The Journal of Risk and Insurance

DOI: 10.1111/j.1539-6975.2009.01336.x

Medical Insurance Coverage and Health Production Efficiency

Laurie J. Bates
Kankana Mukherjee
Rexford E. Santerre

Abstract

Conventional economic theory predicts that medical insurance coverage causes an inefficient production of health because of *ex ante* and *ex post* moral hazard effects. However, no research has empirically examined the magnitude of the inefficiency. This study empirically examines the impact of medical insurance on the technical efficiency of health production at the metropolitan level. The underlying health production function allows for preventive care, curative care, and behavioral factors. Data envelopment analysis determines relative technical efficiency. The multiple regression results indicate that insurance coverage generates inefficiency but the efficiency loss appears to be relatively small on the extensive margin.

Introduction

Among various issues, high medical care costs dominate much of the political debates taking place during the recent presidential election. Given that medical costs now comprise 16 percent of gross domestic product, the attention seems to be well deserved. However, many health economists and policymakers are not entirely convinced that the medical spending represents an economic burden on society. For example, Cutler and McClellan (2001) show that the aggregate benefits of just two major medical technologies outweighed all medical care costs during the last two decades.

While medical care may pay for itself on average, conventional theory suggests that medical insurance may generate inefficiencies at the margin because of both *ex ante* and *ex post* moral hazard effects (Ehrlich and Becker, 1972; Breyer and Zweifel, 1997). *Ex ante* moral hazard refers to the effect of insurance on preventive actions taken before states of sickness occur, such as healthy lifestyles, preventive care, or early detection

Laurie J. Bates is with the Department of Economics, Bryant University, Smithfield, RI 02917. Kankana Mukherjee is with the Economics Division, Babson College, 231 Forest Street, Babson Park, MA 02457. Rexford E. Santerre is with University of Connecticut, School of Business, Department of Finance, 2100 Hillside Road, Unit 1041, Storrs, CT 06269-1041. Rexford E. Santerre can be contacted via e-mail: rsanterre@business.uconn.edu. We thank Stephen Neun and John Nyman for their comments on an earlier version of this article. We also thank the anonymous referees of this journal for their many helpful suggestions for improving the article.

of diseases. In contrast, *ex post* moral hazard captures the size of the financial loss or cost of treatment because of insurance coverage and may show up in terms of longer hospitals stays and additional visits to health-care providers.

If preventive and curative services are substitutes in the production of health, the theory generally views medical insurance as lowering the out-of-pocket price of curative inputs relative to the price of preventive inputs and thereby distorting the choice of inputs. Prevention declines, the probability of sickness rises, and an increased consumption of medical care occurs. The medical costs of maintaining a given level of health increase as a result. For example, Newhouse (1993) finds that households in low coinsurance health plans received more medical care yet possessed virtually the same level of health as those households in high coinsurance plans, *ceteris paribus*.

Because of "nine limiting conditions," however, some researchers note that medical insurance may not have a major impact on the efficiency of health production (Crew, 1969; Schlesinger and Venezian, 1986; Pauly and Held, 1990; Kenkel, 2000; Nyman, 2003; Dave and Kaestner, forthcoming). First, health-care providers may possess market power. The resulting restriction of output negates the typical *ex post* moral hazard effect of medical insurance towards overconsumption. Second, the *ex ante* moral hazard effect may be small because medical insurance does not completely cover the utility loss associated with sickness (pain and suffering). Third, preventive inputs may remain attractive because the choice of health inputs actually involves completely preventing versus incompletely curing illness. The attractiveness of preventive inputs, however, is limited by the fact that prevention can never reduce the probability of illness to zero.

Fourth, medical insurance premiums may be risk-rated. While that might be true for individual medical insurance plans, it is not true for the employer-sponsored and public medical insurance plans that cover most people in the United States. Fifth, health insurers such as managed care organizations (MCOs) may invest directly in prevention to reduce the probability of a loss. Sixth, employers may offer subsidized worksite health promotion activities such as smoking cessation programs. Seventh, people may tend to transition frequently between insured and uninsured status. However, people may also game this transition by shifting expensive medical treatments into the insured period. Eighth, a health promotion effect may result as people visit their primary care givers more frequently because of medical insurance and their primary care givers point out cost-effective ways of generating health. Finally, medical insurance may promote efficient *ex post* moral hazard by providing low-income individuals with financial access to life-saving medical care they could not otherwise afford.

Thus, the theoretical literature to date generally views medical insurance as distorting the choice of health inputs. However, some researchers believe that the distortionary impact of medical insurance on the efficient production of health may be limited by these nine conditions. Yet, no previous study has empirically examined the effect of medical insurance on health production efficiency. Information of this kind would be valuable to policymakers when deliberating on policies expanding medical insurance coverage to those currently uninsured. If large efficiency losses are expected, policy cost estimates should be increased to reflect these losses or medical insurance policies should be designed to curtail such effects.

This study fills this void in the literature by using data envelopment analysis (DEA) to measure aggregate health production efficiency for a number of metropolitan areas for the years 2002 through 2006. Preventive care, behavioral, and curative care health inputs are specified in the underlying health production function. The DEA finds that most metropolitan areas are similarly efficient at producing health. That is, while some metropolitan areas operate with 6 to 12 percent relative technical inefficiency, the average metropolitan area is only 2.3 percent inefficient at producing health. The effect of medical insurance coverage on health production efficiency is analyzed in the next phase of the study using multiple regression analysis while testing and controlling for the endogeneity of insurance coverage. As conventional theory suggests, medical insurance coverage reduces the efficiency of health production although the efficiency loss does not appear to be very large. In particular, the results indicate that at most, a 1 percentage point increase in the percentage of the population insured results in only a 0.5 percentage decline in the efficiency of health production, on average.

The next section of this article discusses the conceptual framework, sample of metropolitan areas, variables, and data used in the empirical analysis. The following section provides the specifics of and results from the DEA. The multiple regression analysis exploring the impact of medical insurance on health production efficiency is taken up in the "Analysis of Medical Insurance Coverage and Health Production Efficiency" section. The last section offers a summary, some policy implications, and ideas for future research in this area.

CONCEPTUAL MODEL, SAMPLE, DATA, AND VARIABLES

Beginning with Auster, Leveson, and Sarachek (1969) and elaborated upon greatly by Grossman (1972), economists have considered that the creation and maintenance of health involves a production process. This theory treats consumers as producing their own health by combining and employing various health inputs. The process can be captured in a health production function that indicates the maximum amount of health that can be generated from a mix of inputs such as curative care, preventive care, behavioral factors, socioeconomic status, and environmental setting, conditioned upon baseline health and a given state of technology. Like in any production process, some substitutability or complementarity may exist among the inputs in the health production function.

