expand this analysis of moral hazard into an entirely new theory of the demand for health insurance.

REINTERPRETING CONVENTIONAL ANALYSIS

As a background against which to introduce his own model, Nyman first lays out what he regards as the five main anomalies or flaws in conventional theory. "Foremost of these is the presumption that any additional health care consumed because of being insured is welfare decreasing because it implies a movement along the demand curve" (p. 21). Indeed, Nyman (p. 1) claims, "Since the advent of Pauly's influential article, almost all economists (Fuchs, 1996) have believed in a theory that implies that the *voluntary purchase* of health insurance *makes the consumer worse off*" (emphasis added). But Pauly had argued simply that *mandatory* participation could potentially reduce welfare for consumers. In Pauly's (1968, p. 534) words, "inefficiency may well be created if individuals are forced, by taxation, to 'purchase' insurance which indemnifies against some kinds of medical care expense" and thus "the net change in utility from a compulsory purchase of this 'insurance' could well be negative."[3] Nor did the economists surveyed by Fuchs (1996) maintain that a voluntary transaction reduces the buyer's utility; rather, Fuchs found broad agreement with the statement that third-party payments lead patients to use services whose costs exceed their benefits. That is, most economists agreed that moral hazard occurs; but since the consumer garners all the benefits of the extra care and passes much of the cost on to others, there is no evidence that economists believe the insured who engages in moral hazard is individually made worse-off. On the contrary, most economists (especially those with a neoclassical orientation) would doubtless consider it irrational and thus extraordinary for an individual to voluntarily purchase insurance (or anything else) that reduced his or her own utility. Provided that the transaction is undertaken with full information, the sale of insurance should benefit both the buyer and the seller, though parties outside the transaction are sometimes made to bear a portion of the costs.[4]

The second flaw, Nyman (p. 22) argues, is that "conventional expected utility theory of the demand for health insurance specifies the loss as exogenously determined. … For example, if a consumer who becomes ill without insurance purchases $20,000 worth of medical care, the same consumer with insurance would again purchase the same $20,000 of medical care, even though insurance has paid off with an extra $20,000 in income." This would arguably be a correct characterization of conventional theory *in the absence of moral hazard*, but in the wake of Pauly's (1968) essay, a large mainstream literature now emphasizes the notion that consumers utilize additional care because of insurance. Nor does the previous neglect of the income effect of insurance justify Nyman's assertion (p. 22) that "the conventional model implicitly assumes that the consumer has an income

elasticity of demand [for medical care] of 0." Indeed, prior attempts to estimate income-elasticities for medical care, such as those by Feenberg and Skinner (1994) or Manning and Marquis (1996), demonstrate otherwise.

Third, Nyman rejects the conventional proposition that consumers prefer certainty to risk, and are thus motivated to purchase insurance by their risk aversion. As evidence, the author invokes prospect theory, which Kahneman and Tversky (1979) first proposed to explain violations of expected utility. On the basis of experiments conducted by Kahneman, Tversky, and others, Nyman (p. 2) concludes, "consumers actually prefer an uncertain loss to a certain loss of the same expected magnitude." Indeed, he asserts (p. xiv), "Consumer preferences regarding risk are largely irrelevant, and, if anything, inhibit the purchase of health insurance." There is, however, simply too much empirical evidence of risk aversion to dismiss conventional wisdom so easily. Indeed, the prevalence of risk aversion has been recognized since the time of Bernoulli (1738), and has been widely documented during recent decades in studies of health insurance (Friedman, 1973), property insurance (Szpiro, 1986a, b), life insurance (Eisenhauer and Halek, 1999; Halek and Eisenhauer, 2001), and numerous non-insurance contexts. And, as shown below, Nyman's alternative theory still appears to depend subtly on aversion to risk.

A fourth and related issue concerns the convention of equating risk aversion with the diminishing marginal utility of income, a proposition originally introduced by Bernoulli (1738) and later formalized by Arrow (1971). Nyman suggests that the two concepts are different, and can be distinguished by the following thought experiment. Suppose that an individual assigns a utility value of 0 to \$50,000 and a utility value of 1 to \$150,000, and is then asked how much utility \$100,000 would generate. Because of diminishing marginal utility, the individual might reasonably assign twice as much extra utility to the first increment of \$50,000 as to the second, so that the utility of \$100,000 is 0.67, determined in an environment of certainty. If the same individual is then asked to equate the utility of \$100,000 with the utility of a gamble having a π probability of winning \$150,000 and a $(1 - \pi)$ probability of winning \$50,000, Nyman argues that she may not require a probability of $\pi = 0.67$ in order to be indifferent between the certain wealth and the gamble. It is true, of course, that inconsistent decision-making is possible, and preferences may not be stable. But given that $U(\$100{,}000) = 0.67$ where U denotes utility, setting $U(\$100{,}000) = \pi(1) + (1 - \pi)(0)$ is logically equivalent to setting $0.67 = \pi$. Thus, it is not the theory that should be regarded as anomalous, but rather any observed violation of it.

