Evidence of Adverse Selection in the Individual Health Insurance Market

Mark J. Browne*

ABSTRACT

In this article the author tests the market for individual health insurance to determine if adverse selection is present. Evidence is found that low risk consumers do purchase less insurance in the individual market than they do in the group market. This finding is consistent with adverse selection existing in the individual market for insurance. However, the finding may also be explained by low risks being off their demand curve in the group market.

A second finding of the study is that equilibrium in the individual market is characterized by a subsidization of high risk consumers' insurance purchases by low risks. This is consistent with the pooling equilibrium models of Wilson and Miyazaki but not the separating equilibrium model of Rothschild and Stiglitz.

The number of people without health insurance in the United States is estimated to be in excess of 36.8 million.[1] Since the possibility of the need for high cost health care faces everyone, the uninsured population stands at risk of, at best, significant financial setback should the need for expensive health care treatment become necessary. At worst, the uninsured may be unable to acquire access to an increasingly expensive but at times life-giving health care delivery system.

Numerous reasons have been advanced to explain why people do not have medical insurance. The reasons include the risk-taking nature of some individuals, financial inability of some to afford medical insurance, ignorance of the risks covered by medical insurance, as well as adverse selection. Dahlby (1983) found that adverse selection leads to reduced insurance consumption by low risks in the automobile insurance market. Beliveau (1981) tested for adverse selection in the life insurance market and similarly found that adverse selection leads to reduced insurance consump-

* Mark J. Browne is Assistant Professor of Risk Management and Insurance at The University of Georgia.

The author gratefully acknowledges the helpful comments of Patricia Danzon, J. David Cummins, Mark Pauly, Howard Kunreuther, Jerry Rosenbloom, the participants at the 1990 Risk Theory Seminar, the Editor, and two anonymous referees. Financial support was provided by the Huebner Foundation.

[1] *Source Book of Health Insurance Data: 1989*

tion by low risks. Conrad, Grembowski, and Milgrom (1985) found evidence that adverse selection occurs in the market for dental insurance. Adverse selection is "the tendency of high risks to be more likely to buy insurance or to buy larger amounts than low risks" (Cummins, et al., 1982). This results from an asymmetry in market information in favor of the buyer of insurance.

The purpose of this article is to test for adverse selection in the individual medical insurance market and to determine whether the equilibrium in the insurance market is a separating or pooling equilibrium if adverse selection is present in the market. The major theoretical models of equilibrium in a market with adverse selection are reviewed and are followed by a test for adverse selection in the individual health insurance market. Then, the nature of the equilibrium in the market is examined. The paper concludes with a discussion of the major findings of the study.

Equilibrium Models Of A Market With Adverse Selection

Seminal work on asymmetric market information was done by Akerlof (1970). He offered a theoretical reason for the lack of insurance owned by large numbers of the elderly prior to Medicare. His justification was that insurers were unable to distinguish between high and low risk applicants for insurance. If an insurer (unable to distinguish between the two groups) prices its policy for the elderly at an average price for all elderly individuals, only those who are worse than average risks will buy insurance. Since this will result in losses for the insurer that had anticipated the better than average risks also purchasing the policy, prices will be raised. Of the group which purchased insurance previously only the worse than average risks will purchase insurance again at the higher price. Prices will again need to be raised and the process will continue. Eventually only the very worst risks will purchase insurance at extremely high prices or the whole market for insurance will collapse.

One model of equilibrium in an insurance market in which asymmetric information exists between the insuring parties was proposed by Rothschild and Stiglitz (1976). In their model the existence of asymmetric information results in an unusual competitive equilibrium or in no equilibrium at all. They assume risk-averse consumers of insurance can be divided into two classes, low risks and high risks. As in all of the equilibrium models to be discussed, low risk and high risks have the same underlying utility function but differ in their probability of suffering a loss. An individual insurance seeker knows to which group he or she belongs. However, the profit-maximizing insurer can not differentiate between high and low risks. The market is competitive with free entry so that all insurance contracts which are expected to make a profit are offered. Therefore, all contracts are supplied for which a demand exists and for which a profit is expected to be made. Equilibrium is of the Nash type which means no contract in the equilibrium set makes negative expected profits and there is no contract outside the equilibrium which will make a non-negative profit. Finally, each

firm assumes that other firms offer contracts independent of its actions. In the Rothschild-Stiglitz model equilibrium is characterized by the low risks and high risks purchasing different insurance policies. The high risks purchase a policy which provides complete insurance protection whereas the low risks purchase a policy that provides less than complete insurance coverage. Each policy is priced actuarially fair and each breaks even. Wilson (1977) introduced an equilibrium concept known as Wilson E2 equilibrium. It contains the assumptions that firms consider the consequences of their actions on the behavior of other firms when they introduce a new policy and firms offer only policies which make non-negative expected profits. Under these assumptions the resulting equilibrium is either a separating equilibrium as in the Rothschild-Stiglitz model or a pooling equilibrium involving a subsidy of the high risks insurance coverage by the low risks. Which type of equilibrium will evolve depends upon the level of risk aversion of the low risk individuals.

Unlike the Rothschild-Stiglitz model the Wilson model assumes firms have the necessary foresight to withhold offering contracts that will be profitable in the short term but that will become unprofitable when other policies are removed from the market. This assumption makes it possible for a pooling equilibrium to occur in a competitive insurance market–a result not obtained by Rothschild and Stiglitz. Intuitively, a pooling equilibrium will result when the costs of cross subsidization are less than the transaction and information costs of achieving a separating equilibrium and the gain in utility for the low risks derived through risk reduction exceeds the costs of cross subsidization. Miyazaki's (1977) model of market equilibrium with adverse selection allows insurers to incur losses on individual policies as long as they break even on their total book of business. Under the assumptions of the Miyazaki model, market equilibrium will be a separating equilibrium involving a low risk to high risk wealth transfer with low risks buying one policy and high risks another. Insurers will incur losses on policies sold to high risks but earn profits on policies sold to low risks.

