Insurance and the utilization of medical services

Jonathan Meer[a], Harvey S. Rosen[b],*

[a] Department of Economics, Stanford University, Stanford, CA 94305, USA
[b] Department of Economics, Princeton University, New Jersey, Princeton, NJ 08544-1021, USA

Abstract

Most data sets indicate a positive correlation between having health insurance and utilizing health care services. Yet the direction of causality is not at all clear. If we observe a positive correlation between the utilization of health care services and insurance status, we do not know if this is because people who anticipate poor health buy more insurance (or take jobs with generous medical coverage), or because insurance lowers the cost of health care, increasing the quantity demanded.

While a few attempts have been made to implement an instrumental variables (IV) strategy to deal with endogeneity, the instruments chosen have not been entirely convincing. In this paper we revisit the IV estimation of the reduced form relationships between insurance and health care utilization taking advantage of what we argue is a good instrument—the individual's self-employment status. Our main finding is that a positive and statistically significant effect of insurance continues to obtain even after instrumenting. Indeed, instrumental variables estimates of the impact of insurance on utilization of a variety of health care services are larger than their non-instrumented counterparts.

The validity of this exercise depends on the extent to which self-employment status is a suitable instrument. To argue this case, we analyze panel data on transitions from wage-earning into self-employment and show that individuals who select into self-employment do not differ systematically from those who remain wage-earners with respect to either the utilization of health care or health status. While this finding does not prove that self-employment status is an appropriate instrument, it is encouraging that there appear to be no underlying differences that might lead to self-employment per se affecting health services utilization.
© 2003 Elsevier Ltd. All rights reserved.

Keywords: Insurance; Self-employment; Health care services

> You'd better have some medical insurance or you're gonna die. That's right. Everybody says, oh, you've got to eat right, exercise. No, you don't. You need some coverage.
> –Chris Rock

Introduction

The Robert Wood Johnson Foundation (RWJF) sponsors a web site called Covering the Uninsured

(*CoveringTheUninsured.org*). The top of the home page delivers this message: "When you're uninsured, life turns out differently." In the center of the page is a picture of a downcast little girl. The text to the left of the girl states, "Her mom gets cancer. They find the tumor early. Her mom is OK." To the right it says, "Her mom gets cancer. She's diagnosed too late. Her mom is gone." The point is clear: Health insurance increases an individual's utilization of health care services which, in turn, leads to better health care outcomes. The RWJF web site certainly reflects the conventional wisdom in policy debates about health insurance.

There are two links in the causal chain of the conventional wisdom, from insurance to the utilization of health care services, and from utilization to health status. This paper focuses on the first link. A number of

*Corresponding author. Tel.: +1-609-258-4022; fax: +1-609-258-6419.

E-mail addresses: jmeer@stanford.edu (J. Meer), hsr@princeton.edu (H.S. Rosen).

previous papers have examined the relationship between insurance status and health services utilization. Most have been observational studies, which analyze outcome differences between insured and uninsured populations. These papers generally show that having insurance increases the utilization of health care services (Brown, Bindman, & Lurie, 1998). However, Levy and Meltzer (2001) note that observational studies "are hopelessly confounded by both observable and unobservable differences between patients who do and do not have health insurance." For the usual reasons, this can lead to inconsistent estimates of the impact of insurance coverage on health care utilization.

More generally, as Gruber (2000) points out, "insurance coverage itself may be a function of health status, leading to endogeneity bias in estimates of the effects of insurance on health and the utilization of medical care." However, the direction of the bias is not clear a priori. Anticipation of relatively high utilization of medical services might lead an individual to seek insurance, which would tend to impart an upward bias to the estimated impact of insurance on utilization. On the other hand, insurers may be able to identify people who will be intensive users of medical services and either decline to offer insurance or charge such a high price that they do not purchase it. An anecdote along these lines appeared recently in the *Wall Street Journal* (April 9, 2002). The story concerned a woman who had been paying $417 per month for health insurance, but whose rate increased to $1881 per month after she was diagnosed with breast cancer. To the extent the tendencies present in the story are typical, the estimated impact of insurance is biased downwards.

In this paper, we examine the link between insurance and the utilization of health care services using an instrumental variables (IV) strategy to estimate the reduced form relationship between insurance and health care utilization. We take advantage of what we argue is a good instrument—the individual's self-employment status. The next section describes the construction of the data set, which is drawn from the 1996 through 1998 waves of the Medical Expenditure Panel Survey (MEPS). The subsequent section discusses econometric issues and presents the results. We find that IV estimates of the impact of insurance on the utilization of a variety of health care services are greater than estimates that ignore endogeneity.

