ELSEVIER

Journal of Clinical Epidemiology 57 (2004) 284–293

Journal of Clinical Epidemiology

Instrumental variables technique: cigarette price provided better estimate of effects of smoking on SF-12

J. Paul Leigh[a,*], Michael Schembri[b]

[a]Center for Health Services Research in Primary Care, and Department of Epidemiology and Preventive Medicine, 2103 Stockton, Suite 2224, University of California, Davis, Sacramento, CA 95817, USA

[b]Family and Community Medicine, University of California, San Francisco, Box 0900, San Francisco, CA 94145-0900, USA

Accepted 11 August 2003

Abstract

Objective: Debate surrounds the usefulness of the instrumental variables (IV) technique for medical research. The choice of an instrument for the technique has been contentious. This study estimated the effects of smoking on physical functional status. We chose an especially valid and strong instrument: cigarette price.

Study Design and Setting: The data were a nationally representative cross-sectional sample of 34,288 persons aged 30 to 91 in 1996–1997. The sample was drawn from the Community Tracking Study. Number of cigarettes smoked per day was predicted by the average cigarette price for the state in which the subject resided. The outcome measure was physical functional status and was measured by the SF-12 physical functional index.

Results: In multivariable models we found the following: cigarettes per day was strongly and negatively associated with the SF-12 index ($P < .001$); cigarette price was strongly and negatively associated with cigarettes per day ($P = .002$); the predicted cigarettes per day (the IV) was strongly and negatively associated with the SF-12 index in linear regression and tobit regression ($P = .047$ and $P = .021$).

Conclusion: Estimated coefficients from the IV method suggested that the effect of smoking on physical functional status was substantially larger than estimates that relied on conventional methods.

Keywords: Econometrics; Two-stage least squares; Tobit

1. Introduction

The instrumental variable (IV) technique, long a workhorse in economic analyses, has great potential in medical and epidemiologic research. A few IV studies have received some medical and epidemiologic attention [1–9]. Many more have been published in Health Services journals [10–13]. One problem has plagued the adoption of IV techniques in medical and epidemiologic research. Medical researchers have been skeptical of the IV technique because many of the "instruments" (unique variables) used in IV studies have been invalid and weak [14]. The concern about invalid and weak instruments has also been voiced by leading econometricians about many applied economic studies [15–18]. The main contribution of this study is to investigate a significant health problem with an instrument that is likely to convince many skeptics. Our problem is estimating the effects of smoking on physical functional status. Our instrument is cigarette price. An additional contribution is the intuitive explanation of the IV technique we offer in the Methods section.

It is important to generate the best possible estimates of the effects of smoking on health. Problems for individuals and societies can result when estimates are either too high or too low [19]. Smoking is a known cause of a number of diseases including lung cancer and chronic obstruction pulmonary disease (COPD). It is strongly suspected to be one of the causes of many others including heart disease, stroke, most cancers, and congenital anomalies [20]. It exacerbates even more diseases including diabetes, HIV, depression, respiratory infection, chronic liver disease, arthritis, nephritis, and ulcers [20]. Finally, smoking is a risk factor for accidents [21]. Many of these are among the leading causes of disability: heart disease, major depression, road-traffic accidents, stroke, COPD, lower respiratory infections, HIV, and arthritis [22,23]. With the possible exception of major depression, these diseases and injuries affect physical functioning, that is, physical disability. Our health measure

* Corresponding author. Tel.: 916-734-8542; fax: 916-734-8731.
E-mail address: pleigh@ucdavis.edu (J.P. Leigh).

0895-4356/04/$ – see front matter
doi: 10.1016/j.jclinepi.2003.08.006

is one of the most popular measures of physical function: the Physical Component Summary (PCS) of the SF-12 [24].

The IV technique is most easily understood in the context of treatment and outcome variables [25]. In observational data, some of the variation in the treatment variable can be polluted by reverse causality, unobserved variables, or measurement error. The IV essentially captures that portion of the treatment variable that is not polluted. This cleansed portion can then be used to assess the true effect of the treatment on the outcome.

A valid instrument satisfies two qualifications. First, it is logically related to and statistically correlated with a treatment variable. Second, there is no logical reason why the instrument should be directly related to the outcome other than the instrument's effect on the treatment. A strong instrument is *strongly* statistically correlated with a treatment. Cigarette price is such a variable. It is logically related to and strongly correlated with smoking: higher prices result in less smoking, on average. On the other hand, the price of cigarettes would not appear to be logically directly related to an individual's health. The only logical relation would be an indirect one: price affects cigarette use that, in turn, affects health. Cigarette price therefore appears to satisfy the conditions for a valid and strong instrument.

We used price-per-pack data from the 50 states and DC matched to persons who resided in those states and DC. Data on individuals were drawn from the Community Tracking Study (CTS) [26]. We used statistical software that not only estimates correct standard errors in IV techniques, but also correct standard errors for data that have geographic clusters.

2. Data and method

2.1. Data source

The CTS is an ongoing data collection effort conducted by the Center to Study Health System Change and financed by the Robert Wood Johnson Foundation [26]. We used the Household Survey portion of the CTS for the years 1996–1997. The CTS Household Survey is a large, nationally representative survey of the civilian, noninstitutionalized population consisting of 60,446 persons. Telephone interviews were conducted as well as in-person interviews for those without telephones. Data were drawn from persons living in 60 randomly selected communities nationwide. Extensive information was available on medical care use, activities of daily living, current and past smoking status, demographics, and medical insurance. Interviews and questionnaires were conducted in English and Spanish. Because physical functional status measures such as the SF-12 were designed for adults, and because functional decline does not generally occur among persons in their 20s, we restricted attention to persons age 30 and over. We also required that all persons in our sample had information on the SF-12 and smoking status. These restrictions resulted in a sample of 34,326. We did not exclude persons with missing values on the control variables (age, gender, race, education, and so on). We used imputed values provided by the CTS. Less than 2% had imputed values.