In each of the three models of equilibrium previously discussed the low risk individuals purchase less insurance than they would in a market free of adverse selection. The models differ in regards to risk class subsidization. The Rothschild-Stiglitz model entails a separating equilibrium with low risks and high risks buying different insurance policies both of which break even individually. The Wilson and Miyazaki models differ from the Rothschild-Stiglitz model in that there is an across risk class subsidization in both models. In the Wilson model both high and low risks purchase the same policy for the same price. A low risk to high risk subsidization results as the high risks have a greater incidence of loss. In the Miyazaki model high risks and low risks purchase different policies. The policies purchased by high risks generate losses for the insurer while the policies purchased by the low risks generate profits. Therefore, the high risk to low risk subsidization in the Miyazaki model is across policies.

Although each of the theories of market equilibrium in a market with asymmetric information differ, there are similarities between them. Evidence of certain peculiarities in the market equilibrium does not necessarily prove the existence of a specific type of equilibrium. However, such information may indicate an asymmetry of information in the market.

The adverse selection theoretical models lead to two major testable hypotheses. They are:

Hypothesis One: In an insurance market characterized by asymmetric information low risks will purchase less insurance than in a market free of adverse selection.

Hypothesis Two: In an insurance market with asymmetric information there will be a low risk to high risk wealth transfer. The Rothschild-Stiglitz definition of equilibrium precludes a subsidy. Those of Wilson and Miyazaki predict a subsidy.

Adverse Selection in the Individual Health Insurance Market

The first hypothesis of this study, that adverse selection is present in the market for individual health insurance, will be tested by comparing the actual amount of insurance purchased by low-risk families in the individual market with an amount that is predicted they would have purchased had they obtained their insurance in the group market. A lower consumption of insurance in the individual market than in the group market by low risks is consistent with the hypothesis that low risks purchase less insurance in a market with adverse selection than in a market with less adverse selection.

The group health insurance market is believed to suffer less from adverse selection than the individual market. In the group market people typically receive their health insurance through employment. Because group health insurance coverage is in many cases largely an employer paid employee benefit, the low risks can leave the pool only by foregoing a very large subsidy of their insurance purchase or by terminating employment. Since health insurance coverage is incidental to employment in most situations, adverse selection should be comparatively less in the group market than in the individual market.[2]

Because of the constraints on choice in the group market, alternative explanations cannot be ruled out for low risks consuming less insurance in the non-group market than in the group market. Due to constrained choice in the group market, there is a possibility that heterogeneous health insurance groups may form. This may entail low risks purchasing more insurance than they would otherwise choose and a subsidy of the high risks in the group by the low risks. If homogenous groups formed low risks would receive the amount of insurance they most preferred and there would be no cross risk subsidization within the insurance purchasing group.

---

[2] Nearly 90 percent of employees in 1977 worked in firms which offered group health insurance as an employee benefit. Employment related groups account for the vast majority of private health insurance policies sold. See Cafferata (1983).

Table 1
Comparison of Policy Provisions
In The Group and Individual Markets

Group Market Policy Provisions	*Maximum*	*Q3*	*Q2*	*Q1*	*Minimum*
Major Medical Max Benefit (N=10,616)	2,000,000	250,000	250,000	50,000	50
Max Benefit Dr. Office Visits (N=501)	250,000	500	350	250	35
Deductible Dr Office Visits (N=11,914)	5,100	100	100	50	0
Max. Benefit Hospital Room and Board (N=327)	250,000	7,941	2,647	1,059	80
Hospital Deductible (N=13,503)	8,100	0	0	0	0
Max Benefit Outpatient Diagnostic Services (N=4,615)	250,000	200	100	100	10
Individual Market Policy Provisions					
Major Medical Max Benefit (N=500)	1,000,000	250,000	25,000	10,000	10
Max. Benefit Dr. Office Visits (N=43)	10,350	1,200	350	250	100
Deductible Dr. Office Visits (N=559)	10,000	250	100	0	0
Max Benefit Hospital Room and Board (N=46)	20,000	10,000	2,647	2,647	476
Hospital Deductible (N=1,310)	1,500	0	0	0	0
Max. Benefit Outpatient Diagnostic Services (N=311)	25,000	150	100	50	8

Variations in N for the different policy provisions result from different policies having different provisions. Data are from the National Medical Care Expenditures Survey.

In both the group and individual markets a variety of different policies were sold at the time the data for this study were collected. Table 1 contains a comparison of group and individual policy provisions. The table illustrates that in both markets policies offering extensive coverage and policies offering comparatively little coverage were available. The availability of policies offering various amounts of insurance in the market is consistent with but not proof of adverse selection. The table also shows that the maximum deductibles purchased in the non-group market exceed those of the group market. The maximum limits purchased in the group market exceed those of the non-group market.

The data used in this study come from the National Medical Care Expenditure Survey. Over 14,000 randomly chosen households participated. The data set includes information on demographic characteristics, medical expenses and medical usage of individuals. Insurance coverage information including premiums and policy characteristics is also included in the data set. However, the National Medical Care Expenditure Survey did not collect insurance information from all households in the study. Further, demographic and loss experience information was not complete for all households. Consequently, the subsample used in this study consisted of 2,515 families in the group market and 225 families in the individual market. The National Medical Care Expenditure Survey was conducted during 1977 and 1978.

The test for determining if low risks consume less insurance in the individual market than they would in a market with less adverse selection involves a comparison of the insurance coverage actually purchased by low risks in the non-group market with an amount that is predicted they would have purchased had they purchased their insurance coverage in the group

market. As previously discussed adverse selection is believed to be less in the group health insurance market than in the individual market.

Testing the hypothesis involves several steps. First, a measure of insurance coverage is defined. Second, an insurance prediction equation is estimated for families with group coverage. Third, the group market health insurance prediction equation is used to predict the amount of insurance individual market consumers would have purchased had they acquired insurance in the group market. Fourth, the discrepancy between insurance actually purchased in the individual market and predicted purchases is analyzed. The difference is decomposed into various factors which may account for the discrepancy. A discrepancy between predicted and actual purchases that is greater for low risks than for high risks is consistent with the hypothesis of reduced insurance consumption by low risks in a market with adverse selection.