Of course, the validity of these results depends on the extent to which self-employment status is a suitable instrument. Ensuing section discusses the case for this instrumentation strategy. The centerpiece of this discussion is the analysis of panel data on transitions from wage-earning into self-employment. We show that individuals who select into self-employment do not differ systematically from those who remain wage-earners with respect to the utilization of health care

and health status. That is, there appear to be no underlying differences that might lead to self-employment per se affecting health services utilization are discussed. The penultimate section discusses the possibility that, regardless of the effect on utilization, lack of insurance presents serious financial difficulties. Again instrumenting using self-employment status, we find that this is generally not the case, although the evidence is mixed. Finally, we conclude with a summary and suggestions for future research.

Data

The data for this paper are drawn from the 1996, 1997 and 1998 waves of the MEPS, which has individual-level information on insurance coverage, utilization of health services, health status and self-employment status. The three waves consist of two overlapping 2-year panels. We exclude persons younger than 18 and older than 62 in 1996. Those under 18 are unlikely to have a strong connection to the labor market, and those over 62 are facing retirement and have different health care options. This leaves us with a sample of 37,331 observations, comprising 23,851 individuals, of whom 9.27% are self-employed. There are more observations than individuals because most people were followed for 2 years.

Each year, respondents are asked about their insurance status, utilization of health care, and a variety of other questions. Certain utilization questions, asked yearly, are not asked in 1997, and some questions deal with tests and procedures that are appropriate for only one gender or the other. Therefore, some models are not estimated with the entire sample.

The MEPS provides information on the utilization of a variety of health care services. They include visits to providers (such as physicians and dentists) and preventative care (such as blood pressure checks, flu shots, physicals, and breast exams). Respondents are asked at the end of the year if, for example, they visited a doctor or had a mammogram in that year. The insurance question is similarly asked on an annual basis. Subjects are also asked whether the insurance is public or private, and whether it is provided by the employer or self-provided. If the individual has any kind of policy, we characterize him or her as being insured. Further, we construe an individual as being insured if the source of insurance is the spouse.

A description of the variables used in this analysis and the associated summary statistics can be found in Table 1. There is substantial variation in the utilization rates for various medical services. About 65% of the sample visited the office of a health care provider during the year; 43% had a physical exam; 72% had their blood pressure checked; but only 18% had a flu shot.

Table 1
Summary statistics

Variable	Description	Mean	Standard deviation
Office-based provider	Respondent had an office-based provider visit in the last year.	0.6517	0.4764
Chiropractor	Respondent visited a chiropractor in the last year.	0.03662	0.1878
Prescription	Respondent received a prescription for drugs in the last year.	0.6023	0.4894
Alternate care	Respondent used some form of alternate care (e.g. massage therapy or acupuncture) in the last year.	0.05425	0.2265
Night in hospital	Respondent spent at least one night in a hospital in the last year.	0.05920	0.2360
Outpatient hospital stay	Respondent had an outpatient visit to a hospital in the last year.	0.004768	0.06889
Dentist	Respondent visited a dentist in the last year.	0.3728	0.4836
Optometrist	Respondent visited an optometrist in the last year.	0.04642	0.2104
Blood pressure	Respondent had his or her blood pressure checked in the last year.	0.7227	0.4477
Cholesterol check	Respondent had his or her cholesterol level checked in the last year.	0.3938	0.4886
Physical	Respondent had a physical in the last year.	0.4282	0.4948
Flu shot	Respondent had a flu shot in the last year.	0.1792	0.3835
Prostate	Male respondent had a prostate exam in the last year.	0.2180	0.4129
Breast	Female respondent had a breast exam in the last year.	0.6354	0.4813
Mammogram	Female respondent had a mammogram in the last year.	0.5166	0.4998
Pap smear	Female respondent had a pap smear in the last year.	0.6221	0.4849
Midwest	Respondent lives in the Midwest.	0.2227	0.4161
South	Respondent lives in the south.	0.3535	0.4781
West	Respondent lives in the west.	0.2405	0.4274
Family size	Respondent's annualized family size.	3.212	1.603
Male	Respondent is male.	0.5156	0.4998
Black	Respondent is black.	0.1318	0.3383
Other race	Respondent is neither white nor black.	0.04682	0.2113
GED	Respondent has a GED.	0.04762	0.2129
High school degree	Respondent has a high school degree.	0.5343	0.4988
College degree	Respondent has a college degree.	0.1591	0.3657
Graduate degree	Respondent has a graduate degree.	0.06982	0.2548
Other degree	Respondent has some other degree.	0.03807	0.1914
Age	Respondent's age.	38.15	11.68
Log family income	Log of respondent's annual family income.	10.67	0.8343
Corporate	Respondent is self-employed and incorporated all year.	0.02604	0.1592
Sole proprietor	Respondent is self-employed and a sole proprietor all year.	0.05692	0.2317
Partner	Respondent is self-employed and in a partnership all year.	0.009697	0.09799
Insurance status	Respondent is insured all year.	0.8244	0.3805

These summary statistics are calculated from the MEPS for 1996–1998. Only individuals between the ages of 18 and 62 in 1996 are included.