2.2. Dependent, control variables, and instrument

Cigarette use was measured as the number of cigarettes smoked per day, on average, over the past 30 days. Persons who quit within the past 30 days were regarded as smokers. The CTS collected information on their cigarettes per day prior to quitting.

The physical functional status variable was measured as the PCS of the SF-12 [24]. This summary was a composite score for questionnaire answers pertaining to whether and how much a subject's health or conditions limit some daily activities such as moving a table, pushing a vacuum cleaner, or climbing stairs, as well as answers to a general question on overall health. Answers were weighted according to an algorithm developed by Ware et al. [24]. The PCS was designed to be normally distributed. In our sample, it had a mean and standard deviation of 48 and 11. The instrument was the average price per pack within states and DC for 1997. These data were drawn from the Centers for Disease Control and Prevention, which in turn, derived them from the (now defunct) Tobacco Institute.

Control variables were not arbitrarily selected. We used the most popular control variables identified in the literature on estimating the effects of smoking on physical functional status, provided those variables were available in the CTS [27–29]. Our structural equations assumed that smoking depended on cigarette price, age, age squared, race, gender, education, marital status, income, employment status, children, and insurance status; and that health depended on all of the above except cigarette price.

2.3. Method

Several problems face researchers attempting to estimate effects of treatment on outcome using observational data. First, it could be that the outcome variable results in the treatment. For example, a rapid physical functional decline could scare someone and result in the person quitting smoking. The smoking that lead to the disability may have stopped months or years before, yet the person would likely still have the disability. Without adjustment for this reverse causality, conventional methods would underestimate the effect of smoking on health. Second, random measurement error for the treatment variable can result in an underestimate of the effect of treatment on outcome. Third, there may be some unobserved, perhaps unmeasureable, variable or set of variables that could influence or be influenced by both the treatment and the outcome. In most cases, we would want to exclude the unobserved variables. For example, risk aversion, a frequently unobserved variable, might lead people to never smoke and to maintain good health. Any correlation between smoking and health that did not remove risk aversion would overestimate the effect of smoking on health.

On the other hand, there may be spillover effects [30–32]. Second-hand smoke can cause disease in innocent bystanders. In addition, the more smokers in a geographic area, the less likely there is a social stigma associated with smoking and the more likely others will decide to smoke. Any correlation between *individual* smoking and *individual* health that did not include spillover effects would underestimate the effect of all smoking on health [30–32]. In the case of spillovers, we would want to include, not exclude, the unobservables. An IV, especially derived from a geographic instrument, would include spillover effects [30–32]. Technically, any of these three problems (reverse causality, measurement error, unobservable variables) result in biased estimates [33].

In theory, the IV technique solves all three problems. A valid instrument acts as a randomization device. In a randomized trial, we can think of a fair coin toss as deciding who is selected into treatment and who is not. In a valid IV analysis, the instrument assigns subjects to either treatment or no treatment using an assignment mechanism that is independent of the outcome. For example, the presence of unmeasured risk aversion in observational data represents a clear violation of random assignment. Persons in the treatment group (smokers) would likely be less risk averse than persons in the control group (nonsmokers). Our IV technique attempted to place subjects into treatment and control groups based upon cigarette price in a geographic area. Loosely speaking, subjects who were strongly influenced by risk aversion to become smokers or nonsmokers were excluded from the IV analysis.

The IV technique can be shown with two diagrams. Fig. 1 is drawn from Newhouse and McClellan (1998) [34]. The scheme in Fig. 1 assumes that price is the instrument. The arrows in Fig. 1 imply causality. Price affects smoking, which in turn, affects health. What is important about Fig. 1 and what makes price a valid instrument is that there are no arrows pointing to price. In the language of economics, price is exogenous. These directional arrows are merely assumptions. They must be established by appealing to logical arguments. Economists rely on economic theory. Perhaps the most basic tenet of economic theory is that the quantity demanded of a product (cigarettes) is inversely related to its price. This is the “law of demand.” As the price of cigarettes increases, theory (logic) suggests a decline in quantity of cigarettes consumed. This is common sense. This argument is one reason why so many states levy cigarette taxes. In addition to providing revenue, some governments want to reduce smoking.

A valid instrument must be theoretically related to the treatment variable but not, theoretically at least, be directly related to the outcome. Invalid instruments are those for which causality arrows might run from: (1) the treatment to the instrument; (2) from the outcome to the instrument; (3) from the instrument to the outcome; or (4) from other variables to the instrument, treatment, and outcome. A weak instrument is weakly statistically correlated with the treatment variable. Unfortunately, weak and invalid instruments frequently appear in the literature. For example, we used mother’s educational attainment as an instrument for adult child’s own educational attainment in assessing the correlation between education and health [35]. But mother’s educational attainment likely influences the child’s health in early years, which certainly influences the adult’s health in later years. Thus, the instrument (mother’s educational attainment) affects the outcome (health) for reasons other than the effects of the instrument on the treatment.