Because this article examines *ex ante* and *ex post* moral hazard effects, we focus mainly on the use of behavioral factors and curative and preventive care in the health production process in the first stage of the DEA. Controls for other less discretionary factors influencing health status, such as socioeconomic status (e.g., education and income) and environmental conditions, which are more beyond the control of the individual consumer at a particular point in time, as well as exogenous factors such as age, race, and chronic medical conditions, are made in the second stage of the DEA.[1]

For a number of reasons, we evaluate and explain the efficiency of health production at the metropolitan level. First, Zweifel and Manning (2000, p. 415) note that while *ex ante* moral hazard may be "hardly noticeable at the individual level, it may cause

[1] Of course, people can change their jobs, relocate to a healthier climate or location (i.e., with less crime or pollution), or become more highly educated in the longer run.

the number of sickness episodes to differ a great deal in the aggregate." Second, baseline health, resulting from someone's genetic predisposition and other factors, plays an important role in the estimation of health production functions but remains hard to observe and measure. Use of aggregate data helps to minimize this baseline health problem because the distribution of underlying health conditions may be more similar across metropolitan areas than across more disaggregated levels, at least within a particular region (Bates, Mukherjee, and Santerre, 2006). Third, the necessary data for medical insurance availability, preventive care, lifestyle factors, and health status are reasonably available over time and across metropolitan statistical areas (MSAs) from the Center for Disease Control. Use of a panel data set helps to control for unobservable heterogeneity among MSAs.

We sought to include the greatest number of MSAs for as many years as possible in our sample so the results would be representative of the entire United States and the analysis would capture the effects of technological change on health production. But data on behavioral risk factors are consistently available for only 60 metropolitan areas for the years 2002 to 2006, resulting in a maximum total of 300 MSA observations. While the number of observations may appear to be low and perhaps unrepresentative of the entire U.S. population, these 60 MSAs reflect about one-third of the population in the United States and range geographically from Portland, Maine in the Northeast to Tucson, Arizona in the Southwest and from Portland, Oregon in the Northwest to Miami, Florida in the Southeast. Consequently, the sample of MSAs may be fairly representative of the urban population in the United States.

In the empirical analysis, the percentage of the population reporting as being in good or better health represents the measure of health status or output in the underlying health production function. Studies find that poor self-reported health status correlates strongly with mortality after controlling for other factors (Idler and Benyamini, 1997), although mood can sometimes create distortions in self-assessment (Schmidt et al., 1996). Despite that limitation, Rosen and Wu (2004) find that their results regarding the effect of health status on portfolio choice remains unchanged when other measures of health status are used or controls are made for mental health status. In addition, the second stage of the DEA controls for some chronic conditions, disability rate, age, poverty, and other demographic and economic factors that may affect mental well-being.

Five behavioral and preventive care health inputs are specified in the underlying health production function. They are the percentage of the population that (1): does not binge drink, (2): has never smoked, (3): exercised within the last month, (4): is neither overweight nor obese, and (5) is 65 years of age and older and received a flu shot.[2] The number of hospital inpatients days and outpatient visits per capita serve as the two measures of curative care health inputs. Other measures of outpatient care

[2] Like any type of medical care, a flu shot can be treated as an output from a medical care production standpoint. Within this context, various types of inputs such as physicians and nurses produce units of medical care. At the same time, medical care represents an input from a health production perspective. That is, the typical consumer combines and transforms behavioral, preventive care, and curative care inputs into some final output of health. Thus, to the consumer, a flu shot represents an input in the health production process.

utilization such as physician consultations outside the hospital are unavailable at the metropolitan level. The hospital utilization data come from the American Hospital Association (AHA). Unfortunately, the Bureau of the Census changed its definition of the MSA in 2003 but the AHA did not incorporate the change until 2005. Thus, data on the two curative care measures are unavailable for some MSAs during the earlier years.

The Appendix provides descriptive statistics for the various variables described above as well as for others used in the remainder of the analysis. Averaged across the sampled MSAs, approximately 86 percent of the population reported themselves as being in good or better health. Health status varies widely, however, with 28 percent of the population in at least one MSA reporting as being in fair or poor health. The relatively high exercise but low not-fat statistics are worth mentioning because these two preventive care measures seem at odds with one another. Finally, the relatively low inpatient days per capita and high outpatient visits per capita reflect the trend of hospitals to respond to managed care and regulatory pressures by treating patients more on an outpatient than inpatient basis.

DEA OF MSA HEALTH PRODUCTION EFFICIENCY

In the production efficiency and productivity literature, first introduced by Debreu (1951), Farrell (1957), and Koopmans (1951), efficiency represents a normative measure and is defined as the ratio of the *optimal* input bundle to the *actual* input bundle or as the ratio of the *actual* output to the *optimal* output. To measure the "optimal" input or output, the production frontier must first be specified. In this respect, the literature has primarily followed two approaches—stochastic frontier analysis (SFA) and DEA.[3]

The SFA approach is an econometric approach, which requires the researcher to specify a functional form for the efficient frontier, as well as, the distributional assumptions for the inefficiency and error terms. For this approach, deviations from the efficient frontier are considered as resulting from inefficiency or random factors. The other approach, DEA, developed by Charnes, Cooper, and Rhodes (1978), is a nonparametric approach and uses mathematical programming to create a piecewise-linear *best practice frontier* based on the observed data. For this method, all deviations from the frontier are treated as inefficiency such that the DEA scores capture the relative technical efficiency of the decision-making units (DMUs).

We adopt DEA for estimating the health production efficiency of MSAs because this method estimates the frontier based on a minimum number of regularity conditions with respect to the technology without making arbitrary assumptions about the functional form of the health production frontier.[4] Further, because the efficiency of health production is evaluated at the MSA level rather than at the individual level, the problem that random influences could be confounded with inefficiency is considerably reduced. Random influences are reduced because shocks relating to health status at the individual level, such as motor vehicle accidents, are averaged out to some degree at the metropolitan level. Since its introduction, DEA has been used in several studies

[3] Other approaches, such as Free Disposal Hull (FDH) analysis are sometimes used.
[4] See Ray (2004) for a detailed exposition of DEA.

concerning the health-care sector with Brockett et al. (2004) and Bates, Mukherjee, and Santerre (2006) among the most recent. Hollingsworth, Dawson, and Maniadkis (1999), Chilingerian and Sherman (2004), and Worthington (2004) provide comprehensive reviews of the literature on efficiency measurement in the health-care sector using nonparametric methods.