Analysis

Preliminary issues

We wish to estimate how a variety of medical services utilization measures depend on insurance status and other covariates. In our models of the utilization of various health care services, the dependent variable, Y, takes a value of one if the individual used the service in question during that year, and zero otherwise. The independent variables in our basic model include an indicator variable for insurance status (Ins), region, family size, age, age-squared, gender, race, and education. We only include covariates that are very likely to

be exogenous. Age clearly affects both health and utilization (Lakdawlka & Philipson, 1998), and education has been linked with both health and ability to pay (Taubman & Rosen, 1982). Race (Kass, Weinick, & Monheit, 1996), region (Skinner, Fisher, & Wennberg, 2001), family size (Taubman & Rosen, 1982), and gender (Verbrugge, 1985) have all been shown to have important effects on the variables of interest. Year effects are also included. We use the conventional probit model:

$$\Pr(Y = 1) = \Phi[\alpha_0 + \alpha_1(Ins) + \alpha_2(X) + \varepsilon], \qquad (1)$$

where X is the vector of exogenous covariates, ε is the error term, and $\Phi[\bullet]$ is the cumulative normal distribution. In computing confidence intervals for our parameter estimates, we wish to account for possible within-individual correlation of the errors and right-hand side variables. To do this, we perform a clustered procedure, with all years of an individual serving as the cluster.

To begin, we estimate the model without any correction for the endogeneity of insurance. The results, presented in Table 2, show that insurance has a positive and significant effect on most measures of utilization. (Only the coefficients and *t*-statistics for the insurance variable are presented. The estimated coefficients on the other covariates are available on request.) For example, having insurance increases the probability of visiting an office-based care provider by 24.9 percentage points. Insurance also increases the likelihood of having a physical exam, mammogram, and a variety of other preventive procedures by as much as 25 percentage points. One cannot reject the hypothesis that it has no effect on visiting a chiropractor or using alternate care. This is unsurprising: insurance plans rarely cover these services, so it stands to reason that it should have no effect on their utilization. Insurance also has a statistically significant effect on hospital usage: 3.3 percentage points for overnight stays and 0.37 percentage points for outpatient visits. These figures are substantial given the small baseline proportions of hospital visits (see Table 1).

However, as noted above, insurance status may well be endogenous to utilization decisions. To deal with this in an IV framework requires an instrument that is well correlated with the endogenous variable but not with the error term in the second stage (Bound, Jaeger, & Baker, 1995). The individual's self-employment status seems to fit the criteria for a good instrument. There is good reason to believe that the individual's self-employment status is correlated with whether or not he or she has insurance. In particular, several previous studies have shown that there is a strong negative correlation between self-employment and medical insurance (Health Insurance Association of American, 2000; Perry & Rosen, 2001b).

Table 2
Probit estimates of health services utilization (marginal effects)

Dependent variable	Sample size	Insurance effect
Office-based provider	37583	*0.2491*
		(32.9)
Chiropractor	37583	0.002425
		(0.860)
Prescription	27249	*0.2122*
		(23.3)
Alternate care	22497	−0.003548
		(−0.910)
Night in hospital	37583	*0.03286*
		(10.5)
Outpatient hospital stay	37583	*0.003739*
		(4.27)
Dentist	37583	*0.2021*
		(27.0)
Optometrist	37583	*0.02398*
		(8.38)
Blood pressure	22336	*0.2004*
		(23.5)
Cholesterol check	21977	*0.1746*
		(18.6)
Physical	22269	*0.1795*
		(19.4)
Flu shot	21929	*0.07298*
		(10.2)
Prostate exam	10117	*0.1294*
		(11.8)
Breast exam	11073	*0.2327*
		(17.0)
Mammogram	5096	*0.2575*
		(11.3)
Pap smear	11081	*0.2289*
		(16.7)

Figures show the effect of a discrete change from 0 to 1 in insurance coverage for the whole year. Each coefficient is from a probit equation in which the left-hand side variable is the variable in the associated row, and the other right-hand side variables include: region, family size, race, sex, education, age, age squared, and year effects. *t*-statistics are in parentheses. Those estimates significant at the 5% level are italicized.