Fig. 2 is inspired by a similar picture in Kennedy [36]. We, however, extended his analysis. Fig. 2 shows the variation in the three variables. The overlapping variations are indicated by red, green, blue, and brown. Silver indicates that portion of health not correlated with either smoking or price. Brown is the variation in price that overlaps with smoking but not health. The overlap between smoking and health are three colors combined (red + green + blue). If we did not need the IV technique (i.e., if no reverse causality could pollute the estimate, if there were no unobserved variables, and if there were no measurement error) then the combined variation (red + green + blue) would allow us to estimate the true causal effect. But suppose the red area reflects the reverse causality or unobserved variables or measurement error. This red area would bias the estimate if we included

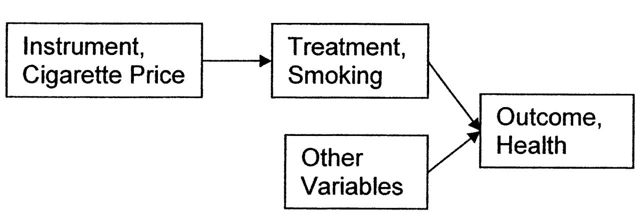

Fig. 1. Source: Newhouse, McClellan, 1998 [34].

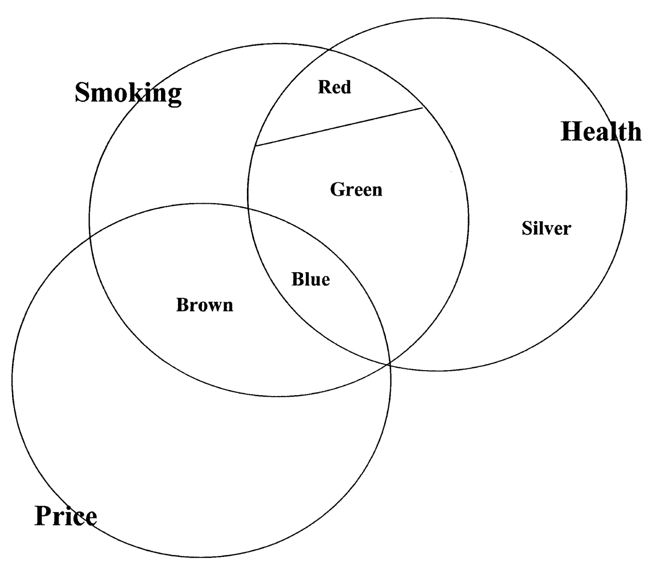

Fig. 2. Venn diagram drawn from Kennedy [36].

it and we mistakenly assumed all of the colors (red + green + blue) combined measured the effect of smoking on health. What we need is green and blue combined, but not red. The instrument provides us with part of what we need—blue. The variation measured by blue is the instrumental variable. Cigarette price is the instrument. The bigger blue is (without crossing into red or silver) the stronger is our instrument and instrumental variable. We are tempted to note that the blue area represents the best linear unbiased estimator, B.L.U.E., but the IV estimator is merely consistent, not B.L.U.E. [33]. Technically, the IV estimator is biased (in finite samples) but consistent (as samples go to infinity). A regression estimator that does not use the IV is biased and inconsistent.

Fig. 2 also is useful in indicating a common mistake new researchers make applying IVs. New researchers sometimes simply replace the treatment variable with the instrument in the regression explaining the outcome. They regress health on price. But this is a mistake because this regression of health on price would use brown and blue to estimate the effect of smoking on health. But brown is not correlated with health. The correct IV uses only that portion of *smoking* that is correlated with health. That portion is blue, and it corresponds to the predicted values of smoking in a regression of smoking on price. It is the portion of variation in smoking accounted for by price. The IV is the predicted value of smoking.

Notice all three circles intersect at the southeast corner of brown and blue. If the price circle were moved a little to the right, some of the price circle would overlap with the silver health circle. This would suggest an invalid instrument. It would suggest that there is some correlation between price and health that is independent of smoking.

The most important equation in the IV technique in our case is a regression of health on the predicted values of smoking. The correct standard of error of this regression, however, requires that the actual values of smoking rather than the predicted values be used to calculate the variance–covariance matrix [37]. A regression of health on the predicted values of smoking will yield the correct (consistent) coefficient, but not the correct standard error. Software is available to correctly estimate these standard errors (in *stata*, programs are *ivreg* and *svyivreg*). The *stata* program *svyivreg* simultaneously adjusts standard errors for geographic clustering and sample weights as well as the IV technique. The program *svyivreg* automatically uses all of the exogenous variables in the first stage regression (smoking on price, age, race, gender, and so on). Whereas the structural equation for smoking may depend only on cigarette price, the reduced form equation used to predict smoking must include all exogenous variables, that is, all variables other than smoking and physical functional status [33].

Smoking is recorded as number of cigarettes per day. Seventy-six percent (76%) of subjects had zeros. A tobit estimator might be more appropriate to estimate the smoking equation [38]. However, *stata* (nor any other software with which we are familiar) does not have a program to estimate the correct IV standard errors using tobit and adjusting for clustering. Nevertheless, a regression of health on the predicted values of smoking (from tobit) will yield the correct (consistent) coefficient.

We preferred to run regressions both ways: using linear regression that accounted for unique IV standard errors with *svyivreg* (Table 1) and second, using tobit that simultaneously accounted for geographic clusters and weights (*intreg*) but did not account for unique IV standard errors (Table 2).

We proceeded with three regressions in Tables 1 and 2 (1, 2, 3). These regressions are explained below. Table 2 did not include results on other exogenous variables. Table 2 results were shortened because results on all other exogenous variables were consistent with those in Table 1.

1. Regress cigarettes per day on price and all exogenous variables (age, race, gender, insurance variables, and so on) to obtain predicted values. Call these Cighat.
2. Regress SF-12 on cigarettes per day and all exogenous variables, except price.
3. Regress functional status SF-12 on Cighat (but not cigarettes per day) as well as all other exogenous variables (age, race, gender, insurance variables, and so on).