Indeed, Nyman refers to a concave utility function defined over riskless income as a Bernoulli function to distinguish it from the expected

utility theory of von Neumann and Morgenstern, and thereby separate diminishing marginal utility from risk aversion. But the two concepts cannot be so easily divided. Daniel Bernoulli's (1738) stated purpose in developing the theory of diminishing marginal utility was to explain aversion to financial risk; specifically, to offer a resolution to a gambling puzzle posed by his cousin Nicholas Bernoulli.[5] Thus, from its inception, diminishing marginal utility has been inextricably linked with risk aversion. Arrow's (1971) subsequent use of concavity to measure risk aversion essentially represents a direct extension of Bernoulli's innovation.[6]

Finally, Nyman points out that several studies have concluded that existing coinsurance rates are too low to effectively preclude moral hazard, and have recommended raising the rates as a means of enhancing social welfare. Nyman (p. 27) presents this as his fifth anomaly, writing, "The question therefore arises: Why are such insurance policies voluntarily purchased if they make the purchaser worse off? …[E]ither this voluntary purchase of unsubsidized insurance represents an irrational act, or the conventional theory is, in some fundamental way, flawed." But this assertion confuses the welfare of society at large with the welfare of the individual who appropriates resources for which others must contribute partial payment. Free riding is not necessarily irrational, and Pauly (1968, p. 535) even emphasized that moral hazard was a consequence of "rational economic behavior." When the difference between the gain to the individual and the cost to society is recognized, the alleged anomaly disappears.[7, 8]

AN ALTERNATIVE THEORY

The new theory postulates that the central rationale for buying insurance is the individual's desire to obtain an income transfer from the risk pool if she becomes ill. This is a valid observation, but it begs the question of why the consumer would pay a loaded premium upfront for a smaller expected transfer in the future. One possibility is that the consumer seeks to smooth out consumption (or wealth) across time by sacrificing a little when healthy to be compensated in the event of injury or illness; that is, to avoid the risk of a potentially large and perhaps unaffordable medical bill in the future. In that case, however, the consumer exhibits the classic symptoms of risk aversion, which Nyman rejects. Instead, Nyman argues that the demand for health insurance is derived from the access it provides to medical care, which generates more utility than does the income spent on premiums. Thus Nyman contends that insurance buyers do not need to be especially risk averse, though he is not prepared to dismiss the principle of diminishing marginal utility, and indeed, his model requires it; hence

his insistence that one can have diminishing marginal utility without risk aversion.

Formally, the model operates as follows.[9] In the absence of insurance, a consumer with initial income of Y^0 maximizes the utility function $U^S(M, Y)$ when sick, where M is medical care (priced at $P = 1$) and Y is residual income available for purchases of other goods. The budget constraint is $Y^0 = M + Y$. The first-order condition is reported as $U^S_M/U^S_Y = -1$, though in fact the ratio of marginal utilities should be positive. The second-order condition is not stated, but it can easily be derived as $U^S_{MM} - U^S_{MY} - U^S_{YM} + U^S_{YY} < 0$, which holds most conveniently when, as in standard microeconomic theory, the cross-partials are non-negative and the second derivatives are negative. At least to that extent, the model implicitly requires concavity of utility.

Similarly, an insured consumer maximizes the $U^S(M, Y)$ utility function when sick, subject to the constraint $Y^0 - R = cM + Y$, where R is the insurance premium and c is the coinsurance rate. Here, the first-order condition is reported to be $U^S_M/U^S_Y = -c$. Of course, the ratio should again be positive, but if R was constant, no further adjustment to the first-order condition would be needed. However, Nyman also treats R as a function of M (so that R is not truly fixed unless M is known), and indeed, the author sets $R = \pi(1-c)M^i$, where π is the probability of illness and M^i is the quantity of medical care consumed when insured and ill. This framework effectively requires the insurance company to correctly anticipate the extent of moral hazard and incorporate that level of usage into the premium. Assuming that to be possible, the first-order condition implicitly defining the optimal level of medical care for an ailing insured would be $U^S_M/U^S_Y = \pi(1-c) + c$. Under this specification, the model implies that a marginal increase in either the probability of illness or the coinsurance rate will reduce the demand for medical care because the premium and coinsurance expenses borne by the insured will increase.

