From Sickness to Health: The Twentieth-Century Development of U.S. Health Insurance

Melissa A. Thomasson

Department of Economics, Miami University and NBER
E-mail: thomasma@muohio.edu

This paper explores the institutional development of health insurance in the United States. By combining a qualitative history of the development of the market with an empirical analysis of a panel of health insurance data from 1931–1955, the paper identifies a number of factors that influenced the growth of the health insurance market. While demand factors such as increasing income and improvements in medical technology certainly contributed to the growth of the market, supply side factors were also important. There is evidence that hospitals may have contributed to the growth of health insurance as a means of smoothing revenues during the Great Depression. State-level policies that allowed the Blue Cross and Blue Shield plans to operate as nonprofits also spurred market growth, as did federal government policies that promoted the link between employment and health insurance. © 2002 Elsevier Science (USA)

I. INTRODUCTION

By 1920, 16 European countries had adopted some form of nationalized, compulsory health insurance that provided income replacement and medical care in the event of accident or illness.² In the United States, the market developed as a primarily private, employment-based system despite attempts to implement compulsory health insurance plans in the 1910s, 1930s, and 1940s. Modern health insurance did not develop in the United States until the late 1920s, and it was only after 1940 that the market experienced substantial growth. As shown in Fig. 1, only 12.3 million Americans had health insurance coverage in 1940, but the market exploded in size in the 1940s and 1950s. By 1960, 122.5 million

¹ Many thanks to Bill Collins, George Davis, Price Fishback, Gerald Friedman, Bart Wilson, and seminar participants at North Carolina State University for providing excellent comments and suggestions. This paper benefited tremendously from the comments of two anonymous referees. Financial support from NSF Grant SBR-9632121 and a grant from the University of Arizona are gratefully acknowledged. All errors are my responsibility.

² Germany established the first nationalized system of health insurance in 1883, followed by Austria (1888), Hungary (1891), Norway (1909), Serbia (1910), Britain (1911), Russia (1912), and the Netherlands (1913). See Millis, p. 50, or Starr, p. 237, for greater discussion.
people—nearly 10 times the number covered in 1940—were enrolled in private health insurance plans (Source Book of Health Insurance Data, 1976–1977, p. 22).

What factors contributed to the initial development of the health insurance market in the late 1920s and the tremendous growth of the market that occurred after 1940? Why did a private system of employment-based health insurance develop in the United States when compulsory insurance was popular in Europe? By combining a qualitative history of the market with a panel of state-level data from 1931–1955, this paper identifies several factors that were important in market development and growth. On the demand side, results show that increases in the demand for health insurance resulted from rising income and improvements in medical technology. Government policies that promoted a link between health insurance and employment lowered the real price of health insurance and further stimulated demand in the 1940s and 1950s.

Catalysts on the supply side were also important. Hospitals faced with fluctuating demand developed an institutional innovation in insurance, the prepayment plans that later became Blue Cross. Several years later, physicians followed suit with Blue Shield. Although physicians were initially opposed to health insurance, they viewed private insurance as a compromise between their ideal of no insurance and the more unpleasant situation of government-sponsored, compulsory insurance. Blue Cross and Blue Shield plans were instrumental in increasing the supply of insurance. Not only did they represent the first modern health insurance plans, but also they proved that the adverse selection and moral hazard problems that were thought to be prohibitively associated with health insurance could be overcome. State-level policies that allowed the Blue Cross and Blue Shield plans to operate as nonprofits (thus freeing them from the usual...
insurance regulations) spurred their growth. By 1940, the market grew further as commercial insurance companies entered the market after viewing the success of the Blue Cross and Blue Shield plans and recognizing that potential problems with adverse selection could be overcome by selling to employee groups.

II. THE GROWTH OF U.S. HEALTH INSURANCE

Prior to 1920, both the demand for and the supply of health insurance were virtually nonexistent. The low demand for health insurance resulted from the fact that the chief financial losses resulting from sickness were associated with lost wages. Actual medical expenditures were small because of the relatively low state of medical technology, and the public had little confidence in the efficacy of medical care. Patients were typically treated at home, and hospitals were charity institutions that the danger of cross-infection gave well-earned reputations as places of death. Movements for compulsory insurance failed repeatedly in the 1910s. The low costs of medical care largely made the need for health insurance “... difficult to justify in the context of the period...” It was felt that the family should be thrifty and save for the rainy day of illness” (Anderson, 1968, p. 86). Thus, instead of health insurance, “sickness” insurance policies designed to protect the insured against loss of income developed, while actual health insurance remained relatively unknown. It was not until the late 1920s that consumers regarded riders covering medical expenditures as more than just frills on disability insurance policies (Goldmann, p. 68).