Development of A Measure of Insurance Coverage

In order to determine the quantity of insurance held by different risk groups and in different markets, it is first necessary to develop a measure of insurance. Insurance has special characteristics which make it difficult to quantify. Insurance is purchased through an aleatory contract. The payoff from the policy for the insured is uncertain and subject to events in the future which may or may not occur. The probability that an insurance contract will provide a loss benefit of a certain amount to an insured will depend upon the risk characteristics of the insured as well as the provisions of the insurance contract.

The quantity of insurance which one holds will be a function of two elements. The first element is the design of the insurance policy, including the deductible, coinsurance rate, and limits of coverage. The amount of insurance can thus be increased or decreased through the level of the deductible, the amount of coinsurance, the limits of coverage, and the stipulated maximum out-of pocket payment provision.

The second element consists of the risk characteristics of the insured. If an insurer sells identical policies to two individuals which it believes to be similar risks, the insured who in fact is a greater risk will receive more insurance. This is so because the insurer is assuming more risk from the individual who is a high risk than from the individual who is a low risk.

The empirical definition of the quantity of insurance must take into account these two elements. The measure used in this study is the expected indemnity benefit from a policy for a given family. This will be referred to as I.

Mathematically I is defined:

$$I_i = E[\text{Benefit}]_{i1} + \ldots\ldots + E[\text{Benefit}]_{i6}$$

$$E[\text{Benefit}]_{ij} = E[\text{MED}]_{ij} - E[\text{OOP}]_{ij} \quad i = 1,\ldots,N \quad j = 1,\ldots,6$$

where each i represents a family and each j represents a medical expense item. The medical expense items include outpatient physician fees, inpatient surgeon fees, inpatient physician (non-surgeon) fees, hospital room and board charges, fees for diagnostic tests, and prescription medicine expenses.

$E[MED]_{ij}$ is the expected medical expense family i incurs for expense item j. $E[MED]_{ij}$ is based upon a prediction model which will be discussed in the next section. $E[OOP]_{ij}$ is the expected amount of money family i will have to pay out-of-pocket for medical expense item j. This is determined by applying the medical expense policy provisions of family i's insurance to the predicted medical expense $E[MED]_{ij}$.[3]

This measure of insurance is appropriate for the study of adverse selection as it represents the indemnity benefits an insured family can expect to derive from its insurance policy. Adverse selection results from the difference in indemnity benefit expectation of the insurer and the insured.[4]

Estimating Expected Expenses

A procedure developed by Duan, et al.(1983) is modified to predict medical expenses. Expenses are predicted for each person in the sample for each of the types of medical services listed earlier. The number of medical services in the analysis was limited by the availability of medical insurance data in the National Medical Care Expenditures Survey. The survey provided extensive insurance coverage information for the medical services used in the study but not for all medical services.

The prediction procedure of Duan, et al. involves developing equations that predict the probability that a medical service of a certain type will be purchased and equations that predict the dollar cost of a medical care purchase given that one is made. The prediction procedure accounts for the thick right tail of the medical expense distribution and for the fact that many individuals have no medical expense during the year. Prediction variables include age, sex, marital status, income, family size, occupation, race, education, region of the country in which the insured resides, the population of the town in which the individual resides, perceived health status, and degree of activity limitation if any. Group size and the employer's share of the health insurance premium are also included as explanatory variables.[5]

---

[3] To illustrate the calculation of the measure of insurance, assume that a family has a predicted prescription medical expense of $500 and a predicted physician fee expense of $400. If the insurance policy held by the family covers the physician fee expense with no cost sharing provisions and covers the prescription medicine expense subject to a deductible of $100 and a coinsurance percentage of 80 up to a maximum out-of-pocket payment of $1,000, then the measure of insurance would by $720 ($400 + 80%($500 - $100)).

[4] Although a more exact definition of insurance would take into account each insured's distribution of medical expenses, the data used in this study did not allow for that refinement.

[5] A similar model was estimated separately for children in the sample. The explanatory variables in the model included age, sex, log of family size, and log of family income.

The expected expenditure on medical care item j for an individual i with characteristics X_i is:

$$E(MED_{ij}|X_i) = P_{ij} * \phi_j * \exp(X_i\delta_{j2}),$$

where, P_{ij} is the probability that insured i will incur a medical expense of type j. The product $[\phi_j * \exp(X_i\delta_{j2})]$ is the expected size of medical expense item j given that this type of expense is incurred.

To determine P_{ij} the following equation is estimated:

$$M_{ij} = X_i\delta_{j1} + \eta_{1ij},$$

where M_{ij} is a dichotomous variable which takes the value of 1 if insured i had a medical expense of type j and 0 otherwise. X_i is the vector of explanatory variables. δ_{j1} is the corresponding vector of parameter estimates. η_{1ij} is the error term of the equation. P_{ij} is expressed:

$$P_{ij} = Pr(MED_{ij}>0|X_i) = 1/(1 + e^{-X\delta_{ij}}).$$

The size of a medical expense given that one is incurred is estimated with the equation:

$$\log(MED_{ij}|M_{ij}>0) = X_i\delta_{j2} + \eta_{2ij},\ \eta_{2ij} \text{ dist. } N(0,\sigma^2_{\eta 2j}),$$

where MED_{ij} is the medical expense of type j incurred by insured i. The log transform of the variable MED_{ij} is made as medical expenses are best approximated by a lognormal distribution. X_i is a vector of explanatory variables. δ_{j2} is the corresponding vector of parameter estimates. η_{2ij} is the error term of the equation. The log transformation necessitates the use of a retransformation term to predict absolute dollar expenses.[6] The retransformation term is written:

$\phi_j = \exp(\sigma^2_{\eta 2j}/2)$.