The model can be summarized graphically by Figure 1, which depicts the choices facing a consumer when she is ill.[10] If uninsured, the consumption possibilities defined over M and Y are given by a budget constraint of slope –1. Having convex indifference curves, the consumer optimizes at point 1, achieving utility level U^S_u with M^u units of medical treatment. With insurance, the budget rotates: the premium reduces discretionary income, but the insurance effectively reduces the per unit price of medical

care for the individual from $P = 1$ to $P = c$. Both the income and substitution effects of becoming insured encourage more consumption of M, and the new optimum is achieved at point 2 with M^i units of care. The difference, $M^i - M^u$, represents total moral hazard. Notice, however, that a *cash* transfer (net of the premium) would have left the slope of the budget constraint equal to –1 as depicted by the dotted line, and the consumer would have optimized at point 3, purchasing M^c units of care. Because the income effect reflects a shift in demand (or an increase in the willingness to pay), Nyman refers to $M^c - M^u$ as "efficient moral hazard," and describes $M^i - M^c$, which results from a price effect (a movement along the demand curve), as "inefficient moral hazard." This distinction provides a valuable insight into the effect of insurance on the utilization of medical care.[11]

As an example, Nyman considers a cancer patient who buys $20,000 worth of medical care without insurance, but who would have utilized $40,000 in care if she had purchased insurance at a premium of $3,000: the net transfer of $37,000 would induce $20,000 in moral hazard. If she would have spent $15,000 out of a $37,000 *cash* transfer on additional care, then $15,000 worth of the extra care utilized under insurance is deemed efficient and the remaining $5,000 expenditure is deemed inefficient. In that event, most of the additional medical treatment ($15,000/$20,000 or 75 percent) is care that the individual would be willing to buy if her income was sufficient. But by the same token, if her income-elasticity of demand for medical care was 0.38 as estimated by Feenberg and Skinner (1994)—a value that Nyman considers conservative but suitable for illustration—and her initial income was $50,000, then a $37,000 cash transfer would induce only a 28.12 percent increase in utilization of care, or an extra $5,624 in spending on M. In this case, $14,376 or about 72 percent of the total moral hazard would be regarded as inefficient.[12]

As the example suggests, the inefficiency associated with moral hazard will vary widely across consumers depending upon their individual circumstances. In particular, because the income effect of any given insurance payout will be more pronounced for a low-income insured than for a high-income insured, holding other variables constant, the inefficiency rises with initial income. Thus, Nyman (p. 156) writes, "For low-income consumers in wealthy countries, or for most consumers in poor countries, the efficient portion of moral hazard tends to dominate the inefficient portion." This aspect of the model is arguably its most important contribution. A direct policy implication is that insuring (or subsidizing insurance among) the poor may cost relatively little in terms of efficiency.

To estimate the average inefficiency for the U.S., Nyman (2003) derives the percentage increase in medical care utilization induced by the price (or