Consider an industry producing a single output y from a vector of m inputs $x = (x_1, x_2, \ldots, x_m)$. Let y_j represent output and the vector x_j represent the input bundle of the jth decision-making unit. Suppose that input–output data are observed for n DMUs. Then the technology set can be completely characterized by the production possibility set $S = \{(x, y) : y \text{ can be produced from } x\}$ based on a few regularity assumptions of feasibility of all observed input–output combinations, free disposability with respect to inputs and outputs, and convexity. In addition, we assume that the production process exhibits constant returns to scale. This implies that all radial expansions as well as (non-negative) contractions of the feasible input–output combinations are also considered feasible.

The input-oriented technical efficiency measure is defined as the ratio of the optimal (i.e., minimum) input bundle to the actual input bundle of a DMU, for a given level of output, holding input proportions constant.[5] The Charnes, Cooper, and Rhodes (1978) DEA model for measuring the input-oriented technical efficiency of a DMU with the input–output bundle (x_0, y_0) is given by the model below

$$\theta^* = \min \theta, \tag{1}$$

subject to

$$\sum_{j=1}^{n} x_{ij}\lambda_j \leq \theta x_{i0} \quad i = 1, \ldots, m \quad \text{(input constraints)}, \tag{2}$$

$$\sum_{j=1}^{n} y_j\lambda_j \geq y_0 \qquad \text{(output constraint)}, \tag{3}$$

$$\lambda_j \geq 0 \qquad j = 1, 2, \ldots, n. \tag{4}$$

In the above model, the output constraint ensures that the resultant output is no lower than what is actually being produced.[6] An efficient DMU will have $\theta^* = 1$, implying that no equi-proportionate reduction in inputs is possible, whereas an inefficient DMU will have $\theta^* < 1$.

[5] Technical efficiency can also be measured based on output orientation where efficiency is defined as the ratio of the observed output to the optimal (i.e., maximum) achievable output.

[6] As mentioned earlier, our conceptualized model involves a single output (health) and seven inputs (i.e., $m = 7$). Accordingly, the DEA model has one output constraint and seven input constraints.

To construct the health production frontier, which generates the efficiency measures, several assumptions are made. First, our assumption of constant returns to scale makes sense in this context because the empirical analysis focuses on decisions made at the extensive rather than intensive margin.[7] Second, a *sequential* frontier is constructed for each year. The DEA literature distinguishes between three types of frontiers: (1) the *contemporaneous* frontier constructed from only the cross-sectional data from a given period, (2) the *sequential* frontier that treats all current and past observations as feasible, and (3) the *intertemporal* frontier based on observations from all periods in the sample (Tulkens and Vanden Eeckaut, 1995).

Specifically, a *sequential* frontier follows from the assumption that any technical change taking place over the years is *nonregressive* so that input–output bundles observed in the past are deemed feasible in later periods, although input–output bundles observed in subsequent years may not be feasible in those prior years. Thus, starting with a reference sample of 46 observations for the year 2002, the reference sample is successively enlarged by including the observations of one more year to create the frontier of each subsequent period. In this manner, the efficiency for the observations in 2006, our last year of data, is estimated by constructing the frontier based on all 263 observations.

For DEA applications with *sequential* frontier, the production possibility set is defined as $S' = \{(x,y) : x \geqslant \sum_{j=1}^{n} \sum_{t=1}^{t} \lambda_j^t x_j^t ; y \leqslant \sum_{j=1}^{n} \sum_{t=1}^{t} \lambda_j^t y_j^t\}$, where n is the number of units observed and t corresponds to the time period at which the DMU is being evaluated.[8] Thus, in constructing the frontier for any current year t, all observations from the current year as well as the past are considered feasible. Conceptually, this amounts to assuming that there is no technical regress, any technical regress will be assimilated with inefficiency by this construction. When technical progress is experienced, however, the *sequential* frontier will shift outward from one year to the next. The advantage of constructing a *sequential* frontier rather than a *contemporaneous* frontier in the present application is that it affords greater degrees of freedom and therefore more variation in the measured efficiencies. Further, an *intertemporal* frontier for all periods is not constructed because technical progress in health production likely occurred for the period under study—but constructing an *intertemporal* frontier would require us to assume that neither technical regress nor progress is present.

Table 1 provides the annual efficiency score for each MSA over the period from 2002 to 2006 for which data are available. In addition, a frequency distribution of the estimated efficiency scores for the entire sample period is presented in Figure 1. A number of insights can be drawn from the exhibits.

First, health production efficiency was at the lowest level of 0.8753 in Charleston, West Virginia in 2006. At the other extreme, health production efficiency was at its most efficient level for 93 of the 263 annual MSA observations. For these MSAs, it is

[7] Recall that the health insurance variable is measured by the percentage of the population with health insurance coverage. Thus, in this context, the extensive margin refers to more people adopting healthy lifestyles and purchasing insurance whereas the intensive margin means more insurance coverage (e.g., lower coinsurance rate) and more exercise or less alcohol consumption per person.

[8] The constraints for the DEA model are revised accordingly.

TABLE 1
Metropolitan Statistical Areas (MSAs) and Estimated Annual Efficiency Scores