---

[6] Note that the equation that predicts the size of expenditure item j for individual i used a logarithmic transformation of actual purchases of item j as the dependent variable. Estimation of predicted values from this equation requires retransformation of the error term of the equation.

$$\log(MED_{ij}|M_{ij}>0) = X_i\delta_{j2} + \eta_{2ij}$$
$$MED_{ij} = \exp(X_i\delta_{j2}) * \exp(\eta_{2ij})$$
$$E(MED_{ij}) = E[\exp(X_i\delta_{j2})] * E[\exp(\eta_{2ij})]$$

Since $E[\exp(X_i\delta_{j2})] = \exp(X_i\delta_{j2})$
and $E[\exp(\eta_{2ij})] = \exp(\sigma^2_{\eta 2j}/2)$

$$E(MED_{ij}|M_{ij}>0) = \exp(X_j\delta_{j2}) * \exp(\sigma^2_{\eta 2j}/2)$$

The retransformation term is accurate provided the error terms of the log regression equation are normally distributed. In the Rand Health Insurance Experiment discussed by Duan, et al. this methodology was used to predict health care purchases. That study found that the error terms were in fact not distributed normally. If the error terms are not normally distributed, inconsistent estimates of the expected expenses will result. However for this model the normal retransformation was found by Duan, et al. to be superior to other workable techniques. Because it is the best workable retransformation technique it is used in this study.

The Structural Model of The Demand For Insurance and Health Care

The structural model used for prediction of medical expenses for insured i is composed of two equations, a medical expense demand equation and a medical insurance demand equation. A two equation system is appropriate as simultaneity has been found to exist in the demand for medical services and insurance (Manning, 1988). The structural equations are:

$$MED_{ij} = A_1X_{1i} + A_2I_i + e_{1i} \quad (1)$$
$$I_i = B_1X_{1i} + B_2X_{2i} + B_3E[MED_{ij}] + e_{2i} \quad (2)$$

MED_{ij} is the medical care expenditure of insured i on medical expense item j. I_i is the amount of health insurance held by family i, X_{1i} is a vector of explanatory demographic characteristics, and X_{2i} is a vector of firm characteristics including group size and the fraction of the premium paid by the employer.

Only the first of the two structural equations in the model is identifiable. It is reasonable to assume that there is no exogenous variable which would affect the demand for medical care and not also affect the demand for medical insurance. Therefore, Equation (2) is not able to be identified. If equation (2) were identifiable coefficient B_3 would indicate the impact of adverse selection on the insurance purchase decision. A_2 would indicate the impact of moral hazard on the purchase of medical care. Since both equations are not identifiable, two-stage least squares estimation is not appropriate for estimating the model.

Two alternative approaches to estimating the medical expenditure and medical expense insurance equations are possible. First the equations may be assumed to be independent of one another. That is:

$$MED_i = A_1X_{1i} + e_{3i} \quad (3)$$
$$I_i = B_1X_{1i} + B_2X_{2i} + e_{4i} \quad (4)$$

Previous studies have shown that this assumption does not hold. The predictions of medical expense and insurance coverage that are made based on this model will be inaccurate to the extent that moral hazard and adverse selection do exist.

The second approach involves the estimation of the reduced form equations of the model represented by equations (1) and (2).[7] This approach accounts for the impact of moral hazard on medical expenses and consequently insurance coverage which the first approach did not. Since previous studies have shown that moral hazard is an important determinant of medical expense this approach is superior to the first presented and is used here. Since the reduced form equations are specified to predict medical expenses and insurance coverage, statistical inferences should not be drawn from them.

---

[7] The reduced form equations are available from the author.

In this study the medical expense prediction equation consists of the two-part model discussed earlier. The exogenous variables used in both parts of the model and for each of the different medical expenses are those from the reduced form equation.

Estimating Expected Benefits

The National Medical Care Expenditure Survey data set contains detailed information about the medical expense insurance policy provisions of many of the people who took part in the survey. Data on insurance coverage were not collected for all of the households in the survey. As previously discussed this information is used in conjunction with the predicted medical expenses of each individual to determine the expected benefits which each person's health insurance policy will provide. The data set reports the insurance policy provisions corresponding to the insurance coverage of each individual as opposed to each family. The expected indemnity benefits of family members are summed to determine a measure of insurance coverage for a family.

The data allow for a computation of what the insurance policies would pay for various size losses of different types. Without having the true policies it is difficult to know exactly how to determine benefit payments. However, enough information is available to make a reasonable estimate of what the policies would pay. Of particular concern are provisions which may exist but on which the survey did not collect information. Basic policy provisions common to most policies were recorded. Unusual and unique policy provisions which would not have conformed to the reporting format are not included in the data set.

Insurance coverage was determined by applying the deductibles and coinsurance rates reported for various services to the estimated expenditures for each service by each insured.[8] In making this computation it was assumed that no claims were previously applied against the policy. The computation procedure also made use of the various limits reported as well as the major medical provisions of an insured's insurance coverage if applicable. Because policy provisions call for benefits to be determined in different ways computation of expected benefits was done accordingly. For example, because some policies have such features as per day reimbursement limits or per visit reimbursement limits and others have limits in terms of total cost, the benefit determination had to be done taking into account the unique provisions in the insurance policy of each insured.

---

[8] Since a medical expense distribution for the individual insureds was not able to be constructed, the measure of insurance is based on the expected mean expenses of each insured. To the extent that insured's have loss distributions which differ from one another in higher moments the measure of insurance will suffer. For instance, the procedure does not distinguish a difference in insurance coverage between an individual with a 50 percent chance of no loss and a 50 percent chance of a $1,000 loss and an individual with a 100 percent chance of a $500 loss.

The amount of health insurance held by family i is therefore a function of the expected medical expenditure for each item of medical expense by each family member and the provisions of the insurance policy held by each family member. The computation of insurance coverage is not affected by whether or not family members have different insurance policies with different provisions. To determine the amount of insurance a family has the amounts of insurance held by each individual in the family are summed. If a family has only one member who has individual coverage then the measure of insurance for that family is equal to the amount of insurance held by that individual. Use of the policy provisions in determining the level of insurance coverage allows for insureds with identical predicted health care expenses to have differing amounts of insurance to the extent that their insurance policy provisions differ. This would be likely as there is a wide variance in insurance coverage among the insureds in the data set.

Note that if a type of medical expense is not covered by insurance, the item is still included in the calculation of insurance coverage or lack thereof. In that situation for that service the insured would pay the total cost and the insurance benefit would be zero. Although the data set on the medical expense insurance policy provisions has several minor drawbacks, this unique information allowed for the reasonably precise measure of insurance used in this study.