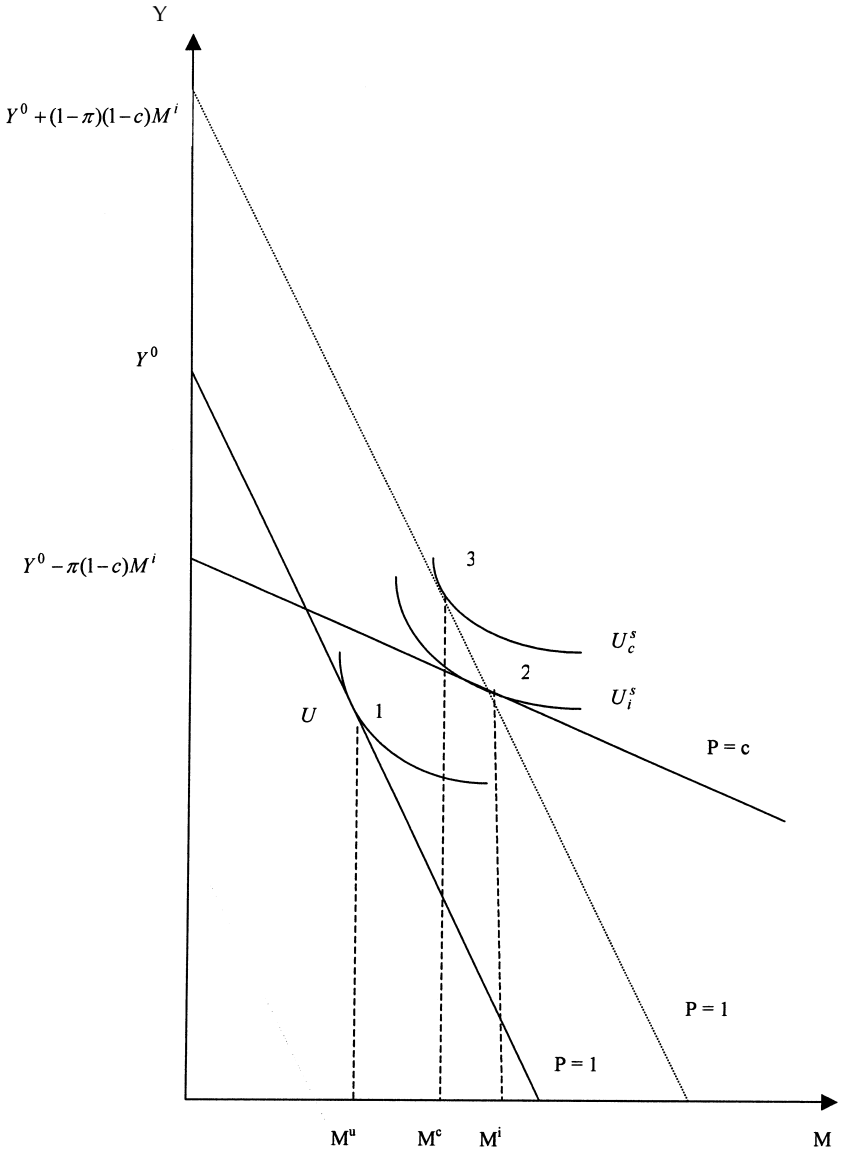

Fig. 1. Efficient and inefficient moral hazard.

substitution) effect; the same derivation appeared earlier in Nyman and Maude-Griffin (2001). Using the notation of Figure 1, inefficient moral hazard, as a proportion of uninsured care, is given by

$$(M^i - M^c)/M^u = (\eta + \theta\varepsilon)(c-1)/[1 + (1-c)\theta\varepsilon],$$

where ε denotes the income-elasticity of the demand for medical care, η is the price-elasticity of demand, and θ is the post-transfer share of expenditures devoted to medical care. Ignoring income effects (so that $\varepsilon = 0$), total moral hazard as a percentage of uninsured purchases can be represented by $(M^i - M^u)/M^u = \eta(c-1)$, and the ratio $(M^i - M^c)/(M^i - M^u)$ indicates the proportion of moral hazard that is inefficient. Using a 31 percent coinsurance rate, the Feenberg and Skinner (1994) income-elasticity of 0.38, a price-elasticity of $\eta = -0.18$ as estimated by Manning and Marquis (1996), and a budget share of 32 percent, Nyman and Maude-Griffin (2001) estimated that health insurance in the U.S. typically raises medical expenditures by 12.4 percent, with 30 percent of the increase being inefficient, as shown in panel A of Table 1.

Roughly the same relative inefficiency is obtained by Nyman (2003); as shown in panel B, however, his parameter values now include a price-elasticity of $\eta = -0.32$, as calculated by Newhouse (1993) from the Rand Health Insurance Experiment (HIE), and a budget share of 54 percent, yielding substantially greater moral hazard. Yet θ could also have been estimated from the HIE. After rescaling monetary values to 1984 dollars, consumers in the HIE had a mean income of $8,912. On average, HIE patients in the fee-for-service plan received net insurance transfers of $3,562 and spent $3,971 on medical care, so their medical budget share can be computed as $\theta = 3971/(8912 + 3562)$ or 31.8 percent, nearly identical to the average budget share originally estimated for the U.S. by Nyman and Maude-Griffin (2001). Similarly, HIE patients assigned to a cost-sharing plan received average net transfers of $2,672 and spent $3,987 on medical care; their budget share can be calculated as $\theta = 3987/(8912 + 2672)$ or 34.4 percent. Thus, a 54 percent budget share seems excessive. Recalibrating the model with the original budget share of 32 percent and Newhouse's price-elasticity of $\eta = -0.32$, panel C of Table 1 estimates total moral hazard to be exactly the same as in Nyman (2003), but shows that 57 percent of it is inefficient. Moreover, Eichner (1998) has estimated the price-elasticity of medical care to be in the neighborhood of $\eta = -0.62$; using this value implies nearly twice as much moral hazard as Nyman (2003) estimates, with 74 percent of it being inefficient, as shown in panel D. Finally, panel E shows that if the Feenberg-Skinner income-elasticity is replaced by the

Table 1. The Relative Inefficiency of Moral Hazard

Panel	Sources	Parameter values	Total moral hazard*	Inefficient moral hazard*	Relative inefficiency
A	Nyman & Maude-Griffin (2001): Manning & Marquis (1996): Feenberg & Skinner (1994):	$\theta = .32$ $\eta = -.18$ $\varepsilon = .38$	.1242	.0372	30%
B	Nyman (2003): Newhouse (1993): Feenberg & Skinner (1994):	$\theta = .54$ $\eta = -.32$ $\varepsilon = .38$	.2208	.0694	31%
C	Nyman & Maude-Griffin (2001): Newhouse (1993): Feenberg & Skinner (1994):	$\theta = .32$ $\eta = -.32$ $\varepsilon = .38$	.2208	.1263	57%
D	Nyman & Maude-Griffin (2001): Eichner (1998): Feenberg & Skinner (1994):	$\theta = .32$ $\eta = -.62$ $\varepsilon = .38$	.4278	.3173	74%
E	Nyman & Maude-Griffin (2001): Eichner (1998): Manning & Marquis (1996):	$\theta = .32$ $\eta = -.62$ $\varepsilon = .22$	.4278	.3617	85%

*Moral hazard is shown as a percentage of uninsured purchases of medical care.
All calculations assume a coinsurance rate of $c = 0.31$.