3. Results

Table 3 presents data on cigarette prices. These prices are available each year. The year 1997 was selected to remain consistent with the CTS data from 1996–1997. The relative ranking across states was virtually the same in 1996.

We present prices for 50 states and the District of Columbia. The CTS, in its 60 communities across the United States, did not include communities from 15 states (Alaska, Delaware, Hawaii, Idaho, Iowa, Kansas, Mississippi, Montana, Nebraska, New Hampshire, New Mexico, North Dakota, Rhode Island, South Dakota, and Vermont). Nevertheless, the 60 communities are representative of the nation [26].

Within Table 3, the three highest prices are in Washington ($2.73), Massachusetts ($2.59), and Michigan ($2.43). The three lowest prices are in Kentucky ($1.56), Virginia ($1.62), and Georgia ($1.62).

Table 4 presents descriptive statistics. The mean and standard deviation for cigarette price were $2.04 and $.31. The variable for cigarettes per day included the majority (76%) of respondents with “0” as a response. The mean for physical function in the SF-12 was 47.99, with a standard deviation of 10.45. Higher scores for functional status reflected better health. The CTS did not collect information on functional status for persons younger than 18.

Table 3 presents results on regressions that accounted for geographic clusters. Three regressions were run. In regression number 1, the dependent variable was cigarettes per day. The key exogenous variable was cigarette price, which

Table 1
Linear regression results accounting for geographic clusters

	Estimated Coefficient and (two-tailed *P*-value)		
Instrument and endogenous variables	Regression 1, number of cigarettes smoked[a]	Regression 2, functional status (SF-12)[a]	Regression 3, functional status (SF-12)[b]
Cigarette price	−.935** (<.002)		
Cigarettes per day		−.054** (<.001)	
Predicted cigarettes per day			−.587* (0.047)
Other exogenous variables			
Age	.255** (<.001)	−.273** (<.001)	−.136** (.080)
Age square	−.003** (<.001)	.002** (<.001)	0.0002 (.064)
African—American	−2.946** (<.001)	−1.327** (<.001)	−2.833** (<.001)
Other race	−.589 (.105)	−.706** (.023)	−1.050** (.014)
Hispanic	−4.220** (<.001)	.345 (.214)	−1.927 (.157)
Female	−1.967** (<.001)	−.390** (<.002)	−1.433** (<.014)
Years of school	−.448** (<.001)	.437** (<.001)	.192 (.169)
Married, spouse present	−1.618** (<.001)	.632** (<.001)	−.212 (.651)
Family income	−.00001** (<.001)	.00001** (<.001)	.00001* (<.018)
Employed	−.643** (<.001)	4.217** (<.001)	3.874** (<.001)
Number of children	−.306** (<.001)	.112 (.067)	−.052 (.649)
Medicare	.257 (.319)	−3.774** (<.001)	−3.632** (<.001)
Medicaid	1.923** (<.001)	−4.968** (<.001)	−3.993** (<.001)
Military	1.77** (<.001)	−2.198** (<.001)	−1.270 (.125)
No insurance	2.143** (<.001)	−.0423 (.860)	1.102 (.131)
Constant[c]	12.107** (<.001)	48.522** (<.001)	54.003** (<.001)
R^2 *F*-prob	0.086** (<.001)	.207** (<.001)	.018** (<.001)

* Indicates significance at .05 level in two-tailed test.

** Indicates significance at .01 level in two-tailed test.

[a] For regressions 1 and 2, *stata* regression command *svyreg* was used.

[b] For regression 3, *stata* instrumental variable regression corrected for weights and clusters *svyivreg* was used. The first stage regression used all exogenous variables, not just cigarette price, to generate the predicted cigarette price variable. Schem 4/20.

[c] Following binary variables omitted: White and no response for race, all other marital state categories, male, private insurance.

was strongly and negatively correlated, as predicted by economic theory. The *P*-value was smaller than .002. The estimated standard error and *t*-statistic (not shown) were .309 and −3.03, respectively. Elasticity, evaluated at the mean for price and quantity of cigarettes, was −.47. Elasticity measures the percentage change in cigarettes per day associated with a percentage change in price.

The other covariates (control variables) perform as expected, thus lending credence to our data and manner in which we specified the equations [39]. CTS women smoked less than men. Education was inversely correlated with smoking. Married persons smoked less than persons in other marital categories. Income was inversely related to smoking. Greater numbers of children in the family was inversely correlated with smoking. Smoking increased with age, but at a decreasing rate (coefficient on age square was negative).

In regression number 2, physical functional status was the dependent variable. The key covariate was cigarettes per

Table 2
Tobit and linear regression results accounting geographic clusters

	Estimated coefficient and (two-tailed *P*-value)		
Instrument and endogenous variables	Tobit regression 1 cigarettes per day[a]	Linear regression 2 functional status (SF-12)[b]	Linear regression 3 functional status (SF-12)[b]
Cigarette Price	−2.322* (.023)		
Cigarettes per day		−.054** (<.001)	
Predicted cigarettes per day[c]			−.237* (.021)
Other Exogenous Variables[d]			

* Indicates significance at .05 level in two-tailed test.

** Indicates significance at .01 level in two-tailed test.

[a] For regressions 1, *stata* regression command *intreg* was used.

[b] For regression 2 and 3, *stata* command *svyreg* was used. *N* = 34324, includes DC (Schem 6/1/03).

[c] Standard error not estimated with IV technique.

[d] All other exogenous variables, age through no insurance, were also included in each of these regressions. Following binary variables omitted: White and no response for race, all other marital state categories, and male, private insurance.