The low level of demand for health insurance was matched by a similarly low level of supply. Health insurance policies were not widely offered by insurance companies, since insurance companies were reluctant to insure against what they perceived as an uninsurable event. An insurable loss was thought to be one which was (1) definite and measurable, (2) unexpected and uncontrolled, and (3) widely disbursed and not subject to catastrophic loss (Eilers, p. 13). When attempting to insure against health losses and medical expenditures, would-be health insurers

---

3 A 1919 State of Illinois Commission report estimated the total cost of sickness for individual wage earners and found that for each dollar of wages lost by individual wage earners, an additional 24 cents had to be added to cover medical expenses (State of Illinois, pp. 15–17). Overall, the total cost per wage earning family of lost wages and medical expenses totaled approximately $75. By way of comparison, the Commission found that a funeral in a wage-earner’s family cost $132.40 (p. 17).

4 A 1918 Bureau of Labor Statistics survey of 211 families living in Columbus, Ohio found their average annual medical bill to be $48.41, of which hospital expenses accounted for only 7.6 percent. Expenditures on physicians and surgeons’ services accounted for 51.6% of all spending, with medicine (14.1%), dentistry (12.4%), nursing (10.1%), and eyeglasses (4.2%) composing the remainder (see Ohio Report, p. 116).

5 These proposals failed for several reasons. They were typically opposed by insurance companies, physicians, and druggists. For greater discussion see Numbers (1978) or Chasse (1994).

6 Emery (1996) states that before the 1930s, fraternal societies were the principal source of sickness and health insurance for white males in the U.S. and Canada, providing sickness benefits and access to a physician in the event of illness. The role of fraternal lodges in insurance declined significantly after 1929. For greater discussion, see also Emery and Emery (1999).
faced a severe adverse selection and monitoring problem. In addition, the loss attributable to sickness and medical treatment was far from definite, since actuarial data on morbidity were not well collected. Commenting on the reluctance of insurance companies to offer health insurance, one author noted that, “. . . to collect on life insurance, the insured must be dead, to collect on accident insurance, he must have had an accident, while to collect on sickness insurance he must have a policy” (McCahan, p. 187).

However, by the late 1920s, both the demand for and the supply of health insurance increased as several changes occurred that tended to enlarge the role that medicine played in people’s lives and to shift the focus of treatment of acute illness from homes to hospitals. Urban homes tended to be smaller than rural ones, with less room to care for sick family members (Faulkner, p. 509). Advances in bacteriology and medical technology also contributed to the rise of hospitals as treatment centers. With improvements in anti-infection techniques, surgical possibilities increased dramatically. While surgery was often performed in private homes until the 1920s, hospitals became natural centers not only for surgery, but also for X-rays and laboratories. The growing atmosphere of medicine as a science further helped to encourage sick people to visit physicians and hospitals. Rosenberg notes that “by the 1920s . . . prospective patients were influenced not only by the hope of healing, but by the image of a new kind of medicine—precise, scientific and effective” (p. 150). This scientific aura began to develop in part as licensure and standards of care among practitioners increased. Following the publication of Abraham Flexner’s critical report on the status of medical education in 1910, more rigorous standards for physician education and licensure were implemented.7

In addition to limiting the supply of physicians, Flexner’s report and the subsequent reforms in medical education may have increased the demand for physicians’ services by increasing the quality of physicians’ skills, thus increasing the cost of medical services.8 Further, the variation in expenditures increased as technology progressed and a greater number of illnesses became treatable. Important breakthroughs such as the antibiotic action of sulfonamide drugs in

---

7 The AMA recruited Abraham Flexner to evaluate the status of medical education. Flexner heavily criticized the system, stating that the methods of medical education prior to 1910 had “. . . resulted in enormous over-production at a low level, and that, whatever the justification in the past, the present situation . . . can be more effectively met by a reduced output of well trained men than by further inflation with an inferior product” (Flexner, p. 16). Flexner argued for stricter entrance requirements, better facilities, higher fees, and tougher standards. Following the publication of the Flexner Report, the number of medical schools in the United States dropped from 131 in 1910 to 95 in 1915. By 1922, the number of medical schools in the United States had fallen even further to 81 (Journal of the American Medical Association, August 12, 1922, p. 633).