Manning and Marquis (1996) estimate of $\varepsilon = 0.22$, then 85 percent of the moral hazard is inefficient. Thus, the wide variation in empirical estimates of the parameter values makes it difficult to determine accurately how much moral hazard commonly arises, and what proportion of it is inefficient.[13]

The moral hazard analysis is concerned exclusively with the consumer's behavior in a state of illness, but Nyman recognizes that the initial decision to insure depends on utility in both the healthy and unhealthy states. In particular, the individual weighs the expected utility from remaining uninsured,

$$EU_u = \pi U^S(M^u, Y^0 - M^u) + (1-\pi)U^h(0, Y^0),$$

against the expected utility from being insured,

$$EU_i = \pi U^s(M^i, Y^0 + (1-\pi)(1-c)M^i - M^i) + (1-\pi)U^h(0, Y^0 - \pi(1-c)M^i),$$

where the superscripts *s* and *h* denote sick and healthy states, respectively; insurance is then purchased if $EU_i > EU_u$. Here, Nyman accepts the basic premise of expected utility. In this framework, however, utility is state-dependent, so that the functional forms of utility in states *s* and *h* may not be the same. This allows medical care to enter the utility function directly, in contrast to the indirect utility function defined only over final wealth.[14] Nyman notes that if these assumptions were relaxed, so that utility was state-independent and the first argument of each felicity function was suppressed, we would return essentially to the standard model. Then if $M^i = M^u = M^*$ (that is, if moral hazard is absent), expected utility is

$$EU_u = \pi U(Y^0 - M^*) + (1 - \pi)U(Y^0)$$

when the consumer is uninsured and

$$EU_i = \pi U(Y^0 - M^* + (1 - \pi)(1 - c)M^*) + (1 - \pi)U(Y^0 - \pi(1 - c)M^*)$$

when the consumer is covered by insurance at the $R = \pi(1 - c)M^i$ premium. Then a consumer with a concave utility function will purchase insurance if the coinsurance rate is sufficiently low.[15]

In conventional analyses, however, the decision to insure is commonly nested within the decision of optimal coverage. That is, EU_i is maximized by choosing *c*; at an actuarially fair premium this leads to full coverage, or an optimal coinsurance rate of $c = 0$, whereas with a loaded premium, partial coverage ($c > 0$) would be optimal, and a corner solution of $c = 1$ would imply a decision to remain uninsured.[16] In contrast, the choice Nyman models is between buying a policy with a given coinsurance clause or remaining uninsured, an all-or-nothing decision.[17] Ironically, access to otherwise unaffordable care is discussed extensively as a primary motive for buying insurance, though this motive might be better addressed by a model that allows a marginal decision: without the ability to choose the optimal level of coverage, an insured could still be left with unaffordable medical bills in the event of illness, should the required co-payment exceed the consumer's wealth. Perhaps this explains why Nyman later makes reference to a marginal decision when he suggests (p. 136–137), "If it is assumed that the consumer does not derive direct benefit from the loading fees, increases in the loading-fee portion of the insurance premium might reduce the quantity of health insurance demanded."

Although he emphasizes the mathematical equivalence between his model and expected utility theory, Nyman objects to risk aversion as the

basis for buying insurance. He therefore delves briefly into prospect theory, where the consumer's value function is assumed to be concave over gains and convex over losses. Framing the consumer's decision in terms of losses—that is, comparing the sure payment of an insurance premium to the uncertain expense of medical bills—Nyman (p. 55) concludes, "insurance should not be purchased according to this specification."[18] In an appendix, he respecifies the consumer's decision as a choice between two gains, and reconciles the purchase of insurance with the concave portion of the value function; but this seems to show that it is the concavity of the objective function (i.e., risk aversion) rather than prospect theory per se that drives the purchase of insurance. After this brief discussion, however, prospect theory disappears from the text, and Nyman eventually states (p. 134) that his own model "differs from both expected utility theory and prospect theory in that it does not rely on switching from concave to convex utility (value) functions to explain behavior."[19]

Though he ultimately endorses a concave utility function, Nyman's artificial separation of risk preference from the shape of the utility function and his rejection of risk aversion as a general characteristic of insurance buyers render the proposed theory unconvincing. An example from chapter 8 is revealing. Nyman compares spending $100 for an insurance policy that pays $200 in the event of a $1,000 medical bill, which occurs with 50% probability, against placing a $100 bet that pays $200 if a tossed coin lands heads-up. Professor Nyman correctly observes that a consumer with a concave utility function would purchase the insurance but not place the wager, but he argues that the difference is not attributable to risk aversion. In his words (p. 129), "There is no difference between these two contracts: both the insurance and gambling contracts represent paying $100 for a 0.5 probability of winning $200. … Both contracts have exactly the same amount of risk associated with them." That assessment, however, ignores the loss of income occasioned by the $1,000 medical bill. In fact, the expected loss of wealth with or without the insurance is $500, but the insurance reduces the standard deviation of final wealth from $500 to $400, which is what makes it attractive to a risk averter. In contrast, the expected gain from the wager is zero, but the wager raises the standard deviation of final wealth from zero to $100, which makes it unattractive to a risk averter. Thus, both decisions are attributable to risk aversion. To circumvent this criticism, Nyman (p. 133) argues that "the medical spending is really not part of the insurance contract" and should not be measured as part of the risk. But of course, the potential medical expenditure that triggers the insurance reimbursement is the origin of the financial risk and is an explicit feature of any health insurance policy.