Table 3
Cigarette prices, 1997

State	Average price ($) per pack
Alabama	$1.81
Alaska[a]	$2.97
Arizona	$2.31
Arkansas	$1.87
California	$2.08
Colorado	$1.86
Connecticut	$2.18
Delaware	$1.79
District of Columbia	$2.37
Florida	$1.92
Georgia	$1.62
Hawaii	$2.60
Idaho	$1.87
Illinois	$2.02
Indiana	$1.67
Iowa	$1.94
Kansas	$1.84
Kentucky	$1.56
Louisiana	$1.86
Maine	$2.20
Maryland	$1.97
Massachusetts	$2.59
Michigan	$2.43
Minnesota	$2.20
Mississippi	$1.76
Missouri	$1.74
Montana	$1.75
Nebraska	$1.95
Nevada	$2.01
New Hampshire	$1.97
New Jersey	$2.08
New Mexico	$1.84
New York	$2.29
North Carolina	$1.68
North Dakota	$2.05
Ohio	$1.73
Oklahoma	$1.84
Oregon	$2.34
Pennsylvania	$1.91
Rhode Island	$2.32
South Carolina	$1.65
South Dakota	$1.93
Tennessee	$1.68
Texas	$2.02
Utah	$2.23
Vermont	$2.14
Virginia	$1.62
Washington	$2.73
West Virginia	$1.71
Wisconsin	$2.11
Wyoming	$1.69

Source: Centers for Disease Control and Prevention.

[a] The following states do not have individuals from CTS residing in them: Alaska, Delaware, Hawaii, Idaho, Iowa, Kansas, Mississippi, Montana, Nebraska, New Hampshire, New Mexico, North Dakota, Rhode Island, South Dakota, and Vermont.

day, which was strongly and negatively correlated with physical functional status, as expected. The *P*-value was less than .001. The standard error and *t*-statistic (not shown) were .0081 and −6.71, respectively.

There are additional noteworthy results in regression 2. Functional status decreased with age, but at a decreasing rate. African–Americans and persons in the "other race" category reported lower functional status than Whites. Women and persons insured by Medicare, Medicaid, or the military reported lower status than men and persons with private insurance. Education, being married with spouse present, income, being employed, and number of children in the family, were all positively and strongly related to higher functional status. Again, these results were consistent with the literature on functional status and good health in general [27–29]. In regression number 3, physical functional status was again the dependent variable. This regression adjusted standard errors in accordance with the IV technique. With the exception of age squared, married, schooling, and military insurance, every covariate that was statistically significant in regression 2 was statistically significant in regression 3, and carried the same sign. The key covariate was the IV: predicted cigarettes per day. It was negatively correlated with functional status and had a *P*-value of .047 (standard error = 0.2953; *t*-statistic = −1.99). The estimated coefficient, −.587, was roughly 10 times larger in absolute value than the cigarettes-per-day coefficient in regression 2 (−.054).

Table 2 presents results on regressions that correspond to those in Table 1: a tobit regression (column 1) similar to linear regression 1 in Table 1; a linear regression (column 2) identical to regression 2 in Table 1; and a linear regression (column 3) similar to linear regression 3 in Table 1. But in Table 2, only results on key variables are presented. All three partial correlations in Table 2 generated *P*-values below .05 in two-tailed tests. In column 1, the tobit coefficient on cigarette price was −2.322, which is more than double the size of the linear regression coefficient in Table 1 of −.935. However, these coefficients are not directly comparable [33]. "Marginal effects" from the tobit are comparable [33]. We estimated the marginal effect to be −0.557. The elasticity associated with −.557 is −0.28. The estimated coefficient on predicted cigarettes per day (the IV) was −.237. This is less than half the size of the coefficient in Table 1 that did not use Tobit, but over four times the size of the coefficient in column 2 that did not use the IV technique.

Whereas there are only 36 different values for cigarette prices, there are likely to be many more values for predicted cigarettes per day. Whenever an exogenous variable such as age or years of school, and so on, changes its value (whenever we compare two different people in the CTS), predicted cigarettes will also change (will be different for the different people). This would not be true if cigarette price *only* were the predictor of cigarettes per day.

The magnitude of the effect of cigarettes on functional status can be assessed with comparing a pack-a-day smoker with a nonsmoker. This would require multiplying 20 (cigarettes) with the estimated coefficient, −.587 (see Table 1) and −.237 (see Table 2) to yield roughly −11.7 and −4.7. These −11.7 and −4.7 are indeed large numbers given that

Table 4
Descriptive statistics, CTS household sample, persons age 30+ sample size = 34,326

	Continuous Variables				Binary Variables	
Variable	Mean	Standard deviation	Min.	Max.	Number in category	% in category
Instrument						
Cigarette price	$2.04	$0.31	$1.56	$2.73		
Dependent (endogenous) variables						
Cigarettes per day[a]	4.05	8.99	0	96		
Physical functional status, SF-12	47.99	10.45	10.37	69.04		
Independent (exogenous) variables						
Age	50.98	14.28	30	91		
African–American					3,879	11.3%
Other Race (Native/Asian/Pacific)					1,476	4.3%
White, non-Hispanic					25,745	75.2%
Hispanic, any race					3,124	9.1%
No race/ethnic response					102	<0.1%
Female					18,760	53.1%
Years of schooling	12.92	2.71	6	19		
Married, spouse present					23,107	64.7%
Family income	$44,035	$35,796	0	$150,000		
Employed					21,489	58.3%
Number of children	0.70	1.10	0	7		
Private insurance					22,559	65.7%
Medicare					7,106	20.7%
Medicaid					752	2.2%
Military and other public					721	2.1%
Uninsured					3,188	9.3%

[a] Cigarettes-per-day was constructed. Number with "0" is 26,318 or 76% of the sample.

the mean of SF-12 is 48 and the standard deviation is 10.5. Assuming functional status is roughly normally distributed, z-values of 1.11 (for -11.7) and 0.45 (for -4.7) suggest the smoker had functional status of 37 to 17% worse than that of a nonsmoker.