The Group Market Insurance Prediction Equation

With a measure of insurance developed, the group market insurance prediction equation can be specified. The sample used to estimate the model consists of those families that acquired health insurance coverage in the group market. This market is assumed to be relatively free of adverse selection. The insurance prediction equation that is estimated will be used to predict the insurance purchases that participants in the individual market would make in a market relatively free of adverse selection.

The insurance prediction equation is written:

$$I_i = \beta_g Y_{gi} + e_i$$

where:

I_i is the amount of insurance that insured family i holds in the group health insurance market, Y_{gi} is a vector of characteristics pertaining to insured family i of the group market which are hypothesized to effect the demand for medical insurance, β_g is a vector of coefficient estimates corresponding to the characteristics Y_{gi}, and e_i is a random error term.

Table 2 reports the estimated health insurance prediction equation.

The parameters of the estimated health insurance prediction equation are used to predict insurance purchases by members of the individual health insurance market:

$$I_{pi} = \beta_g Y_{ii} + e_{pi}$$
$$E[I_{pi}] = \beta_g Y_{ii}$$

where:
I_{pi} is the predicted insurance purchase of family i in the individual market based on estimated group market parameters, β_g is the vector of parameters estimated with the group market health insurance demand equation, and Y_{ii} is a vector of characteristics pertaining to insured family i of the individual market which are hypothesized to effect the demand for insurance.

Table 2
Group Market Health Insurance Prediction Equation
Dependent Variable: I_i
(R^2=.54, N= 2,515)

Variable	*Coefficient*	*(t-stat)*
INTERCEPT	- 625.57	(-5.88)
EDUCAT (Years of education)	-0.37	(-0.14)
LFAMINC (Log of household income)	24.38	(2.31)
LFAMSIZE (Log of family size)	31.88	(2.27)
AGE	10.47	(18.51)
SEX (1 if Female)	-13.99	(-0.60)
MARITAL STATUS (1 if Married)	110.43	(5.30)
AREA2 (SMSA>500,000, but not largest 16)	-50.51	(-2.58)
AREA3 (SMSA<500,000)	-10.69	(-0.54)
AREA4 (Not SMSA, largely urban)	-33.08	(-1.59)
AREA5 (Primarily rural)	-40.89	(-1.70)
BLACK (Black)	69.41	(2.95)
HISP (Hispanic)	-116.93	(-3.24)
OTHER (Not white, black or Hispanic)	176.94	(3.32)
LIMMAJ (Severe activity limitation)	1184.28	(19.72)
LIMAMT (Less severe limitation)	710.06	(19.18)
LIMMIN (Activity limitation is minimal)	149.96	(3.68)
PHSGOOD (Reported health status, good)	92.28	(6.82)
PHSFAIR (Reported health status, fair)	301.59	(13.05)
PHSPOOR (Reported health status, poor)	611.47	(12.35)
HOPER (Occupation is operator)	88.99	(3.40)
HTRANS (Occupation is transportation)	50.48	(1.56)
HSERV (Occupation is service)	122.20	(4.53)
HLABOR (Occupation is labor)	43.36	(1.27)
HFLABOR (Occupation is farm labor)	169.72	(1.76)
HFOWN (Occupation is farm owner)	-282.27	(-3.69)
HUNK (Occupation is unknown)	10.16	(0.38)
HMGR (Occupation is manager)	69.58	(2.88)
HSALES (Occupation is sales)	14.19	(0.43)
HCLERK (Occupation is clerk)	53.58	(1.80)
HCRAFTS (Occupation is crafts)	90.85	(3.75)
NEWENG (Region is New England)	267.40	(7.70)
MIDATL (Region is Mid-Atlantic)	155.50	(6.28)
ENCENT (Region is East, North Central)	133.86	(5.89)
WNCENT (Region is West, North Central)	111.66	(3.96)
SSATL (Region is South, South Atlantic)	59.52	(2.49)
SEAST (Region is Southeast)	-66.33	(-2.00)
SWEST (Region is Southwest)	-51.13	(-1.72)
MNT (Region is Mountain)	-48.42	(-1.30)
GROUPSZP (Group Size)	0.000388	(1.86)
ESHARE (Employer paid share of premium)	62.31	(3.01)

Omitted Control Variables
MALE, NOT MARRIED, AREA1 (largest 16 SMSAs), WHITE, LIMNONE (No activity limitation), PHSEXC (Reported health status, excellent), HPROF (Occupation is professional), PACIFIC (Region is Pacific).

An important assumption of this methodology is that families in the group market obtain insurance coverage suitable to their tastes. The assumption requires that employers choose health insurance policies consistent with the tastes of their employees and that employees are attracted to firms that offer a compensation package they desire. Among the reasons typically justifying employee benefit plans are the employers' concern for employees' welfare, the corporate need to attract and obtain capable employees and the demands of labor groups. Cost minimization requires that firms attract employees at the lowest possible cost through appropriate design of the benefits package. These reasons suggest that employers will attempt to provide health insurance coverage consistent with employees' tastes. Because of the likelihood that most firms employ workers with heterogeneous tastes for health insurance, a firm's group insurance policy is likely not optimal for all workers. Goldstein and Pauly (1976) discuss employee benefit determination when firms have heterogeneous workers. To the extent that insureds who obtain coverage in the group market are off their demand curves, the conclusions of this analysis must be modified.

The Demand Discrepancy Decomposition

The hypothesis that adverse selection exists in the individual market for health insurance is tested by comparing actual insurance purchases by low risks with predicted insurance purchases. As discussed previously the predicted purchases are derived from the demand equation specified with group market data. Adverse selection is indicated if actual insurance consumption by low risks is less than predicted consumption.

The hypothesis that low risks purchase less insurance in a market with adverse selection than in a market with less adverse selection is tested by comparing predicted purchases in the market with little if any adverse selection to actual insurance purchases in the market believed to have adverse selection. The test for this relationship is:

$$D_i = (\text{Predicted}_i - \text{Actual}_i) = a_1 LR_i + a_2 X_i + e_i.$$

D_i is the difference between the predicted insurance purchase and the actual insurance purchase of family i of the individual market.