The concept of risk aversion need not be limited to wealth fluctuations, however, and the access motive that Nyman endorses may even be viewed as a reflection of the consumer's aversion to health risks. If the consumer knew with certainty that she would never need medical treatment, she would presumably not be willing to pay for health insurance. It is the *risk* of becoming ill (at an uncertain time and with unpredictable severity and duration) that prompts a desire for access to medical care. In this respect, risk aversion (with regard to health) and the access motive may be viewed as complements rather than alternatives.[20] Even so, however, a general theory should also be able to account for wealthy individuals who purchase insurance not because they would otherwise lack access to expensive treatment, but simply because paying a premium when healthy helps them to avoid devastating medical bills in the event of an illness.

Indeed, if Nyman were correct in arguing that consumers are generally not risk averse, the implications would extend far beyond health insurance. Purchases of all forms of insurance and countless other behaviors would require an alternative explanation. Consider the simple example of travelers' checks. Consumers who fear losing or being robbed of cash often pay a modest fee to carry travelers' checks, which have essentially the same probability of being lost or stolen but which guarantee replacement in full. Since one cannot plausibly separate the utility of travelers' checks from the financial security they provide, the only sensible interpretation is that those who buy them are risk averse.

CONCLUSION

Despite the fact that many individuals in developed economies now receive health insurance as either an employee benefit or a public entitlement, there is undoubtedly a place in the economic literature for a theory of the demand for privately purchased health insurance. The established canon posits that a risk-averse individual facing a known premium rate and probability of illness maximizes expected utility by choosing an optimal level of coverage. The contract allows the risk averter to pay a certain premium in the present in order to avoid uncertain but potentially large (and perhaps otherwise unaffordable) medical bills in the future. After the contract has been established, moral hazard may arise either because the insured takes fewer health precautions (raising the probability of illness and the frequency of medical care utilization) or because the insured utilizes a greater quantity of medical care when ill (possibly on a physician's advice) than had been anticipated. Thus, moral hazard is tradition-

ally viewed as an unintended change in behavior that occurs after an individual becomes insured.

In Nyman's view, most or all of what has been called moral hazard is planned from the outset. The insurance company correctly anticipates the amount of medical care that an insured would use, establishes a coinsurance rate, and sets a premium based on the probability of illness. The consumer likewise anticipates the amount of care she would use, and decides whether to purchase the policy or remain uninsured. If she buys the policy and subsequently becomes ill, she utilizes precisely the amount of care that had been anticipated, which may be only slightly more than she would have purchased with an equivalent cash subsidy instead of insurance. Thus, Nyman's model imbues consumers with greater foresight than existing theory attributes to them; neither the probability of illness nor the expected quantity of care differs between the time when coverage is purchased and the time at which treatment is received. Indeed, both parties to the insurance contract seem able to correctly predict the physician's recommendation for treatment. Yet, these foresighted consumers are said to prefer uncertainty to certainty, and purchase insurance despite their preference for risk.[21] As a consequence of such assumptions, the attempt to elevate moral hazard into a general theory of the demand for health insurance that replaces the role of risk aversion falls short of the mark.

As a contribution to the *theory of moral hazard*, however, Nyman's model is valuable. It clearly demonstrates that some portion of the additional medical care utilized under insurance represents an efficient use of resources, because the consumer's willingness to pay for care differs between the insured and uninsured positions. An important policy implication is that extending insurance to the indigent may be substantially less inefficient than previously believed. As Nyman readily acknowledges, though, the likely magnitudes of the relevant parameters have not yet been estimated with sufficient precision to provide a reliable gauge of the relative inefficiency of moral hazard; this aspect of the model therefore requires and deserves further empirical investigation.

NOTES

[1] Arrow (1963) had also mentioned moral hazard, but argued that it had been adequately minimized by the introduction of coinsurance clauses in health insurance contracts.

[2] De Meza (1983) compared insurance to self-insurance financed by saving and borrowing; the later models assume credit constraints that preclude saving or borrowing.

[3] Some forms of moral hazard also impose (confer) negative (positive) externalities on others in society (Eisenhauer, 1996), but that is not directly addressed either by Pauly or Nyman.