MSA	Efficiency Score 2002	Efficiency Score 2003	Efficiency Score 2004	Efficiency Score 2005	Efficiency Score 2006
Albuquerque, NM	0.926	0.922	0.904	0.929	0.939
Asheville, NC	NA	NA	NA	0.926	0.903
Atlanta-Sandy Springs-Marietta, GA	1.000	1.000	0.985	0.991	0.978
Baltimore-Towson, MD	0.993	0.970	0.971	0.997	0.968
Birmingham-Hoover, AL	0.949	0.962	1.000	1.000	0.938
Boise City-Nampa, ID	1.000	0.990	0.993	1.000	0.990
Bridgeport-Stamford-Norwalk, CT	NA	NA	NA	0.955	0.968
Burlington-South Burlington, VT	1.000	0.993	0.999	1.000	0.980
Charleston, WV	0.953	0.911	0.893	0.988	0.875
Charlotte-Gastonia-Concord, NC-SC	0.940	0.957	0.954	0.954	0.926
Cheyenne, WY	1.000	1.000	0.999	1.000	1.000
Chicago-Naperville-Joliet, IL-IN-WI	1.000	1.000	0.976	1.000	0.974
Cincinnati-Middletown, OH-KY-IN	NA	NA	NA	0.999	0.975
Columbia, SC	0.983	0.956	0.951	1.000	0.941
Denver-Aurora, CO	0.996	0.983	0.998	0.986	0.989
Des Moines-West Des Moines, IA	1.000	1.000	0.960	1.000	0.978
Dover, DE	1.000	1.000	0.973	0.977	1.000
Durham, NC	NA	NA	NA	0.932	0.887
Fargo, ND-MN	1.000	1.000	1.000	1.000	0.991
Greensboro-High Point, NC	NA	NA	NA	0.963	0.926
Hartford-West Hartford-East Hartford, CT	0.967	0.969	0.977	0.964	0.965
Houston-Sugar Land-Baytown, TX	0.993	0.972	0.925	0.970	0.966
Huntington-Ashland, WV-KY-OH	0.895	0.910	0.949	0.970	1.000
Indianapolis-Carmel, IN	0.982	0.941	0.935	0.966	0.959
Jackson, MS	0.953	1.000	0.933	0.962	0.980
Kansas City, MO-KS	0.983	0.962	0.971	1.000	0.953
Las Vegas-Paradise, NV	1.000	1.000	1.000	1.000	0.998
Lincoln, NE	1.000	0.983	0.981	1.000	1.000
Little Rock-North Little Rock, AR	0.920	0.946	0.960	0.929	0.959
Louisville, KY-IN	1.000	0.945	0.947	0.972	1.000
Manchester-Nashua, NH	NA	NA	1.000	1.000	1.000
Memphis, TN-MS-AR	1.000	1.000	1.000	0.993	0.960
Miami-Fort Lauderdale-Miami Beach, FL	NA	NA	1.000	1.000	1.000
Milwaukee-Waukesha-West Allis, WI	0.975	0.989	0.977	0.955	0.980
Minneapolis-St. Paul-Bloomington, MN-WI	1.000	1.000	1.000	1.000	1.000

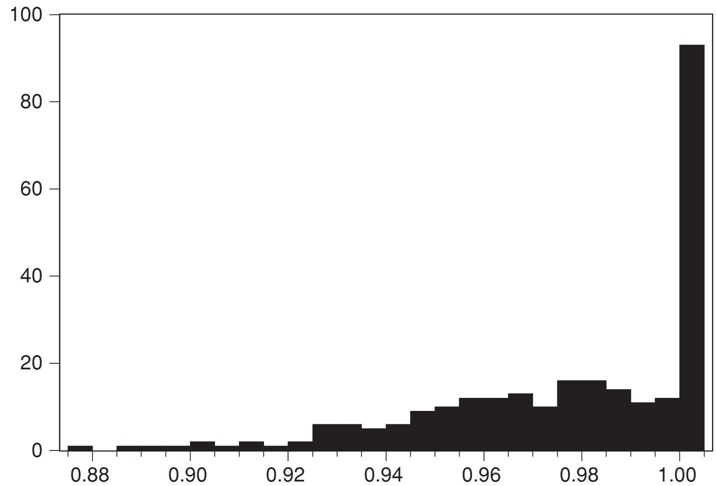
(*Continued*)

TABLE 1
Continued

MSA	Efficiency Score 2002	Efficiency Score 2003	Efficiency Score 2004	Efficiency Score 2005	Efficiency Score 2006
Nashville-Davidson-Murfreesboro, TN	0.981	0.958	0.988	1.000	1.000
New Haven-Milford, CT	NA	NA	NA	0.998	0.949
New Orleans-Metairie-Kenner, LA	1.000	0.994	0.944	0.980	0.960
Ogden-Clearfield, UT	NA	NA	1.000	1.000	1.000
Oklahoma City, OK	1.000	1.000	0.949	0.994	0.942
Omaha-Council Bluffs, NE-IA	0.994	1.000	1.000	0.978	0.978
Phoenix-Mesa-Scottsdale, AZ	1.000	0.979	1.000	0.990	0.940
Pittsburgh, PA	0.957	1.000	0.976	1.000	0.981
Portland-South Portland-Biddeford, ME	NA	NA	NA	0.948	0.960
Portland-Vancouver-Beaverton, OR-WA	1.000	1.000	0.982	0.982	1.000
Providence-New Bedford-Fall River, RI-MA	NA	NA	NA	1.000	1.000
Rapid City, SD	0.965	0.998	1.000	0.974	0.998
Reno-Sparks, NV	0.990	0.964	0.945	0.961	0.954
Richmond, VA	1.000	1.000	1.000	1.000	0.968
Salt Lake City, UT	NA	NA	NA	0.955	0.941
San Francisco-Oakland-Fremont, CA	NA	NA	1.000	1.000	1.000
Sioux Falls, SD	1.000	0.980	0.960	1.000	1.000
Springfield, MA	0.941	1.000	0.911	0.989	0.976
St. Louis, MO-IL	1.000	0.961	0.976	0.967	0.954
Topeka, KS	1.000	1.000	0.981	1.000	0.995
Tucson, AZ	0.980	0.951	0.934	0.933	0.935
Tulsa, OK	1.000	0.965	0.961	0.968	0.940
Virginia Beach-Norfolk-Newport News, VA-NC	NA	NA	1.000	1.000	1.000
Wichita, KS	1.000	1.000	0.986	0.987	0.957
Worcester, MA	1.000	1.000	0.986	0.981	0.977

not possible to reduce the inputs proportionally while still maintaining the level of health. Second, the efficiency score averages 0.977 for the sample, meaning that the typical MSA experienced about 2.3 percent relative inefficiency in the production of health. Finally, 32 of the 60 MSAs transitioned on or off the efficiency frontier over the 5-year period whereas another seven remained on the frontier for the 5 or fewer years of data observed. The remaining 21 MSAs remained off the efficiency frontier throughout the period.

FIGURE 1
Frequency of DEA Scores for Health Production Efficiency



ANALYSIS OF MEDICAL INSURANCE COVERAGE AND HEALTH PRODUCTION EFFICIENCY

The main intent of this article is to examine how medical insurance coverage influences the observed differences in the efficiency of health production across the MSAs as reflected by the estimated efficiency scores in Table 1. Conventional economic theory predicts that medical insurance coverage leads to an inefficient production of health because such coverage generates both *ex ante* and *ex post* moral hazard. To isolate the pairwise correlation between medical insurance coverage and health production efficiency, it is important to control for confounding factors by using multivariate analysis.