In this context, though, one should note that the self-employed are not a homogeneous group with respect to the institutional environments in which they function. They operate in different organizational forms—sole proprietorships, partnerships, and corporations—and the probability of being insured could vary with organizational form. In particular, those who are incorporated might be more likely to have insurance for two reasons. First, their expenditures for health insurance are fully deductible; for members of partnerships and sole proprietors, they are not (at least during our sample period). Second, to the extent that corporate enterprises have more employees, the owners can

purchase insurance at advantageous group rates.[1] Indeed, in our data, self-employed individuals who are organized as corporations have an insurance rate of 89.5%, as compared with 67.2% for sole proprietors and 71.3% for partners. Hence, organizational form potentially provides useful information, so instead of characterizing self-employment status by a single dichotomous variable, we use three, one for each of the organizational forms. The omitted group consists of all individuals who are not self-employed, that is, wage earners and the unemployed.[2] In our sample 2.6% of the individuals have incorporated businesses, 5.7% are sole proprietors, and 0.97% are in partnerships, for a total of 9.27% self-employed.[3]

The other key issue in assessing the adequacy of self-employment status as an instrument is whether it exerts an independent effect on health care utilization. While there is no obvious reason that this should be the case, a possible problem is that there might be unobservable differences between wage earners and the self-employed that affect utilization and health. Perhaps people who are too unhealthy to hold jobs as wage-earners opt for self-employment. Alternatively, some sort of 'animal spirit' may drive both health and the propensity to be self-employed. Previous research using a variety of data sets has addressed this issue, and shown that neither health status nor utilization of health care services is a good predictor of whether or not a wage-earner will make a transition to self-employment (Holtz-Eakin, Penrod, & Rosen, 1996; Perry & Rosen, 2001a). We update these results in the Expenditure section and confirm that they hold in our data—there is no selection on the basis of health-related variables. This exercise does not constitute a formal test for the adequacy of self-employment as an instrument; indeed, no such test exists. That said, while these findings cannot definitely exclude the possibility of unobservable heterogeneity, they certainly provide no support for the notion that people who select into self-employment are systematically different with respect to health-related attributes.

Instrumental variables estimates

As usual, we implement the two-stage estimation procedure by first estimating a reduced form regression of insurance status on all the covariates of the model. Importantly, in this first-stage regression, one can strongly reject the joint hypothesis that the three self-employment variables have zero coefficients. Indeed, the associated *F*-statistic, 219.8, more than satisfies the usual criteria for an adequate fit in the first stage (Cawley & Weight, 2000). Moreover, for each health service, an overidentification test cannot reject the null hypothesis of exogeneity of the instruments.[4] While this, in itself, does not prove that the instruments are good, it does indicate that they do not belong on the right-hand side of the equation themselves.

The next step is to estimate the probit equation (1) using the fitted values of the insurance variable from the first stage. Rivers and Vuong (1988) and others have discussed the issues that arise in obtaining consistent standard errors within this framework. The most straightforward solution is to compute bootstrapped standard errors.[5] The bootstrap is based on random sampling; if the data are heteroscedastic, then each sample will have a different distribution, resulting in inconsistent point estimates and standard errors (Efron, 1979). To investigate whether this is a problem in our data, we estimate the bootstrapped model several dozen times to see if the results change substantially. They do not, and we conclude that the data are unlikely to be sufficiently heteroscedastic to render the bootstrap algorithm unreliable.

Table 3 reports the marginal effects for the boot-strapped two-stage probit model.[6] Comparing these results to their counterparts in Table 2, we see that, in general, instrumenting for insurance in the utilization equations increases the magnitude of its effect. For example, the effect of insurance on visiting an

---

[1] See Thomasson (2000) on the advantages of group coverage.

[2] Grouping the unemployed together with wage-earners may be problematic, so we also estimate our models including a dichotomous variable equal to one if the individual is not employed and zero otherwise. This has no impact on our results. To explore this issue further, we simply exclude the unemployed entirely from the sample. This exclusion similarly leaves our results unchanged.

[3] An interesting question is whether the differences in insurance rates by organizational form translate into differences in the utilization of health care services. To investigate this question, we estimate a series of reduced form regressions, essentially substituting the self-employment variables for insurance status in Eq. (1). The results, available upon request, show that incorporated individuals generally do not differ greatly from wage earners in their utilization of medical services, whereas sole proprietors and partners tend to utilize medical services less than wage earners. This is more or less what one would expect given the pattern of insurance coverage across the organizational forms.

[4] For example, for office-based provider visits, the chi-square test with two degrees of freedom is 3.91, which is significant only at the 0.14 level.

[5] As noted above, we perform a clustered procedure to allow for the possibility of within-individual correlation of the errors. To implement this in a bootstrapping context, we first create a list of individuals. For each bootstrap iteration, we then draw a set of individuals from this list, and use all years of data from the selected individuals to construct that iteration's dataset.

[6] Two-stage probit coefficients were calculated using a Stata module developed by Jonah Gelbach of the University of Maryland (Gelbach, 1997).