[4] Nyman's assertion that economists generally view the extra care as frivolous (e.g., designer sunglasses) and distinct from basic care is not supported by Pauly's (1968) model, in which

more units of care *of a given quality* (e.g., additional days in the hospital) are consumed as one moves down the demand curve. Nyman criticizes Rice (1992) for failing to hold quality constant along a demand curve, but other models cannot rightly be accused of this.

[5] The puzzle, subsequently known as the St. Petersburg paradox, concerned the observation that individuals are generally unwilling to pay a large sum of money to participate in a gamble with an infinite expected value—an example of the phenomenon of risk aversion. Bernoulli (1738) hypothesized that a potentially large but risky payoff provided no more utility than a smaller but certain sum, and even proposed the use of expected utility to value risk.

[6] More recently, Werner's (2005, p. 74) axiomatic treatment has shown "that preferences have concave expected utility representation if and only if they exhibit risk aversion and satisfy the independence axiom (sure-thing principle)."

[7] Manning and Marquis (1996, p. 610), for example, are among those who find the prevailing coinsurance rates to be too low; as they explain, "Because consumers would not purchase this additional care if they had to pay its full cost at the margin, the extra services' value to consumers falls short of the *social cost* of producing that care" (emphasis added).

[8] Nyman suggests that income tax deductions for health insurance premiums could explain the purchases, but that these tax incentives are likely to be offset by premium loading. Eisenhauer (2002) notes that both uninsured medical expenses and health insurance premiums are tax deductible, and that under premium loading, the former has a greater effect than the latter, generating a net *reduction* in coverage roughly proportional to the marginal tax rate.

[9] One of the principal difficulties with the book's exposition is that there are a number of distracting mathematical and typographical errors. In what follows, I attempt to clarify the model so that it may be judged appropriately.

[10] Figure 1 here is closest to Figure 3.1 in Nyman's (2003) text.

[11] Blomqvist (2001) notes that when the choice is between full coverage and partial coverage rather than between coverage and no coverage, Nyman's distinction between efficient and inefficient moral hazard is of less consequence. Nevertheless, Nyman's demonstration that moral hazard can be efficient makes a useful contribution to the health insurance literature.

[12] This application of an income-elasticity to analyze a relatively large income change follows Nyman (p. 83), who notes, "If the consumer in the above scenario had an endowed income of $50,000, then the $37,000 in income transfers would represent a 74% increase in income, using the endowed income as the base. With an income elasticity of 1, this would generate a 74% increase in health care spending."

[13] Both Pauly (1968) and Nyman (2003) model individual demand with a fixed marginal cost of medical care. In the marketplace, however, supply curves slope upward, so greater utilization of care involves higher costs per unit. Thus, when many insureds exhibit moral hazard, they collectively exert upward pressure on cost, and this ultimately widens the gap between marginal cost and willingness to pay. In that respect, analyses of individual behavior that assume constant costs are likely to underestimate the aggregate inefficiency caused by moral hazard.

[14] Evans and Viscusi (1991) found empirically that for minor illnesses, utility is state-independent; for severe illnesses, Viscusi and Evans (1990) found utility to be state-dependent.

[15] Printer's errors in the text make it appear that Nyman's Figure 4.2 compares $U(Y^0)$ against $U(Y^0 - \pi(1-\pi)(1-c)M^*)$ and $U(Y^0 - M^*)$ against $U(Y^0 - M^* + \pi(1-\pi)(1-c)M^*)$, though a careful reading indicates that such is not the author's intent.

[16] The effect of loading can be observed by replacing inside the utility function by $\lambda\pi$, where $\lambda = 1$ implies a fair premium and $\lambda > 1$ implies loading.

[17] In an earlier passage, Nyman (p. 11) explains why he focuses on the all-or-nothing case, writing, "In more sophisticated models, the consumer may choose the optimal level of insurance payoff or coinsurance rate. … In most cases, however, the consumer does not have sufficient leeway to choose the exact level of coverage he wants, so models that describe the optimal level of coverage have limited practical appeal."

[18] Part of the reason for this finding is that Nyman uses the objective probability of illness (π) to weight outcomes, whereas prospect theory holds that individuals subjectively over-weight low-probability events.

[19] The regret theory of Loomes and Sugden (1982) is not mentioned, though in some respects that approach is closer to Nyman's. According to regret theory, a consumer contemplates the unhappiness that making the wrong decision would yield. In the present context, this would imply that if she remains healthy, the consumer might regret having paid for insurance; if she becomes ill, she would regret having remained uninsured. The decision would then focus on avoiding the most regrettable outcome.

[20] Just as a concave utility function defined over wealth (or income) represents aversion to financial risk, a multivariate utility function that is concave in each argument represents risk aversion with respect to each argument. Thus, consumers may be risk averse in both wealth and health; for a discussion of multivariate risk aversion, see Kihlstrom and Mirman (1974).