As a result, we control for factors, other than medical insurance coverage, that may influence the efficiency at which health is produced. Based upon health production theory and data availability, the following control variables are specified: population, per capita income, unemployment rate, the percentage of the population with diabetes or asthma, the state-level health maintenance organization (HMO) penetration rate, percentage of the population with disabilities, percentage 65 years and older, percentage with bachelor degrees or higher, percentage in poverty, percentage white, percentage African American, percentage Hispanic, percentage female, and percentage of employees working in the health or education fields, along with time and metropolitan fixed effects.

These variables should help control for the impact of age distribution, education, income and its distribution, race, disabilities, managed care and knowledge (through health/education employment) on the efficiency of health production. Because MSA-level data are unavailable to us, the HMO penetration rate at the state level must be used. By design, HMOs are supposed to practice cost-effective medicine and emphasize well care rather than sick care. Population serves to control for agglomeration economies such as the clustering of hospitals and physician clinics although it may pick up other factors associated with congestion such as greater crime and homicides.

The efficiency of health production and medical insurance coverage may be jointly influenced by the same unobservable factor or factors. For example, both variables may be influenced by the population's baseline health or degree of risk aversion. Specifically, individuals who are more risk averse may simultaneously be more inclined to purchase medical insurance and be efficient producers of health (propitious selection).[9] Also, sicker individuals face a greater incentive to choose medical insurance (i.e., adverse selection) but may be unable physically to efficiently produce health. Therefore, given the likelihood of endogeneity bias, an instrumental variables approach should be used.

Because employer-sponsored insurance remains the dominant form of medical insurance in the United States, variables capturing the size distribution of firms serve as instruments. These employment variables are employees per firms and the number of firms with 10 to 49, 50 to 99, 100 to 999, and 1,000 or more employees. Measures of firm size distribution have previously been used as instrumental variables in the prediction of health insurance indicators (Dranove, Simon, and White, 1998; Baker and Brown, 1999; Town et al., 2007; Bates and Santerre, 2008). State per capita income is specified as an additional instrument to control for the generosity of Medicaid eligibility in the different MSAs. All of these variables are likely to be correlated with medical insurance coverage but not the efficiency at which health is produced—a necessary property for a good instrumental variable.

Table 2 reports the multiple regression results for the first stage predicting the percentage of the metropolitan population with medical insurance coverage. Some observations are missing from the analysis because employment and/or demographic data are unavailable for some of the earlier years. The independent variables collectively explain about 89 percent of the variation in the medical insurance coverage rate.

The multiple regression results indicate that many of the instrumental variables are individually statistically significant. In addition, a Wald F-test shows that the instrumental variables are jointly significant. Higher state per capita income is associated with an increased percentage of individuals insured within an MSA, likely reflecting that state policies regarding Medicaid eligibility may be more generous.[10]

Interestingly, firm size appears to have a nonlinear impact on the percentage of people insured in the various MSAs. When compared to firms with one to nine employees (the default category), the percentage of the population insured first increases, then declines, and increases again with respect to the size of the firms. The nonlinearity probably captures the complex net impact of varying firm size on employment status (e.g., employment duration, casual employment, etc.) and the likelihood of employer-sponsored insurance in conjunction with eligibility for public insurance programs.

[9] See Hemenway (1990) for a discussion of propitious selection.

[10] Notice in Table 1 that 14 MSAs overlap two or more states. We use the per capita income and HMO penetration rate for the first city listed in the MSA name because it is presumably the most populous. For a sensitivity test, we also eliminated overlapping MSAs from the analysis and the results remained qualitatively similar.

TABLE 2
First Stage Results (Dependent Variable: Percentage of Population With Health Insurance)

	Coefficient	*t*-statistic
Constant	62.7601**	2.42
State per capita income	0.0005***	3.11
10 to 49 employees	0.0015***	3.32
50 to 99 employees	−0.0187***	−8.30
100 to 999 employees	0.0052**	2.24
1,000 or more employees	0.0963**	1.98
Employees per firm	−0.6139*	−1.75
Population	6.02E-06	1.39
Per capita income	−3.23E-05*	−1.70
Unemployment rate	0.7514***	3.02
Percent with asthma	−0.2592***	−7.40
Percent with diabetes	0.0241	0.16
State HMO penetration rate	0.0906***	18.11
Percent with disabilities	−0.1379	−0.93
Percent 65 years and older	−0.0933	−0.18
Percent BA+	−0.0427	−0.36
Percent poverty	0.0067	0.04
Percent white	−0.0730**	−2.20
Percent African American	−0.6139***	−2.90
Percent Hispanic	−0.1690*	−1.97
Percent female	0.5034	1.25
Percent health and ED employment	0.0714*	1.90
	Year and Metropolitan Fixed Effects	
F-test for instrumental variables	12.25 (prob. = 0.0000)	
Adjusted R^2	0.886	
Observations	192	

***Statistically significant at the 1 percent level.
**Statistically significant at the 5 percent level.
*Statistically significant at the 10 percent level.

For instance, the negative coefficient estimate on the number of firms with 50 to 99 employees may reflect both the relatively low likelihood of employer-sponsored insurance combined with the general ineligibility for Medicaid insurance because of the wage scale in firms of that size.

Some of the results for the control variables are worth noting. First, the medical insurance coverage rate appears to be directly related to the state HMO penetration rate. Perhaps, more individuals have the wherewithal to purchase medical insurance because of the lower premiums that result from the cost-effective practices of, and competition among, HMOs (Wickizer and Feldstein, 1995; Baker et al., 2000).[11] A seemingly odd finding is the positive and statistically significant coefficient estimate

[11] But we cannot rule out reverse causality even though the HMO penetration rate is measured at the state rather than MSA level.

on the unemployment rate. But the direct relationship may indicate that previous employed but uninsured individuals, and their dependents, become eligible for Medicaid insurance coverage once unemployed. Finally, in agreement with previous research on racial and ethnic disparities (e.g., Lillie-Blanton and Hoffman, 2005), the results show that African Americans and Hispanics are more likely than white individuals to be uninsured.