4. Discussion

4.1. Implications

When comparing results on cigarettes per day and predicted cigarettes per day in regressions 2 and 3 in both Tables 1 and 2 two features stand out. First, P-values (and standard errors) are larger for the predicted cigarettes than the actual cigarettes. Second, the coefficients, in absolute value, are an order of magnitude larger for predicted cigarettes than for actual cigarettes. The IV method thus suggests that the true effect of smoking on health is larger than conventional methods have estimated.

This result was not expected, but can be explained. As mentioned above, reverse causality and spillover effects could produce an IV estimate that exceeds the conventional estimate. The reverse causality could occur when a person quits smoking because he or she is in poor health. That is, many people who are currently disabled may have quit smoking sometime in the past. Our CTS data are restricted to smoking in the past 30 days. Spillover effects could involve second-hand smoke and social stigma. The IV technique will include these spillover effects, especially if the instrument is a geographical variable, as ours is [30–32].

4.2. Literature review

Stuck et al. [40] recently reviewed the literature on the risk factors associated with functional status decline. Twenty-eight factors made it on their short list of the most studied risk factors. Of these 28, only 12 generated the highest rating for a risk factor that was consistently and strongly predictive of functional decline. Smoking was one of 12. None of the studies reviewed by Stuck et al. [40], however, used the IV technique.

For a number of reasons, numerical estimates of association between smoking and functional decline cannot be easily summarized across studies. First, smoking and functional decline studies do not always measure smoking the same way (number of cigarettes, binary smoking, pack-years). Second, functional decline is not measured the same way (SF-12, Stanford Health Assessment Questionnaire, continuous measures, binary measures). Third, statistical techniques differ (linear regression, logistic regression, ordered probit regression). Fourth, samples differ (different age groups, men only, women only). Fifth, we cannot directly compare our estimated coefficients with those in the literature. Epidemiologic research tends to focus on estimating odd ratios and relative risks when both the dependent and independent variables are binary. But when they are continuous, as they are here, epidemiologists tend not to focus on the size of the coefficient and sometimes do not ever report it.

There is a growing body of literature on IV applications in medical research and epidemiology [1–9]. It is worth noting that two of these studies rely on *geography* to help create an instrument. McClellan, McNeal, and Newhouse [1] use distance from the hospital. Earle et al. [4] use differential use of chemotherapy across metropolitan areas. Moreover, both of these studies have generated a great many citations. A number of the earlier IV applications provided little rationale for their choice of instruments [5–7]. More recent applications [2,3] give more attention to justifying their choice of instruments. However, the justification relies more on statistical tests (Hausman Test) than on theoretic rationale.

Our large estimate IV effects have precedent in the emerging literature using IV techniques to estimate the effects of education on health. This literature generally finds that the IV technique estimates larger effects for education than conventional techniques [30–32]. Finally, our estimated cigarette price elasticities (−.47, −.28) compare favorably to others in the literature (−0.46 in Hu et al. [41] and −0.3 to −0.5 in Keeler et al. [42]).

4.3. Strengths and limitations

We begin with strengths. Few, if any, IV studies by economists account for geographic clustering [43–45]. But this is important, as our study suggests, especially if the IV is geographically specific. Another major strength is that our instrument, cigarette price, is simple and plausible. A third strength is that the SF-12, smoking, and cigarette price are all continuous variables. Problems are introduced if either treatment or outcome is not continuous [15]. A fourth strength is the CTS data set, which is popular and highly regarded by medical researchers [46,47].

There are limitations. The IV estimator is consistent but not unbiased. This underscores the necessity of using large samples. Our sample size was 34,288. We selected ages 30 and over. Many studies of functional decline look at persons age 50 and over. However, to increase sample size, we selected age 30. A number of studies on functional status have looked at middle-aged persons (Haapanen et al. [48], aged 35–63 years; Philappaerts, Lefevre [49], aged 30–40 years; Idler et al. [50], aged 25–74 years; Michel et al. [51], aged 25–74 years; Ueda et al. [52], aged 40–75 years; Garraway et al. [53], aged 40–79 years).

Another limitation involves mobility. People move. Current state of residence may not have much influence on a person's smoking habits. This is especially true for retirees who may have relocated to a new state. Persons typically begin smoking in their teen years. On the other hand, they quit anytime after that. There are only 36 different data points for cigarette price. However, normally, this would suggest it is difficult to find low *P*-values. But low *P*-values were found. That is, the regression technique already accounts for the fact that there are so few different observations on cigarette price. Moreover, when all exogenous variables were included in the cigarette price regression, the predicted cigarettes per day had much more variation than just 36 values.

District of Columbia residents may leave and buy cigarettes in Virginia or Maryland. This can happen in any state. Evans and Ringle [43] found this bootlegging effect was small. Lewit, Coate, and Grossman [54], on the other hand, found bootlegging was a significant problem among youth. If there is significant bootlegging from low tax states to high tax states then both our conventional model estimates and the IV model estimates will have a downward bias on the effects of cigarettes on physical functional status.

An incorrect criticism of our study is that we should have used taxes, not prices. Economic theory suggests that it is prices (which include taxes) that are the most relevant for the consumer. Taxes are relevant for policy. But this is not the question addressed in this study.