Table 3
Two-stage probit estimates of health services utilization (marginal effects)

Dependent variable	Sample size	Insurance estimate
Office-based provider	37331	*0.3517*
		(5.45)
Chiropractor	37331	*−0.1591*
		(−2.65)
Prescription	27077	*0.3541*
		(4.69)
Alternate care	22340	*−0.3198*
		(−3.65)
Night in hospital	37331	*0.05369*
		(3.84)
Outpatient hospital stay	37331	0.0066395
		(0.366)
Dentist	37331	*0.2076*
		(4.30)
Optometrist	37331	*0.03790*
		(3.10)
Blood pressure	22177	*0.5194*
		(7.10)
Cholesterol check	21818	*0.3989*
		(9.74)
Physical	22110	*0.3769*
		(7.40)
Flu shot	21776	*0.2270*
		(11.0)
Prostate exam	10062	*0.2301*
		(6.84)
Breast exam	10977	*0.4383*
		(3.35)
Mammogram	5060	*0.3301*
		(2.40)
Pap smear	10985	*0.3787*
		(2.76)

Figures show the effect of a discrete change from 0 to 1 in insurance coverage for the whole year. Each coefficient is from a probit equation in which the left-hand side variable is the variable in the associated row, and the other right-hand side variables include: region, family size, race, sex, education, age, age squared, and year effects. The instruments for insurance are indicators for operating as a corporation, sole proprietor, or partner. *t*-statistics are in parentheses, for bootstrapped standard errors. Those estimates significant at the 5% level are italicized.

office-based provider increases from 24.9 percentage points to 35.2 percentage points. The change between the two specifications is even more pronounced for the preventative care measures. The insurance effect on the probability of having a cholesterol check, for example, increases by 22 percentage points. It appears, then, that ignoring endogeneity leads to underestimating the impact of insurance coverage on utilization. As noted earlier, this is consistent with a scenario in which

individuals who are likely to be intensive users of health services find themselves discouraged from obtaining insurance for any of a variety of reasons.

Alternative specifications

We have shown that when one takes into account the endogeneity of insurance status, the magnitude of the insurance effect on the utilization of a variety of health care services increases. We estimate a number of variations on our basic model in order to assess the robustness of these results.

Income

A positive correlation between health status and income is well-documented, but the direction of causation is controversial. (see, for example, Feinstein, 1993; Smith & Kington, 1997; Smith, 1999). In light of this controversy, we choose not to include income as a right-hand side variable in our basic models. However, given the widespread belief that low income leads to less or lower quality medical care, it seems worthwhile to re-estimate the model including family income as a covariate. The IV results, available upon request, indicate that none of our substantive findings changes.

In connection with income, an intriguing finding from the Rand experiment (Newhouse, 1993) is that changes in the extent of insurance coverage have some impact on the health status of low-income people, but not for the rest of the population. In our context, this suggests that one might want to estimate a model that includes an interaction between insurance status and an indicator for whether or not the individual has a low income. We create a dichotomous variable that equals one if the individual's family income is in the bottom quartile, and augment our basic models from Table 3 with both this indicator variable and its interaction with insurance status. We find that the interaction term is negative for most health services, and generally statistically insignificant or only marginally significant. Most importantly, including the interaction term does not affect our substantive results with respect to insurance and the utilization of various health services.

Marital status

It is easy to imagine that marital status affects health-related outcomes. Married people may have more stable home environments, better diets, and so on. However, it is equally easy to imagine that the direction of causality runs the other way. For example, healthy people may be more likely to find mates than unhealthy people (Gold-man, 2001). Because of the possibility of joint determination, we leave marital status out of our basic model. But marital status has been included in other studies

(e.g., Gruber & Madrian, 2002), so it seems instructive to estimate a set of models that include a dichotomous marriage indicator. The results, available upon request, are neither qualitatively nor quantitatively very different from those of the basic model.

Econometric specification

A possible problem with our results is that they are a consequence of the particular assumptions underlying the two-stage probit model. In the probit model, the two-stage procedure generates consistent estimates only if the error terms in both the first- and second-stage equations are jointly normally distributed, and both equations are correctly specified. In a linear probability model, the conditions for consistency are less stringent— the right-hand variables in the first-stage equation have to be uncorrelated with the error term in the second-stage equation, but consistent estimates may be obtained even if some variables that belong in the first-stage equation are omitted. Therefore, despite the well-known limitations of the linear probability model, it seems worthwhile to use it to check our estimates.

The results, reported in Table 4, are very similar to those obtained using the probit. For example, in Table 3 the insurance effect on the probability of visiting an office-based provider is 35 percentage points; in Table 4 it is 31 percentage points. Thus, we feel confident that our results are not an artifact of the assumptions behind the two-stage probit model.