The National Medical Care Expenditures Survey asked study participants if they perceived their health status as excellent, good, fair, or poor. LR_i is a binary variable which takes on the value of 1 if family i is a low risk. A low-risk family is defined as one whose head of household reported a perceived health status of excellent, good, or fair in the survey. High-risks are those who reported their perceived health status as poor. Insurance companies do not generally collect information on perceived health status for use in developing their estimates of losses they will have to pay. To the extent that individuals do use their perceived health status in forming estimates of their anticipated medical care consumption there is an

information asymmetry in the health insurance market which may result in adverse selection.

The variable LOW RISK is the basis for testing for adverse selection in the individual medical expense insurance market. If adverse selection is a more severe problem in the individual market than in the group market, low risks will purchase less insurance in the individual market than would have been predicted by the group market demand equation. Each of the earlier mentioned models of equilibrium in a market with adverse selection predicts this.

X_i is a vector of other variables pertaining to family i which may explain the discrepancy between predicted insurance purchases and actual insurance purchases. Several of these variables are observable by insurers and are used in classifying insureds. These include age, sex, family size, and marital status. Since the expense loading in an insurance policy is in part multiplicative[9], factors which increase the rate charged to an insured will also increase the expense loading. The expense loading in the group market is averaged over all members of a group. However, in the individual market the expense loading will be a function of the individual insured's expected loss. Therefore, it is expected that risk factors, through their impact on the expense loading in the premium, will explain some of the variation between insurance consumption in the individual and group markets.

Income level may also explain some of the discrepancy between predicted and actual consumption because there are tax advantages to purchasing health insurance in the group market which are not mirrored in the individual market. The tax advantages inherent in purchasing group health insurance increase with higher levels of income. Due to the tax advantages available in the group market, actual insurance purchases in the individual market should be lower than predicted insurance purchases in the group market. This difference should increase with income. A positive correlation between income and the demand discrepancy is therefore expected.

Prohibitions on the use of certain types of data in classifying and rating insureds create asymmetries of information in the market for health insurance. Race of the insured is one factor which insurers are often prohibited from using. The inclusion of race classification variables in the discrepancy decomposition accounts for this asymmetry. The variables classify insureds as white, black, Hispanic, or other. The final category includes Native Americans and Orientals. In some states sex and marital status are also prohibited variables. As previously discussed these variables are used for rating and underwriting in states which do not have regulatory

---

[9] Multiplicative loading is a method of insurance pricing in which insurer expenses are determined as the product of the insured's expected loss and an expense loading factor. Multiplicative loading reflects the fact that certain expenses incurred by the insurer are proportional to the loss size. These expenses include sales commissions, premium taxes, and loss adjustment expenses. Bluhm (1988) provides a detailed discussion of health insurance premium calculation.

prohibitions against their use. A shortcoming of the National Medical Care Expenditures Survey data set is that families are not identified by state. Therefore, state dummy variables and demographic classification variables can not be interacted.

The activity limitation variables discussed earlier were also included as explanatory variables in the demand discrepancy decomposition equation. Activity limitations are observable and could therefore be used by insurers both in underwriting and rating where not prohibited by law. Since underwriting in the group insurance market is typically less stringent than in the individual market, people with activity limitations may not be able to acquire insurance in the individual market even though they may be able to do so in the group market. The activity limitation variables may therefore account for some portion of the demand discrepancy between the two markets.

Finally a group of occupation classification variables are included in the decomposition equation. The loading fee on group insurance policies decreases as group size increases. The expense saving is achieved through lower marketing and administrative expenses. Since the optimal firm size varies across industries, workers in different industries will be in different sized groups. The relative additional expense of acquiring insurance in the individual market as opposed to the group market will therefore vary across industries. Occupation may therefore explain part of the discrepancy between predicted and actual insurance consumption.

Table 3
Actual and Predicted Insurance Purchases
of Low-risk and High-risk Families in the Individual Market

	MEAN	*S.D*	*MAX*	*Q3*	*Q2*	*Q1*	*Min*
			High Risks (N=9)				
Actual Insurance	2077.51	1663.66	5133.61	3331.68	1339.44	642.83	417.01
Predicted Insurance	1598.94	655.56	2404.18	2218.72	1693.69	913.81	835.69
			Low Risks (N=216)				
Actual Insurance	357.38	414.14	2661.32	484.65	223.37	92.70	0.00*
Predicted Insurance	380.93	417.90	2016.22	525.54	360.83	99.19	0.00*

Several families were predicted to have negative amounts of insurance. As discussed in the article, if a family's predicted level of insurance was negative it was set equal to zero.

Table 3 provides a comparison of the actual insurance purchases of those who bought their insurance in the individual market with the amount that is predicted they would have purchased had they obtained their insurance in the group market.[10] The insurance prediction equation specified earlier

[10] The insurance prediction equation predicted that some members of the individual market would purchase negative insurance if they acquired their insurance in the group market. In these cases the amount of insurance that was predicted would be purchased was set equal to

Table 4
Medical Expense Insurance Prediction Equation Independent Variables
Group and Individual Market Means