The CTS has no data on specific health problems such as whether the respondent has cancer, circulatory disease, COPD, etc. Nor does the CTS have data on other health habits such as drinking or exercising or body mass index. The National Health Interview Survey (NHIS) has these data, but it is the policy of the NHIS not to provide geographic state data on individuals to researchers. The lack of health data sets with geographic information is problematic for implementing IV techniques.

Another potential limitation involves the inclusion of the four variables reflecting insurance status. Insurance status may be endogenous. Smokers may be risk takers, in general, and therefore, more likely to go uninsured. However, insurance status is likely to affect health, that is, to be included in the structural equation explaining health. As a result, the reduced form equation explaining cigarettes per day would still have the insurance status variables as covariates. We, nevertheless, allowed for the possibility that the insurance status variables might be influencing our results. We ran regressions excluding them. Our fundamental findings remained unchanged: predicted cigarettes per day generated *P*-values below .05, and the IV estimates were at least four times larger than the non-IV estimates.

Finally, the standard error on predicted cigarettes per day in Table 2 did not adjust for the IV technique. But we are not aware of any IV Tobit program that also adjusts for geographic clustering. We did test whether the standard error in a linear regression of SF-12 on predicted cigarettes per day and exogenous variables that did not adjust the variance–covariance matrix (*svyreg*) was appreciably different from the standard error in a similar linear regression that did adjust the variance–covariance matrix. The standard error on predicted cigarettes per day in the unadjusted regression was .2535 and .2953 in the adjusted regression. This is only a 16% increase. A similar increase on the standard error in Table 2 would result in a *P*-value far below .05. We, therefore, do not think it is likely that estimated standard of errors in Table 2 would alter our fundamental findings.

5. Conclusion

We have used a large national sample to investigate the effects of smoking on functional status with the IV technique. We hope our instrument, cigarette price, and our intuitive explanations of the technique together with other studies [1–9], persuade some physicians and epidemiologists of the usefulness of instrumental variables.

Acknowledgments

This study was funded in part by a grant from the National Institute for Occupational Safety and Health (OH07338-01). We are indebted to Tonya Lange and Dina McHugh for word processing.

References

[1] McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction reduce mortality? JAMA 1994; 272:859–66.
[2] Chen Q, Kane RL. Effects of using consumer and expert ratings of activities of daily living scale on predicting functional outcomes of post acute care. J Clin Epidemiol 2001;54(4):334–42.
[3] Rizzo JA, Coady MA, Elefteriades JA. Procedures for estimating growth rates in thoracic aortic aneurysms. J Clin Epidemiol 1998;51 (9):747–54.
[4] Earle CC, Tsai JS, Gelber RD, Weinstein MC, Neumann RP, Weeks JC. Effectiveness of chemotherapy for advanced lung cancer in the elderly: instrumental variable and propensity analysis. J Clin Oncol 2001;19(4):1064–70.
[5] Neale MC, Walters E, Heath AC, Kessler RC, Perusse D, Eaves LJ, Kendler KS. Depression and parental bonding—cause, consequence, or genetic covariance. Genet Epidemiol 1994;11(6):503–22.
[6] Ruel MT, Habicht JP, Pinstrupandersen P, Grohn Y. The mediating effect of maternal nutrition knowledge on the association between maternal schooling and child nutritional-status in Lesotho. Am J Epidemiol 1992;135(8):904–14.
[7] Ruel MT, Rivera J, Habicht JP, Martorell R. Differential response to early nutrition supplementation – long-term effects on height at adolescence. Int J Epidemiol 1995;24(2):404–12.
[8] Gifford AL, Bormann JE, Shively MJ, Wright BC, Richman DD, Bozzette SA. Predictors of self-reported adherence and plasma HIV concentrations in patients on multidrug antiretroviral regimens. J Acquir Immune Defic Syndr 2000;23(5):386–95.
[9] Leigh JP, Ward MM, Fries JF. Reducing attrition bias with an instrumental variable in a regression model. Stat Med 1993;12:1025–8.
[10] Brooks JM, McClellan M, Wong HS. The marginal benefits of invasive treatments of acute myocardial infarction: does insurance coverage matter? Inquiry 2000;37(1):75–90.
[11] Hadley J, Rabin D, Epstein A, Stein S, Rimes C. Post hospitalization home health care use and changes in functional status in a Medicare population. Med Care 2000;38(5):494–507.
[12] Harris KM, Remler DK. Who is the marginal patient? Understanding the instrumental variables estimates of treatment effects. Health Serv Res 1998;33(5):1337–60.
[13] Johnson RW, Crystal S. Uninsured status and out-of-pocket costs at midlife. Health Serv Res 2000;35(5):911–32.
[14] Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 2000;29(4):722–9.
[15] Angrist JD, Krueger AB. Split sample instrumental variables estimates of the return to schooling. J Bus Econ Stat 1995;13(2):225–35.
[16] Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. J Am Stat Assoc 1995; 90:443–50.
[17] Angrist JD, Krueger AB. Instrumental variables and the search for identification: from supply and demand to natural experiments. J Econ Perspect 2001;15(4):69–85.
[18] Hahn JY, Hausman J. Weak instruments: diagnosis and cures in empirical econometrics. Am Econ Rev 2003;93(2):118–25.
[19] Viscusi WK, Magat WA, Huber J. Smoking status and public responses to ambiguous scientific risk evidence. South Econ J 1999;66(2): 250–70.
[20] Mackay J, Eriksen M. The tobacco atlas. Geneva, Switzerland: World Health Organization; 2002.
[21] Gauchard GC, Chan N, Touron C, Benemghar L, Dehaene D, Perrin P, Mur JM. Individual characteristics in occupational accidents due to imbalance: a case–control study of the employees of a railway company. Occup Environ Med 2003;60(5):330–5.
[22] Murray CJL, Lopez AD. Alternative projections of mortality and disability by cause 1990–2020. Global burden of disease study. Lancet 1997;349(9064):1498–505.
[23] Callahan LF, Rao J, Boutaugh M. Arthritis and women's health: prevalence impact, and prevention. Am J Prev Med 1996;12(5):401–9.
[24] Ware JE, Kosinski M, Keller CD. SF-12: how to score the SF-12 physical and mental health summary scales. 2nd ed. Boston: The Health Institute; 1995. p. 51.
[25] Imbens GW. Sensitivity to exogeneity assumptions in program evaluation. Am Econ Rev 2003;93(2):126–32.
[26] Kemper P, Blumenthal D, Corrigan JM, Cunningham PJ, Felt SM. The design of the community tracking study in a longitudinal study of health system change and its effects on people. Inquiry 1996;33: 195–206.
[27] House JS, Lepkowski JM, Kinney AM, Mero RP, Kessler RC, Herzog AR. The social stratification of aging on health. J Health Soc Behav 1994;35(3):213–34.
[28] Parker MG, Thorslund M, Lundberg O, Karehalt I. Predictors of physical function among the oldest old—a comparison of three outcome variables in a 24-year follow-up. J Aging Health 1996;8(3): 444–60.
[29] Liu X, Liang J, Muramatsu N, Sugisawa H. Transition in functional status and active life expectancy among older people in Japan. J Gerontol 1995;50(6):S383–94.
[30] Grossman M. Education and nonmarket outcomes. In: Hanushek E, Welch F, editors. Handbook of the economics of education. Amsterdam: Elsevier Science; 2003.
[31] Lleras-Muney A. The relationship between education and adult mortality in the U.S. Cambridge, MA: National Bureau of Economic Research Working Paper #8986; 2002.
[32] Currie J, Moretti E. Mother's education and the intergenerational transmission of human capital: evidence from college openings and longitudinal data. Cambridge, MA: National Bureau of Economic Research Working Paper #9360; 2002.
[33] Greene WH. Econometric analysis. New York: Macmillian; 2000.
[34] Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. Annu Rev Public Health 1998;19:17–34.
[35] Berger MC, Leigh JP. Schooling, self-selection and health. J Hum Resource 1989;24:433–55.
[36] Kennedy P. A guide to econometrics. 3rd ed. Cambridge, MA: M.I.T. Press; 1992.
[37] Wallace TD, Silver JL. Econometrics. Reading, MA: Addison-Wesley; 1988.
[38] Leigh JP, Fries JF. Tobit, fixed effects, and cohort analyses of the relationship between severity and duration of rheumatoid arthritis. Soc Sci Med 1983;36(11):1495–502.
[39] Chaloupka FJ, Wechsler H. Price, tobacco control policies, and smoking among young adults. J Health Econ 1997;16(3):359–73.
[40] Stuck AE, Walthert JM, Nikolaus T, Bula CJ, Hohmann C, Beck JC. Risk factors for functional status decline in community-living elderly people: a systematic literature review. Soc Sci Med 1994;48 (4):445–69.
[41] Hu TW, Ren QF, Keeler TE, Bartlett J. The demand for cigarettes in California and behavioral risk factors. Health Econ 1995;4(1):7–14.