Transitions into self-employment

As suggested above, some unobserved variable might drive both health status and self-employment decisions. One can imagine, for example, that relatively unhealthy people are unable to hold jobs and hence enter self-employment. Alternatively, perhaps healthy, energetic people are particularly well-suited for running their own businesses. To the extent that either type of selection takes place, self-employment status is not a suitable instrument.

Examining the determinants of transitions from wage earning into self-employment can be instructive in ascertaining whether some underlying variable drives both health and self-employment status. Consider a group of wage-earners during a given time period. If the probability that an individual in this group transits to self-employment in the subsequent period is independent of his or her health status at the outset, then one can feel some confidence that selection into self-employment on the basis of health is not driving our results. On the other hand, if healthier or less healthy individuals are more likely to make transitions into self-employment, the interpretation of our findings becomes problematic.

Table 4
Two-stage least squares estimates of health services utilization (marginal effects)

Dependent variable	Sample size	Insurance estimate
Office-based provider	37331	*0.3143* *(5.40)*
Chiropractor	37331	*−0.1245* *(−3.57)*
Prescription	27077	*0.3249* *(4.44)*
Alternate care	22340	*−0.2214* *(−4.68)*
Night in hospital	37331	*0.06742* *(2.71)*
Outpatient hospital stay	37331	0.01106 (1.72)
Dentist	37331	*0.2001* *(3.37)*
Optometrist	37331	*0.05091* *(2.19)*
Blood pressure	22177	*0.4577* *(6.06)*
Cholesterol check	21818	*0.4827* *(6.29)*
Physical	22110	*0.4140* *(5.34)*
Flu shot	21776	*0.3744* *(6.15)*
Prostate exam	10062	*0.3363* *(4.50)*
Breast exam	10977	*0.4308* *(3.01)*
Mammogram	5060	*0.3175* *(2.08)*
Pap smear	10985	*0.3747* *(2.57)*

Figures are the effect of a discrete change from 0 to 1 in insurance coverage for the whole year. Each coefficient is from a linear probability model in which the left-hand side variable is the variable in the associated row, and the other right-hand side variables include: region, family size, race, sex, education, age, age squared, and year effects. The instruments for insurance are indicators for operating as a corporation, sole proprietor, or partner. *t*-statistics are in parentheses. Those estimates significant at the 5% level are italicized.

Using data from the Survey of Program Participation and the Panel Study of Income Dynamics, Holtz-Eakin, Penrod and Rosen (1996) find no evidence for this latter hypothesis. Similarly, Perry and Rosen's (2001a) analysis of the 1996 and 1997 MEPS data suggests that the absence of health differences between wage earners and the self-employed does not appear to be due to the fact that people with relatively good health tend to select into self-employment.

In this section, we update these analyses and examine whether the results hold for the sample of individuals

Table 5
Probit estimates of transitions into self-employment (marginal effects)

Specification	Sample size	Health status	Office based provider visit
1	12974	0.0005902 (0.201)	—
2	12974	—	0.0001546 (0.0850)
3	12974	0.0006140 (0.207)	0.0001851 (0.101)

Each coefficient is from a probit equation in which the left-hand side variable is the probability of being self-employed in year t, and on the right-hand side are the following variables dated period $t - 1$: region, family size, race, sex, education, age, and age squared. The sample consists only of those individuals who were wage-earners in period $t - 1$. Hence, each coefficient shows the probability of making a transition from wage-earning to self-employment, ceteris paribus. Coefficients are the effect of a discrete change from 0 to 1 for the indicator variables for initial health status and for whether the individual visited a health services provider. t-statistics are in parentheses. Those coefficients significant at the 5% level are italicized.

used in this study. Specifically, the overlapping 2-year structure of the MEPS allows us to construct a pooled data set for two sets of transitions, from 1996 to 1997 and 1997 to 1998. There are 19,744 individuals with 2 years of complete data. Approximately 1% of wage earners leave their jobs in each year to become self-employed.

We model the probability that an individual who is a wage-earner in year $t - 1$ makes a transition to self-employment in period t. The sample consists of individuals who are wage-earners in year $t - 1$. The dependent variable is an indicator variable for whether the individual is self-employed (in any organizational form) in year t.[7] The right-hand side includes the covariates used in our canonical model dated year $t - 1$, a time effect to take into account any systematic differences between the two transition periods, and some measure of the individual's health status and/or utilization of health services in year $t - 1$.