Variable	Definition	Group Mean (N=2,515)	Individual Mean (N=225)
EDUCAT	Years of education	12.33	12.28
INCOME	Family income	21,765	22,219
AGE	Years of age	40.17	44.57
SEX	1 if female	.164	.276
MARITAL STATUS	1 if married	.719	.587
AREA2	1 if in Area2	.256	.240
AREA3	1 if in Area3	.198	.169
AREA4	1 if in Area4	.210	.196
AREA5	1 if in Area5	.126	.196
BLACK	1 if black	.089	.027
HISP	1 if Hispanic	.033	.013
OTHER	1 if other race	.014	.004
HUNK	1 if occupation unknown	.095	.240
HMGR	1 if manager	.108	.067
HSALES	1 if in sales	.047	.076
HCLERK	1 if a clerk	.071	.040
HCRAFT	1 if a craftmen	.160	.116
HOPER	1 if an operator	.121	.044
HTRANS	1 if in transportation	.057	.036
HSERV	1 if service worker	.096	.116
HLABOR	1 if a laborer	.050	.031
HFOWN	1 if farmowner	.007	.089
HFLABOR	1 if farmlaborer	.004	.009
PHSGOOD	1 if reported good health	.396	.364
PHSFAIR	1 if reported fair health	.097	.098
PHSPOOR	1 if reported poor health	.018	.040
LIMMAJ	1 if major limitation	.012	.044
LIMAMT	1 if amount limitation	.031	.067
LIMMIN	1 if minor limitation	.024	.009
NEWENG	1 if in corresponding	.047	.049
MIDATL	1970 Census region	.144	.258
ENCENT	"	.231	.222
WNCENT	"	.089	.080
SSATL	"	.208	.151
SEAST	"	.055	.058
SWEST	"	.073	.080
MNT	"	.038	.049
LFAMSIZE	Family size	3.219	2.840
GROUPSZP	Group size	19386	1
ESHAREC	Employer's Share	.740	.024

was used to predict the insurance purchases reported here. The table reports the comparison separately for high and low risk families.

Table 4 reports the means of the independent variables of the health insurance prediction equation. The group and individual market means are provided.

zero. The test for adverse selection discussed below was done twice, once setting the insurance level to zero for these families and once not adjusting the insurance level. The results of the test were invariant to the procedure followed. The results reported are for the test with the insurance level set to zero for the families predicted to have negative insurance.

Table 5 reports the estimated parameters and t-statistics of the demand discrepancy decomposition equation. The dummy variable LOW RISK is positive and significant. The positive sign of the coefficient indicates that the difference between predicted and actual insurance purchases is greater for low risks than for high risks. This finding is consistent with adverse selection being present in the individual market. An alternative explanation of this finding is that low risks receive more insurance through group insurance purchase than they desire.

Table 5
DEMAND DISCREPANCY DECOMPOSITION EQUATION
Dependent Variable: (Predicted Insurance - Actual Insurance)
(R^2=.36, N=225)

Variable	Coefficient	(t-stat)
INTERCEPT	-1022.95	(-4.03)
LOW RISK	375.01	(3.12)
EDUCAT	-0.68	(-0.08)
LOG OF INCOME	59.28	(2.53)
SEX	23.11	(0.33)
AGE	3.37	(1.84)
MARITAL STATUS	-148 45	(-2.24)
BLACK	-47.43	(-0.37)
HISP	-15 62	(-0.08)
OTHER	600.86	(1.92)
LIMMAJ	-456.08	(-4.07)
LIMAMT	345.05	(3.95)
LIMMIN	491.60	(2.18)
HOPER	-60.89	(-0.52)
HTRANS	78.24	(0.59)
HSERV	70.78	(0.81)
HLABOR	122.76	(0.85)
HFLABOR	364.36	(1.60)
HFOWN	46.83	(0.49)
HUNK	40.73	(0.52)
HMGR	208.62	(2.07)
HSALES	-133.89	(-1.38)
HCLERK	105.54	(0.89)
HCRAFTS	145.08	(1.64)

The income variable coefficient is of the hypothesized sign and is significant. The tax advantage to purchasing insurance in the group market leads to a greater consumption of medical insurance in the group market than in the individual market. The income variable therefore accounts for a portion of the discrepancy in demand between the two markets.

The age, activity limitation, and marital status variables are all significant and of the hypothesized signs. This indicates that factors which insurers use for underwriting and classification in the individual market that are not similarly used in the group market explain in part the discrepancy in consumption between the two markets. The insureds that are surcharged in the individual market on the basis of rating characteristics obtain more insurance in the group market where they are not surcharged. Note that the parameter estimates of the activity limitation variables decrease as activity limitation becomes more severe. Two conditions could give rise to this

relationship. Consumption in the group market may decrease as activity limitations increase. This would result if individuals were forced to leave the job market and thus lose access to group insurance as limitations became more severe. An alternative explanation is that individuals purchase more insurance in the individual market as activity limitations become more severe.

Of the occupation classification variables only those for the managerial and crafts professions are statistically significant at the 5 percent level.

None of the race variables is significant at the 5 percent level of significance although OTHER is borderline. The conclusion, that state prohibitions on the use of these types of information for the purposes of underwriting and rating leads to adverse selection, is therefore not supported by the data. However, data which does not identify the state of residence of the families surveyed is not well suited for testing this hypothesis.

The empirical results of the hypothesis test are consistent with the theory that adverse selection is present in the individual medical insurance market. Low risks have a lower level of insurance consumption than would be true in a market with less adverse selection. As previously mentioned, the alternative hypothesis that low risks are off their demand curve in the group market and on their demand curve in the individual market also may explain the empirical results.

Subsidization In The Individual Health Insurance Market

The second major hypothesis of this study is that in an insurance market with asymmetric information there will be a low risk to high risk wealth transfer. This occurs through a subsidization of high risks insurance purchases by the low risks.

This hypothesis is tested by comparing the mean contributions to insurer surplus and expenses of the low and high risks. The expected contribution is defined for each family in the sample as:

$$CIS_i = Premium_i - I_i,$$

where, CIS_i is the contribution to insurer surplus and expenses of family i, and $Premium_i$ is the premium paid by or on behalf of family i for insurance coverage. I_i is as defined earlier.

The premium an insurer charges is made up of two components. The first component is the pure premium. The pure premium is the insurer's estimate of expected benefits it will pay to the insured. The second component is a loading to cover the expenses the insurer incurs writing insurance. This includes marketing expenses, administrative expenses, tax expense, as well as a normal return on capital.

I_i is insured family i's estimate of expected benefits to be received from the insurer. In a market with perfect information I_i will equal the pure premium of the insurance policy which will be different for high and low risks. High risks will pay more for insurance coverage than low risks. In a market with asymmetric information the insurer can not differentiate

between high and low risks. Consequently, the pure premium charged to both groups will be equal. In comparison to the case of perfect information the high risks will be undercharged and the low risks will be overcharged for insurance. With asymmetric information the insureds have more information than the insurers, as a result I_i will now differ from the pure premium. The contribution to surplus and expenses will differ by risk group. The low risks will make a greater contribution than they would in a market with complete information. The high risks will make a lower contribution.