Given the debate in the literature concerning which functional form is best for the second-stage DEA model, four different specifications are estimated: ordinary least squares (OLS), two-stage least squares (2SLS), Tobit model, and truncated regression. For example, in a comparison of multiple regression results, Hoff (2007) shows that in some cases OLS may replace tobit as a sufficient second-stage DEA model. Moreover, based upon Monte Carlo experiments, Simar and Wilson (2007) report that truncated regression comes closer to estimating the correct DEA model than Tobit.

A Hausman (1978) test was conducted to determine the appropriateness of 2SLS estimates. Following the approach of Davidson and MacKinnon (1993), the residual from the estimated health insurance equation reported in Table 2 was specified in the second-stage DEA equation for the OLS, Tobit, and truncated models. A t-test was conducted to determine the statistical significance of the residual term in the second-stage DEA equation. The estimated t-statistics were 2.34, 1.24, and 0.52 for the estimated coefficients on the residual term, respectively, suggesting that the null hypothesis of exogeneity could be rejected only in the case of OLS.

The results associated with the second-stage DEA model are reported in Table 3 for the four different specifications. Not surprisingly, the findings indicate that disabled and elderly individuals are less efficient at producing health. The results also indicate that racial/ethnic and gender differences exist with respect to the efficient production of health and that neither income nor education greatly influences the marginal efficiency of health production. Somewhat puzzling, a higher poverty rate is associated with greater health production efficiency, although this result does not hold for the truncated regression model. Perhaps health literacy plays a role such that low-income individuals tend to systematically overestimate their true health status.

Of more importance to this article, the estimated coefficients on the medical insurance coverage rate are negative and statistically different from zero for all equations.[12] This result confirms theoretical expectations that medical coverage reduces the efficiency at which people produce their health. Moreover, the magnitude of the coefficient estimates on the medical insurance coverage rate supports some researchers' suspicions that the efficiency loss from medical insurance coverage may be relatively small because of the aforementioned nine limiting conditions. For example, according to the estimated coefficient on the health insurance variable in the 2SLS model, the largest in magnitude of the four, a 1 percentage point increase in the medical insurance coverage rate lowers the efficiency index by only 0.0045 percentage points, or about one-half of 1 percent on average.

Calculated elasticities also suggest that the relation between insurance coverage and the efficiency of health production is relatively inelastic. Evaluated at the means of

[12] One-tailed tests are conducted given that theory predicts an inverse relationship.

TABLE 3
Second-Stage DEA Results (Dependent Variable: Health Production Efficiency Score)

	Estimated Coefficient (Absolute value of *t*-statistic)		Estimated Coefficient (Estimated value of *z*-statistic)	
	OLS	2SLS	Tobit	Truncated Regression
Constant	1.855***	2.017***	2.150***	3.508***
	(6.79)	(8.29)	(4.30)	(6.02)
Health insurance coverage	**−0.0025****	**−0.0045*****	**−0.0031*****	**−0.0013***
	(2.18)	**(6.86)**	**(2.70)**	**(1.30)**
Population	−6.96E-09	−5.16E-09	−5.79E-08	−1.04E-09
	(0.26)	(0.17)	(1.35)	(0.02)
Per capita income	2.28E-08	4.86E-08	9.75E-07	7.07E-07*
	(0.06)	(0.11)	(1.56)	(1.66)
Unemployment rate	0.0055	0.0065	0.0070	0.0064
	(1.44)	(1.55)	(1.43)	(1.46)
Percent with asthma	0.0007	0.0005	0.0010	0.0022
	(0.63)	(0.45)	(0.62)	(1.40)
Percent with diabetes	0.0025**	0.0029**	0.0039	0.0012
	(2.52)	(2.08)	(1.59)	(0.54)
State HMO penetration rate	0.0003	0.0004	0.0009	−0.0007
	(0.78)	(1.13)	(1.58)	(1.38)
Percent with disabilities	−0.5748***	−0.5897***	−0.7549***	−0.5383***
	(2.72)	(2.90)	(4.24)	(3.08)
Percent 65 years and older	−1.6757**	−1.6137***	−2.006**	−1.5478**
	(2.62)	(2.78)	(2.57)	(2.37)
Percent BA +	−0.0002	−0.0004	−0.0030	−0.0022
	(0.11)	(0.28)	(1.45)	(1.04)
Percent poverty	0.0018**	0.0018**	0.0036**	0.0021
	(2.11)	(2.29)	(2.44)	(1.26)
Percent white	−0.1191	−0.1370	−0.1132	−0.3395***
	(1.20)	(1.42)	(1.02)	(3.40)
Percent African American	−0.4924**	−0.5232**	−0.3528	−0.7650**
	(2.40)	(2.55)	(1.12)	(2.48)
Percent Hispanic	−0.5889	−0.6394	−1.1357*	−1.7847***
	(1.47)	(1.49)	(1.68)	(2.82)
Percent female	−0.5584**	−0.5231**	−0.3289	−2.4892***
	(2.45)	(2.33)	(0.50)	(3.05)
Percent health and ED employment	0.2462	0.2558	0.4761**	0.3156
	(1.07)	(1.10)	(2.56)	(1.26)
	Year and Metropolitan Fixed Effects			
Adjusted R^2	0.515	0.502		
Log likelihood	538.98		316.44	401.69
Observations	187	187	187	128

***Statistically significant at the 1 percent level.
**Statistically significant the 5 percent level.
*Statistically significant at the 10 percent level.

the two variables, the estimated elasticities for the OLS, 2SLS, Tobit, and truncated regression models are approximately −0.22, −0.40, −0.27, and −0.11. The average of the four estimates, or −0.25, suggests that the efficiency of health production would decline by 2.5 percent given a 10 percent increase in the percentage of the population insured. Given that health insurance coverage also provides a reduction in risk exposure, this efficiency loss may be a small price to pay for greater financial security, especially for those low-income individuals and their families who are currently and completely uninsured. Interestingly, Finkelstein and McKnight (2005) find that the introduction of Medicare had little discernable impact on the health of the elderly but that the welfare gains from the reduction in risk exposure was worth one-half to three-fourths of the expense of the Medicare program at that time.

Conclusion

Nyman (2007) argues that health-care policy in the United States over the last 40 years has been preoccupied with reducing moral hazard. The preoccupation has steered U.S. health-care policy toward managed care and cost-sharing mechanisms and away from expanding medical insurance coverage to the approximately 45 million people currently uninsured. Nyman points out that much of this preoccupation has its roots in the RAND health insurance study, which finds a substantial moral hazard effect on the intensive margin.