Row (1) of Table 5 shows the results when self-reported health status is included on the right-hand side; row (2) when an indicator variable for an office visit to a health-care provider is included; and row (3) when both are included. None of the health-related variables is significant in any of the specifications, and all are minuscule in magnitude. For example, the point estimate in row (1) suggests that being in good health makes one only 0.06 percentage points more likely to

make a transition into self-employment, and row (2) indicates that having paid a visit to a health-care provider makes one only 0.02 percentage points more likely to make such a transition. We find similar results when we use other measures of health care utilization. In short, Table 5 confirms previous findings that neither health status nor health services utilization are predictors of whether wage-earners will become self-employed. This does not constitute a formal test that self-employment status is a good instrument. We are nevertheless encouraged that it appears unlikely that some variable is driving both health and the self-employment decision.

Expenditures

We have shown that once simultaneity is taken into account, the impact of health insurance upon the use of a variety of health care services increases. Our focus on this issue reflects perhaps the dominant issue in the public policy debate over the uninsured—their ability to obtain medical services. However, the discussion over health care sometimes loses sight of the key function of insurance, namely, to spread consumption over different states of the world. It is therefore important to ask whether, in the absence of insurance, paying for health care leads to a serious reduction in a household's standard of living.

The MEPS includes information about out-of-pocket expenditures on health care. We wish to estimate how having insurance affects these expenditures, taking into account the endogeneity of insurance; as before, we use a set of self-employment variables as instruments. A sensible way to examine how health care expenditures affect the "standard of living" is to measure their magnitude relative to consumption expenditures. However, the MEPS does not include consumption information. Hence, we use income, which is recorded in the survey. We exclude families whose incomes are below $5000 from this analysis, as we suspect that income is transitorily very low or mismeasured for such families.

To begin, we use ordinary least squares to estimate a regression with out-of-pocket medical expenditures on the left-hand side, and insurance and the same exogenous covariates as in Table 2 on the right-hand side. The coefficient on the insurance variable indicates that it reduces out-of-pocket medical expenditures by a mere $16.50 with a standard error of $12.20. When the model is estimated using two-stage least squares, the effect increases substantially. Having insurance lowers out-of-pocket medical expenditures by $482 (s.e. = $176). It may be more instructive, however, to examine expenditures as a proportion of income. When we estimate an ordinary least squares regression of the proportion of income on the same covariates as above,

---

[7] We also examine transitions from wage earning into particular organizational forms. The results are not affected.

we find that insurance reduces medical expenditures by 0.40% of income (s.e. = 0.065%). The two-stage least squares estimate is a reduction of 0.58% of income with a standard error of 0.46%.

These calculations suggest that out-of-pocket medical expenditures are, on average, not a large proportion of the incomes of the uninsured.[8] One could argue, however, that in the context of medical expenditures, the effect on the average is not quite to the point. A key purpose of insurance, after all, is to protect against a small probability of a major loss. To investigate whether the absence of insurance can lead to extraordinarily high expenditures, we create an indicator that equals one if out-of-pocket expenditures as a proportion of income are greater than 20% and zero otherwise.[9] Conventional probit estimation indicates that having insurance reduces the probability of reaching this threshold by 0.51 percentage points (s.e. = 0.13 percentage points). In addition to being statistically significant, this is a quantitatively substantial effect, given that only 0.65% of the sample experiences out-of-pocket expenditures greater than 20% of their incomes. Instrumenting in a two-stage probit framework, we find that the point estimate on the insurance variable increases in absolute value but is rendered statistically insignificant. Specifically, insurance reduces the probability by 2.3 percentage points, but with a standard error of 3.3 percentage points. The results, then, are inconclusive. While one cannot reject the hypothesis that the effect is equal to zero, the large standard error encompasses some potentially important changes in the probability of a serious diminution in a household's standard of living.

Conclusion

Using the Medical Expenditure Panel Survey, we have examined the effect of health insurance on the utilization of health care services. We began by replicating previous studies which show that insured individuals are more likely to utilize a variety of health care services than those who do not, ceteris paribus. However, as several previous investigators have pointed out, insurance status is likely to be an endogenous variable in this context, so that attaching a causal interpretation to this statistical relationship is problematic. We use an instrumental variables estimation strategy to address this problem, arguing that self-employment status is a suitable instrument. Our main finding is that a positive and statistically significant effect of insurance is obtained even after instrumenting. Indeed, instrumental variables estimates of the impact of insurance on utilization of a variety of health care services are larger than their non-instrumented counterparts.

Given that this study is based upon data from the United States, which has a rather unique set of health care institutions, any generalization of our findings to other countries must be done with great caution. That said, they might be of relevance to those nations that are debating whether or not to scale back their systems of universal health insurance. Specifically, to the extent that our results carry over to other countries, reforms that reduce the incidence of health insurance will also decrease the utilization of health care services. In the same way, policies that increase the availability of health insurance will likely increase utilization of health care services and possibly health care costs.