Evidence that the contribution made by low risks exceeds that of high risks would indicate that the market for individual health insurance is characterized by information asymmetries. Second, it would indicate that the models of subsidizing equilibrium developed by Wilson and Miyazaki better describe the market for individual health insurance than the separating equilibrium model of Rothschild and Stiglitz. In the Wilson and Miyazaki models market equilibrium in a market with adverse selection may entail subsidization of high risks' insurance purchases by low risks. While neither model of equilibrium guarantees a subsidization will occur, both allow for it. The separating equilibrium of Rothschild and Stiglitz does not include the possibility of subsidization. The pooling equilibrium of the Wilson model occurs when both high and low risks purchase the same policy. In the Miyazaki model there is a separating equilibrium with a subsidization across policies.

The hypothesis that subsidization occurs is tested with the equation:

$$CIS_i = b_1 LR_i + b_2 Z_i + e_i,$$

where CIS_i is the contribution to insurer expenses and surplus made by insured i. LR_i is the low risk classification variable described earlier. A positive relationship between LR_i and CIS_i would indicate that low risks subsidize high risks. Z_i is a vector of variables which insurers typically use in underwriting, pricing and classification. These variables include age, sex, marital status, family size, and degree of activity limitation. Recall that regulations which vary by state may prohibit the use of some information. Due to multiplicative expense loading a positive relationship is expected between the contribution to insurer expenses and surplus and the risk factors age, family size, and degree of activity limitation. A negative relationship is expected with the risk factors sex and marital status as females are typically better risks than males and married individuals are typically better risks than singles.

Table 6 reports the results of the subsidization test.

The variable LOW RISK is positive and statistically significant which supports the hypothesis that low risks subsidize the insurance consumption of high risks in the market for individual health insurance. Of the insurer classification variables only the family size variable and two of the activity limitation classifiers are significant. The family size variable is positive as hypothesized. The two activity limitation variables that are significant are the two which correspond to the most severe degrees of activity limitation.

The negative relationship exhibited by these variables may result if insurers are unable to assess accurately the health status of insurance applicants. This would lead to undercharging those with activity limitations.

Table 6
ANALYSIS OF CONTRIBUTION TO INSURER EXPENSES AND SURPLUS
Dependent Variable: CIS_i
(R^2=.56, N=225)

Variable	Coefficient	(t-stat)
INTERCEPT	-629.19	(- 2.76)
LOW RISK	915.94	(5.11)
SEX	14.47	(0.14)
MARITAL STATUS	-147.11	(- 1.35)
FAMILY SIZE	92.23	(3.77)
AGE	-4.22	(-1.67)
LIMMAJ	-1774.29	(-10.32)
LIMAMT	-539.32	(- 4.11)
LIMMIN	279.16	(0.81)

Conclusions

There are two major findings of this study. First, low risk insureds purchase more insurance in the group market than in the individual health insurance market. The lower consumption by low risks in the individual market is consistent with adverse selection theory. However, the lower consumption may also result from low risks being off their true demand curves in the group market.

The second finding of this study is that low risks subsidize the insurance consumption of high risks in the individual health insurance market. This finding is consistent with the models of Wilson and Miyazaki of market equilibrium in a market with adverse selection. However, subsidization is inconsistent with the separating equilibrium model of Rothschild and Stiglitz. The data used in this study are not suitable for testing whether the pooling equilibrium model of Wilson or the separating equilibrium model with cross policy subsidization of Miyazaki better describes the individual health insurance market.

Taken together the findings of this study suggest that adverse selection is present in the market for individual health insurance. Adverse selection leads to reduced insurance consumption by the low risks and a wealth transfer from low risks to high risks.

References

Akerlof, George A., 1970, The Market For 'Lemons': Quality Uncertainty And The Market Mechanism, *Quarterly Journal of Economics*, 84: 488-500.

Beliveau, Barbara, 1981, Two Aspects of Market Signaling. Ph.D thesis, Yale University.

Bluhm, William and Spencer Koppel, 1988, Individual Health Insurance Premiums, in: Francis O'Grady, ed., *Individual Health Insurance* (Society of Actuaries) 57-91.

Cafferata, Gail Lee, 1983, Private Health Insurance: Premium Expenditures and Sources of Payment, National Health Care Expenditures Study Data Preview 17, U.S. Department of Health and Human Services.

Conrad, Douglas A., David Grembowski, and Peter Milgrom, 1985, Adverse Selection Within Dental Insurance Markets, in: Richard M. Scheffler and Louis F. Rossiter, eds., *Advances In Health Economics and Health Services Research* (Greenwich, CT: JAI Press) 171-90.

Cummins, J. David, Barry D. Smith, Neil R. Vance, and Jack VanDerhei, 1983, *Risk Classification in Life Insurance*, (Boston: Kluwer, Nijhoff Publishing).

Dahlby, B. G., 1983, Adverse Selection and Statistical Discrimination, *Journal of Public Economics* 20: 121-30.

Duan, Naihua, Willard Manning, Carl Morris, and Joseph Newhouse, 1983, A Comparison of Alternative Models for the Demand for Medical Care, *Journal of Business and Economic Statistics* 1: 115-26.

Goldstein, G. S. and Mark V. Pauly, 1976, Group Health Insurance As A Local Public Good, in: Richard N. Rosett, ed., *The Role of Health Insurance In The Health Services Sector* (New York: National Bureau of Economic Research) 73-114.

Miyazaki, Hajime, 1977, The Rat Race and Internal Labor Markets, *The Bell Journal of Economics* 8: 394-418.

Rothschild, Michael and Joseph Stiglitz, 1976, Equilibrium In Competitive Insurance Markets: An Essay On The Economics of Imperfect Information, *Quarterly Journal of Economics* 90: 629-49.

Wilson, Charles, 1977, A Model of Insurance Markets With Incomplete Information, *Journal of Economic Theory* 97: 167-207.

Source Book of Health Insurance Data: 1989, 1990, (Washington D.C.: Health Insurance Association of America).