This study uses an alternative approach to estimating the moral hazard effects of medical insurance by analyzing the impact of insurance coverage on health production efficiency on the extensive margin. Based upon an underlying health production function containing preventive, behavioral, and curative health inputs, DEA is used to measure the efficiency of health production across 60 MSAs over the period from 2002 to 2006. Although differences can be observed, the DEA shows, in general, that most MSAs are reasonably similar in terms of health production efficiency.

Next, multiple regression analysis is used to explain and examine the impact of medical insurance on the observed health production efficiency differences. The 2SLS procedure is used given that medical insurance coverage may be endogenous and a panel data set helps control for unobservable heterogeneity, at least to some extent. On the extensive margin, the empirical results from the multiple regression analysis support the view that medical insurance coverage is associated with a moral hazard effect but the magnitude of the effect seems to be relatively small.

Thus, the findings of this study imply that medical insurance might be expanded to those currently uninsured without much concern about the corresponding moral hazard effects. This is a particularly important policy implication given the widespread political discussions concerning universal health insurance currently taking place at the state and national levels. However, this policy implication should be cautiously accepted until further studies have taken place on this issue. Indeed, future research might improve upon this study in a number of ways.

First, future studies could benefit from including additional MSAs in the sample when more longitudinal data become available. Second, when consistent data are made available, researchers might specify additional preventive (e.g., likelihood of periodic prostate and breast exams) and curative (e.g., office visits) health inputs. Third, other measures of health, such as work days lost, might be used for particular population segments when that data become available. Fourth, not only the extent of medical insurance coverage but also the type of medical insurance (HMO, preferred provider organizations, etc.) might be included in the second-stage analysis.

Finally, Simar and Wilson (2007) recently suggest that second-stage DEA results may be biased because of serial correlation caused by efficiency of all observations being similarly determined by those observations on the frontier. As a result, these authors recommend bootstrap procedures in the second stage of the DEA analysis. While our second-stage DEA results are likely not greatly biased because all of the DEA scores are relatively close to 1 and vary little, and our output measure is also skewed to the right and varies little, we encourage others, especially those with the requisite specialized human capital, to pursue their suggested line of inquiry. At the very least, this study provides an initializing template to help organize and execute a more ambitious study on such an important and timely topic.

APPENDIX: DESCRIPTIVE STATISTICS AND DATA SOURCES FOR ALL VARIABLES USED IN THE STUDY

Variable	Obs.	Mean	Standard Deviation	Minimum	Maximum	Data Source
Health	300	85.77	3.42	72	94.5	1
% non-binge drinkers	300	84.93	3.45	73.9	93.5	1
% not-fat	300	40.49	3.77	23.3	49.3	1
% not-smoke	300	54.07	4.89	41.5	75.1	1
% exercise	300	77.4	4.27	65.1	87.1	1
% Flushot	300	69.1	6.74	48.1	88.2	1
Inpatient days per capita	263	0.70	0.26	0.26	2.15	2
Outpatient visits per capita	263	2.00	0.99	0.20	7.71	2
DEA efficiency score	263	0.98	0.03	0.88	1.00	3
% health insurance	300	86.24	4.27	70.5	94.3	1
State per capita income ($)	300	31,944	5,398	22,346	50,787	4
10–49 employees	278	8,089	8,439	503	50,963	5
50–99 employees	278	1,207	1,319	52	7,781	5
100–999 Employees	278	935	1,051	37	6,258	5
1,000+ employees	278	38	44	1	277	5
Employees per firm	278	16.58	2.04	11.39	21.45	5
Population	300	1,459,907	1,617,978	83,135	9,505,748	6
Per capita income	300	34,262	6,179	20,210	71,901	6
Unemployment rate	300	4.85	0.98	2.70	8.30	6
% asthma	300	8.78	2.15	3.60	15.90	1
% diabetes	300	6.84	1.61	3.20	12.70	1
State HMO penetration rate	300	18.70	11.28	0.10	50.50	4
% disabilities	200	14.61	2.60	9.69	25.73	7
% 65 years and older	200	11.46	2.00	7.18	17.57	7
% BA+	200	29.01	4.63	16.20	43.20	7

(*Continued*)

APPENDIX
Continued

Variable	Obs.	Mean	Standard Deviation	Minimum	Maximum	Data Source
% poverty	200	11.85	2.86	6.10	20.40	7
% white	200	77.10	11.53	49.85	96.18	7
% African American	200	13.48	11.90	0.303	46.75	7
% Hispanic	200	9.42	9.68	0.310	43.97	7
% female	200	51.11	0.829	48.61	52.64	7
% health and education employment	200	21.01	3.18	11.93	34.94	7

Note: The data source codes are defined as follows:
1. CDC (various years).
2. Health Forum LLC (various years).
3. Authors' estimates.
4. U.S. Census Bureau (various years) Statistical Abstract of the U.S.
5. U.S. Census Bureau (various years) MSA Business Patterns.
6. Bureau of Economic Analysis (various years).
7. U.S. Census Bureau (various years) American Community Survey.

REFERENCES

Auster, R., I. Leveson, and D. Sarachek, 1969, The Production of Health: An Exploratory Study, *Journal of Human Resources*, 4: 411-436.

Baker, L. C., and M. L. Brown, 1999, Managed Care, Consolidation Among Health Care Providers and Health Care: Evidence From Mammography, *RAND Journal of Economics*, 30(2): 351-374.

Baker, L. C., J. C. Cantor, S. H. Long, and M. S. Marquis, 2000, HMO Market Penetration and the Costs of Employer Sponsored Health Plans, *Health Affairs*, 19(September/October): 121-128.

Bates, L. J., K. Mukherjee, and R. E. Santerre, 2006, Market Structure and Technical Efficiency in the Hospital Services Industry: A DEA Approach, *Medical Care Research and Review*, 63(4): 499-524.

Bates, L. J., and R. E. Santerre, 2008, Do Health Insurers Possess Monopsony Power in the Hospital Services Industry? *International Journal of Health Care Finance and Economics*, 8(1): 1-11.

Breyer, F., and P. Zweifel, 1997, *Health Economics* (Oxford: Oxford University Press).

Brockett, P. L., R. E. Change, J. J. Rousseau, J. H. Semple, and C. Yang, 2004, A Comparison of HMO Efficiencies as a Function of Provider Autonomy, *Journal of Risk and Insurance*, 71(1): 1-19.

Bureau of Economic Analysis, various years, World Wide Web: http://www.bea.gov (accessed April 2, 2008).