However, as suggested at the outset, utilization of health care services is not an end in itself. Policymakers are concerned not only about the link between insurance and the utilization of health services, but also between insurance and health status. Our finding that insurance is associated with greater utilization of health care services does not necessarily imply that insurance leads to better health. Some argue, for example, that access to health care is responsible for only a relatively small part of health, with more important determinants being genetics, environment, and health behaviors (Institute for the Future, 2000, p. 23). Statistical examination of the links between insurance status and health comes up against an endogeneity problem much like the one involved in the analysis of utilization. Dealing with endogeneity in the context of health status is an important topic for future reasearch.

---

[8] An interesting question is whether this finding holds for individuals with particularly severe health problems. The MEPS indicates whether or not the individual has a "priority condition," which is one of a number of serious illnesses such as cancer and heart disease. When we interact a dichotomous variable for the presence of a priority condition with the insurance variable, the IV estimate is −0.40 percentage points. That is, for a person with a priority condition, not having insurance increases out-of-pocket expenditures relative to income by 0.40 percentage points. However, one cannot reject the hypothesis that the interaction term is zero. The main effect of having a priority condition is 1.1 percentage points (s.e. = 0.78 percentage points).

[9] There is, of course, no obvious cutoff for when medical expenditures as a proportion of income become truly onerous. The 20% figure is arbitary, but our substantive results do not change when it is lowered to 15% or raised to 25%.

References

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). The cure can be worse than the disease: A cautionary tale regarding instrumental variables. *Journal of the American Statistical Association*, *90*(430), 443–450.

Brown, M., Bindman, A., & Lurie, N. (1998). Monitoring the consequences of uninsurance: A review of methodologies. *Medical Care Research and Review*, *55*(2), 177–210.

Cawley, J. (2000). *Body weight and women's labor market outcomes*. NBER Working Paper No. 7841. August.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*(1), 1–26.

Feinstein, J. (1993). The relationship between SES and health: A review of the literature. *Milbank Quarterly*, *71*(2), 279–322.

Gelbach, J. (1997). *Probitiv: Stata module to perform instrumental variables probit*. http://www.glue.umd.edu/ ~gelbach/ado/.

Goldman, N. (2001). *Social inequalities in health: Untangling the underlying mechanisms*. Working Paper, Princeton University, February.

Gruber, J. (2000). *Medicaid*. NBER Working Paper 7829, August.

Gruber, J., Madrian, B. (2002). *Health insurance, labor supply, and job mobility: A critical review of the literature*. NBER Working Paper 8817, February.

Health Insurance Association of America (2000). *Source book of health insurance data, 1999–2000*. Washington, DC.

Holtz-Eakin, D., Penrod, J., & Rosen, H. S. (1996). Health insurance and the supply of entrepreneurs. *Journal of Public Economics*, *62*, 209–235.

Institute for the Future (2000). *Health and health care 2010—The forecast, the challenge*. San Francisco: Jossey-Bass Publishers.

Kass, B. L., Weinick, R. M., Monheit, A. C. (1999). *Racial, ethnic differences in health, 1996*. Agency for Health Care Policy and Research, MEPS Chartbook No. 2, AHCPR Pub. No. 99-0001.

Levy, H., Meltzer, D. (2001). *What do we really know about whether health insurance affects health*? Working Paper, University of Chicago, June.

Newhouse, J. P. (1993). *Free for all? Lessons from the RAND health insurance experiment*. Cambridge, MA: Harvard University Press.

Perry, C. W., Rosen, H. S. (2001a). *The self employed are less likely to have insurance than wage earners. So what*? NBER Working Paper 8316, June.

Perry, C. W., Rosen, H. S. (2001b). *Insurance and the utilization of medical services among the self employed*. NBER Working Paper 8490, August.

Rivers, D., & Vuong, Q. H. (1988b). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics*, *39*, 347–366.

Skinner, J., Fisher, E. S., Wennberg, J. E. (2001). *The efficiency of medicare*. NBER Working Paper 8395, July.

Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relation between health and economic status. *Journal of Economic Perspectives*, *13*(2), 145–166.

Smith, J. P., Kington, R. S. (1997). Race, socioeconomic status, and health in late life. In L. G. Martin, B. J. Soldo (Eds.), *Racial and ethnic differences in the health of older Americans*. Washington, DC: National Academy Press.

Taubman, P., Rosen, S. (1982). *Healthiness, education, and marital status*. NBER Working Paper 611, August.

Thomasson, M. A. (2000). *The importance of group coverage: How tax policy shaped US health insurance*. NBER Working Paper 8316.

Verbrugge, L. M. (1985). Gender and health: An update on hypothesis and evidence. *Journal of Health and Social Behavior*, *26*(3), 156–182.