8 According to one study, the average American family had medical expenses totaling $108 in 1929, with hospital expenditures composing 14 percent of the total bill (Falk, Rorem, and Ring, p. 89). In 1929, medical charges for urban families with incomes between $2,000 and $3,000 per year averaged $67 if there were no hospitalizations, but averaged $261 if there were any illnesses that required hospitalization.
1935, penicillin in 1946, and the vaccine for polio in 1955 convinced American families of the effectiveness of medical care. Families with access to physicians and hospital services began to demand more care, and their demand for health insurance increased as they sought new ways to pay for rising expenditures.

At the same time the demand for medical care and health insurance was increasing in the 1920s, the supply of insurance also increased as hospitals and insurance companies developed innovative ways of providing insurance for medical expenditures. The first organizations to offer modern health insurance were not commercial insurance companies, but rather hospitals, as health insurance originally developed as a means to ensure that patients paid their hospital bills. In 1929, the first “Blue Cross” plan was founded when Baylor University Hospital contracted with a group of Dallas teachers and agreed to provide 21 days of hospitalization for a fixed $6.00 annual payment. One official connected with the Baylor plan compared hospital bills to cosmetics, noting that the nation’s cosmetic bill was actually more than the nation’s hospital bill, but that “We spend a dollar or so at a time for cosmetics and do not notice the high cost. The ribbon counter clerk can pay 50¢, 75¢, or $1 a month, yet . . . it would take about twenty years to set aside a large hospital bill” (The American Foundation, p. 1023). During the Great Depression, these plans gained momentum as hospital revenues fell and hospitals used prepayment plans to increase income.9 As one pediatrician in the Midwest noted, “Things went swimmingly as long as endowed funds allowed the hospitals to carry on. When the funds from endowments disappeared the hospitals got into trouble and thus the various plans to help the hospitals financially developed” (American Foundation, p. 756).

These prepayment plans were the first actual health insurance plans to be offered on a widespread scale. By providing benefits in the form of services instead of a cash indemnity, the prepayment plans limited the problem of moral hazard. Marketing the plans to groups of employees allowed the hospital plans to overcome problems associated with adverse selection. In addition, having employers deduct premiums from employee paychecks lowered the administrative costs associated with selling insurance. The American Hospital Association (AHA) encouraged such endeavors ostensibly as a means of relieving “. . . from financial embarrassment and even from disaster in the emergency of sickness those who are in receipt of limited incomes” (Reed, 1947, p. 14). However, the AHA also recognized that the plans could serve hospitals in other ways, and began to organize them under the name Blue Cross. As the AHA developed the guidelines that Blue Cross plans were required to follow, they did so in such a way as to maximize hospitals’ benefits. Since single-hospital plans generated greater competition among hospitals, the AHA designed the Blue Cross guidelines so they would reduce interhospital price competition. Prepayment plans

---

9 Hospital receipts per patient dropped from $236.12 in 1929 to $59.26 one year later, and occupancy declined from 71.28 to 64.12% (Bulletin of the American Hospital Association, July 1930, p. 68).
seeking the Blue Cross designation had to provide subscribers with free choice of physician and hospital, a requirement that eliminated single-hospital plans from consideration.\(^\text{10}\)

Blue Cross plans also benefited from special state-level enabling legislation allowing them to act as nonprofit corporations, to enjoy tax-exempt status, and to be free from the usual insurance regulations. Originally, the reason for this exemption was that Blue Cross plans were considered to be in society’s best interest since they often provided benefits to low-income individuals (Eilers, p. 82). Without the enabling legislation, Blue Cross plans would have had to organize under the laws for insurance companies. If they organized as stock companies, the plans would have had to meet reserve requirements to ensure their solvency. Organizing as mutual companies meant that they would either have to meet reserve requirements or be subject to assessment liability.\(^\text{11}\) Since most plans had few financial resources available to them, they would not have been able to meet the requirements. As noted in the *Yale Law Journal*, “...full compliance with these stringent requirements would seriously hamper the formation and operation of group health plans” (*Yale Law Journal* (1943), p. 171).