Centers for Disease Control and Prevention, various years, *Behavioral Risk Factor Surveillance System Survey Data*. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Charnes, A., W. W. Cooper, and E. Rhodes, 1978, Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, 3(4): 392-444.

Chilingerian, J. A., and H. D. Sherman, 2004, Health Care Applications: From Hospitals to Physicians, From Productive Efficiency to Quality Frontiers, in: W. W. Cooper, L. M. Seiford, and J. Zhu, eds., *The Handbook on Data Envelopment Analysis* (New York: Kluwer Academic Publishers).

Crew, M., 1969, Coinsurance and the Welfare Economics of Medical Care, *American Economic Review*, 59(5): 906-908.

Cutler, D. M., and M. McClellan, 2001, Is Technological Change in Medicine Worth It? *Health Affairs*, 20(5): 11-29.

Dave, D., and R. Kaestner, forthcoming, Health Insurance and Ex Ante Moral Hazard: Evidence from Medicare, *International Journal of Health Care Finance and Economics*.

Davidson, R., and J. G. MacKinnon, 1993, *Estimation and Inference in Econometrics* (New York: Oxford University Press).

Debreu, G., 1951, The Coefficient of Resource Utilization, *Econometrica* , 19(3): 273-292.

Dranove, D., C. J. Simon, and W. D. White, 1998, Determinants of Managed Care Penetration, *Journal of Health Economics*, 17: 729-745.

Ehrlich, I., and G. Becker, 1972, Market Insurance, Self-Insurance, and Self-Protection, *Journal of Political Economy*, 80(4): 623-648.

Farrell, M. J., 1957, The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society*, 120 (Series A, Part 3): 253-281.

Finkelstein, A., and R. McKnight, 2005, What Did Medicare Do (And Was It Worth It?) NBER Working paper No. 11609.

Grossman, M., 1972, *The Demand for Health: A Theoretical and Empirical Investigation* (New York: Columbia University Press for the National Bureau of Economic Research).

Hausman, J. A., 1978, Specification Tests in Econometrics, *Econometrica*, 46: 1251-1271.

Health Forum LLC, various years, *Hospital Statistics* (Chicago: Health Forum LLC).

Hemenway, D., 1990, Propitious Selection, *Quarterly Journal of Economics*, 105(4): 1063-1069.

Hoff, A., 2007, Second Stage DEA: Comparison of Approaches for Modelling the DEA Score, *European Journal of Operational Research*, 181: 425-435.

Hollingsworth, B., P. J. Dawson, and N. Maniadkis, 1999, Efficiency Measurement of Health Care: A Review of Non-Parametric Methods and Applications, *Health Care Management Science*, 2: 161-172.

Idler, E., and Y. Benyamini, 1997, Self-Related Health and Mortality: A Review of Twenty-Seven Community Studies, *Journal of Health and Social Behavior*, 38: 21-37.

Kenkel, D. S., 2000, Prevention, in: A. J. Culyer, and J. P. Newhouse, eds., *Handbook of Health Economics* (Amsterdam: North-Holland), pp. 1675-1720.

Koopmans, T. C., 1951, An Analysis of Production as an Efficient Combination of Activities, in: T. C. Koopmans, ed., *Activity Analysis of Production and Allocation: Cowles Commission for Research in Economics, Monograph, No. 13* (New York: John Wiley and Sons Inc.).

Lillie-Blanton, M., and C. Hoffman, 2005, The Role of Health Insurance Coverage in Reducing Racial/Ethnic Disparities in Health Care, *Health Affairs,* 24(2): 398-408.

Newhouse, J. P., 1993, *Free for All? Lessons From the RAND Health Insurance Experiment* (Cambridge, MA: Harvard University Press).

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

Nyman, J. A., 2007, American Health Policy: Cracks in the Foundation, *Journal of Health Politics, Policy, and Law*, 32(October): 759-783.

Pauly, M. V., and P. J. Held, 1990, Benign Moral Hazard and the Cost Effectiveness of Insurance Coverage, *Journal of Health Economics*, 9(4): 447-461.

Ray, S. C., 2004, *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research* (Cambridge: Cambridge University Press).

Rosen, H. S., and S. Wu, 2004, Portfolio Choice and Health Status, *Journal of Financial Economics*, 72: 457-484.

Schlesinger, H., and E. Venezian, 1986, Insurance Markets With Loss-Prevention Activity: Profits, Market Structure, and Consumer Welfare, *Journal of Risk and Insurance*, 17(2): 227-238.

Schmidt, N. B., M. J. Telch, and T. E. Joiner, 1996, Factors Influencing Health Perceptions in Patients With Panic Disorder, *Comprehensive Psychiatry*, 37: 253-260.

Simar, L., and P. W. Wilson, 2007, Estimation and Inference in Two-Stage Semi-parametric Models of Production Processes, *Journal of Econometrics*, 136: 31-64.

Town, R. J., D. Wholey, R. Feldman, and L. R. Burns, 2007, Revisiting the Relationship Between Managed Care and Hospital Consolidation, *Health Services Research*, 42(February): 219-238.

Tulkens, H. P., and P. Vanden Eeckaut, 1995, Non-parametric Efficiency, Progress, and Regress Measures for Panel Data: Methodological Aspects, *European Journal of Operational Research*, 80(3): 474-479.

U.S. Census Bureau, various years. American Community Survey. World Wide Web: http://www.census.gov/acs/www/Products/index.htm (accessed April 3, 2008).,

U.S. Census Bureau, various years, MSA Business Patterns. World Wide Web: http://censtats.census.gov/cgi-bin/msanaic/msasect.pl (accessed April 2, 2008).

U.S. Census Bureau, various years, Statistical Abstract of the U.S. World Wide Web: http://www.census.gov/compendia/statab/ (accessed April 2, 2008).

Wickizer, T. M., and P. J. Feldstein, 1995, The Impact of HMO Competition on Private Health Insurance Premiums, *Inquiry*, 32(Fall): 241-251.

Worthington, A. C., 2004, Frontier Efficiency Measurement in Health Care: A Review of Empirical Techniques and Selected Applications, *Medical Care Research and Review*, 61: 135-170.

Zweifel, P., and W. G. Manning, 2000, Moral Hazard and Consumer Incentives in Health Care, in: A. J. Culyer and J. P. Newhouse, eds., *Handbook of Health Economics* (Amsterdam: North-Holland), pp. 409-459.