[42] Keeler TE, Hu TW, Bartlett PG, Manning WG. Taxation, regulation, and addition. A demand function for cigarettes based on time-series data. J Health Econ 1993;12(1):1–18.

[43] Evans WN, Ringel J. Can higher cigarette taxes improve birth outcomes? J Public Econ 1999;72(1):135–54.

[44] Markowitz S. The price of alcohol, wife abuse, and husband abuse. South Econ J 2000;67:279–303.

[45] Woolridge JM. Cluster-sample methods in applied econometrics. Am Econ Rev 2003;93(2):133–8.

[46] St. Peter R, Reed MC, Kemper P, Blumenthal O. Changes in the scope of care provided by primary care physicians. N Engl J Med 1999; 341:1980–5.

[47] Landon BE, Reschovsky J, Blumenthal O. Changes in career satisfaction among primary care and specialist physicians, 1997–2001. JAMA 2003;289(4):442–9.

[48] Haapanen N, Miilunpalo S, Vuori I, Oja P, Pasanen M. Characteristics of leisure time physical activity associated with decreased risk of premature all-cause and cardiovascular disease mortality in middle-aged men. Am J Epidemiol 1996;143(9):870–80.

[49] Philippaerts RM, Lefevre J. Reliability and validity of three physical activity questionnaires in Flemish males. Am J Epidemiol 1998;147 (10):982–90.

[50] Idler EL, Russell LB, Davis D. Survival, functional limitations, and self-rated health in the NHANES I epidemiological follow-up study, 1992. Am J Epidemiol 2000;152(9):874–83.

[51] Michel A, Kohlmann T, Raspe H. The association between clinical findings on physical examination and self-reported severity of back pain – results of a population-based study. Spine 1997;22(3):295–303.

[52] Ueda T, Tamaki M, Kageyama S, Yoshimura N, Yoshida O. Urinary incontinence among community-dwelling people aged 40 years or older in Japan: prevalence, risk factors, knowledge and self-perception. Int J Urol 2000;7(3):95–103.

[53] Garraway WM, Russell EBAW, Lee RJ, Collins GN, McKelvie GB, Hehir M, Rogers CAN, Simpson RJ. Impact of previously unrecognized benign prostatic hyperplasia on the daily activities of middle-aged and elderly men. Br J Gen Pract 1993;43(373):318–21.

[54] Lewit EM, Coate D, Grossman M. The effects of government regulation on teenage smoking. J Law Econ 1981;24(3):545–69.