[21] In Nyman's (p. 141) words, "consumers tend to prefer a risk of loss to an actuarially equivalent certain loss. …[I]f risk preferences enter the insurance choice in a decisive way, they would probably serve to override the natural preference for an income transfer when ill, explaining why consumers might not purchase insurance, rather than why they do."

REFERENCES

Arrow, Kenneth J. (1963) "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, 53(5) December: pp. 942–973.

Arrow, Kenneth J. (1971) *Essays in the Theory of Risk-Bearing*. Chicago: Markham.

Bernoulli, Daniel (1738) "Exposition of a New Theory on the Measurement of Risk," translated by L. Sommer, *Econometrica*, 22(1) January, 1954: 2, pp. 3–36.

Blomqvist, Ake (2001) "Does the Economics of Moral Hazard Need to be Revisited? A Comment on the Paper by John Nyman," *Journal of Health Economics*, 20(2) March: pp. 283–288.

de Meza, David (1983) "Health Insurance and the Demand for Medical Care," *Journal of Health Economics*, 2(1) March: pp. 47–54.

Eichner, Matthew J. (1998) "The Demand for Medical Care: What People Pay Does Matter," *American Economic Review*, 88(2) May: pp. 117–121.

Eisenhauer, Joseph G. (1996) "Insurance and Externalities: An Economic Analysis of Premium Taxes," *Journal of Insurance Issues*, 19(1) Spring: pp. 69–77.

Eisenhauer, Joseph G. (2002) "Relative Effects of Premium Loading and Tax Deductions on the Demand for Insurance," *Journal of Insurance Issues*, 25(1) Spring: pp. 47–62.

Eisenhauer, Joseph G. and Martin Halek (1999) "Prudence, Risk Aversion, and the Demand for Life Insurance," *Applied Economics Letters*, 6(4) April: pp. 239–242.

Evans, William N. and W. Kip Viscusi (1991) "Estimation of State-Dependent Utility Functions Using Survey Data," *Review of Economics and Statistics*, 73(1) February: pp. 94–104.

Feenberg, Daniel and Jonathan Skinner (1994) "The Risk and Duration of Catastrophic Health Care Expenditures," *Review of Economics and Statistics*, 76(4) November: pp. 633–647.

Friedman, Bernard (1973) "Risk Aversion and the Consumer Choice of Health Insurance Option," *Review of Economics and Statistics*, 56, May: pp. 209–214.

Fuchs, Victor (1996) "Economics, Values, and Health Care Reform," *American Economic Review*, 86(1) March: pp. 1–24.

Halek, Martin and Joseph G. Eisenhauer (2001) "Demography of Risk Aversion," *Journal of Risk and Insurance*, 68(1) March: pp. 1–24.

Kahneman, Daniel and Amos Tversky (1979) "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47(2) March: pp. 263–291.

Kihlstrom, Richard E. and Leonard J. Mirman (1974) "Risk Aversion with Many Commodities," *Journal of Economic Theory*, 8(3) July: pp. 361–388.

Loomes, Graham and Robert Sugden (1982) "Regret Theory: An Alternative Theory of Rational Choice under Uncertainty," *Economic Journal*, 92(368) December: pp. 805–824.

Manning, Willard G. and M. Susan Marquis (1996) "Health Insurance: The Tradeoff Between Risk Pooling and Moral Hazard," *Journal of Health Economics*, 15(5) October: pp. 609–639.

Newhouse, Joseph P. (1993) *Free for All? Lessons from the Rand Health Insurance Experiment*. Cambridge MA: Harvard University Press.

Nyman, John A. (1999) "The Economics of Moral Hazard Revisited," *Journal of Health Economics*, 18(6) December: pp. 811-824.

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

Nyman, John A. and Roland Maude-Griffin (2001) "The Welfare Economics of Moral Hazard," *International Journal of Health Care Finance and Economics*, 1(1) March: pp. 23–42.

Pauly, Mark V. (1968) "The Economics of Moral Hazard: Comment," *American Economic Review*, 58(3) June: pp. 531–537.

Pauly, Mark V. (1983) "More on Moral Hazard," *Journal of Health Economics*, 2(1) March: pp. 81–85.

Rice, Thomas (1992) "An Alternative Framework for Evaluating Welfare Losses in the Health Care Market," *Journal of Health Economics*, 11(1) May: pp. 85–92.

Szpiro, George G. (1986a) "Measuring Risk Aversion: An Alternative Approach," *Review of Economics and Statistics*, 68(1) February: pp. 156–159.

Szpiro, George G. (1986b) "Relative Risk Aversion Around the World," *Economics Letters*, 20(2): pp. 19–21.

Viscusi, W. Kip and William N. Evans (1990) "Utility Functions that Depend on Health Status: Estimates and Economic Implications," *American Economic Review*, 80(3) June: pp. 353–374.

Werner, Jan (2005) "A Simple Axiomatization of Risk-Averse Expected Utility," *Economics Letters*, 88: pp. 73–77.