The enabling legislation freed the plans from the traditional insurance reserve requirements because the Blue Cross plans were underwritten by hospitals. Hospitals contracted with the plans to provide subscriber services, and agreed to provide service benefits even during periods when the plans lacked funds to provide reimbursement. Under the enabling legislation, the plans “enjoy the advantages of exemption from the regular insurance laws of the state, are freed from the obligation of maintaining the high reserves required of commercial insurance companies and are relieved of paying taxes” (Anderson, 1944, p. 11). As a result, these enabling laws facilitated the availability and growth of prepaid health insurance.\(^\text{12}\)

Despite the success of Blue Cross and prepaid hospitalization policies, physicians were much slower to provide prepaid care. Physicians were traditionally staunch opponents of health insurance because they feared that interference from

---

\(^{10}\) For further discussion and a listing of the AHA requirements, see Eilers, p. 11.

\(^{11}\) Stock companies are companies that are owned by stockholders who are entitled to the earnings of the company. Stock companies are required to hold reserves to guard against insolvency (see Faulkner, pp. 406–429, for a detailed discussion of reserves). Mutual companies are cooperative organizations in which the control of the company and its ownership rest with the insureds. Mutual companies may be required to have reserves or to engage in assessment liability (in which insureds must pay additional amounts if premiums fall short of claims). Both stock and mutual companies pay taxes.

\(^{12}\) Regular insurance companies viewed the enabling laws with “considerable misgiving” (*Michigan State Bar Journal* (1940), p. 165), although most insurance companies did not offer health insurance at the time. The enabling legislation was also crafted to ensure that medical services were not covered in the Blue Cross plans, and that the plans allowed subscribers free choice of physician. New York was the first state to enact such enabling legislation in 1934, and 32 states had adopted special enabling legislation for hospital service plans by 1943. Other states exempted Blue Cross plans by categorizing them strictly as nonprofit organizations (Eilers, pp. 100–107).
a third party would restrict their income and limit their ability to price discrimi-
nate.\textsuperscript{13} Thus, Blue Cross and Blue Shield developed separately, with little coordination between them (McDavitt, 1946). Some physicians were even so opposed to the idea of voluntary health insurance that they ensured that the Blue Cross enabling laws prohibited the plans from covering physicians’ services (\textit{Michigan State Bar Journal} (1940), p. 167).\textsuperscript{14}

However, by the late 1930s physicians were faced with two situations that spurred them to develop their own prepaid plans. First, Blue Cross plans were becoming popular, and some physicians feared that hospitals would move into the realm of providing insurance for physician services, thus limiting physician autonomy. In addition, advocates of compulsory health insurance looked to the emerging Social Security legislation as a logical means of providing national health care. Compulsory health insurance was even more anathema to physicians than voluntary health insurance. It became clear to physicians that in order to protect their interests, they would be better off preempting both hospitals and compulsory insurance proponents by sculpting their own plan (Hedinger, p. 82).

As a result, physicians began to organize a framework for prepaid plans that covered physician services. The American Medical Association (AMA) adopted a set of 10 principles in 1934 “. . . which were apparently promulgated for the primary purposes of preventing hospital service plans from underwriting physician services and providing an answer to the proponents of compulsory medical insurance” (Hedinger, p. 82). Within these rules were provisions that ensured that voluntary health insurance would remain under physician supervision and not be subject to the control of nonphysicians. In addition, physicians retained their ability to price discriminate by charging patients who were subscribers to Blue Shield the difference between their actual charges and the amount for which they were reimbursed by Blue Shield.

These principles were reflected in the actions of physicians as they established enabling legislation similar to that which allowed Blue Cross plans to operate as nonprofits. Like the Blue Cross enabling legislation, these laws allowed Blue Shield plans to be tax-exempt and free from the provisions of insurance statutes. Physicians lobbied to ensure that they would be represented on the boards of all such plans and acted to guarantee that all plans required free choice of physician.

\textsuperscript{13} In the 1910s, physician opposition helped defeat several proposed state bills that would have provided compulsory health insurance. One joke making the rounds in New York summarized physician sentiment, saying that to pick out a doctor in any crowd, all a person had to do was to “whisper ‘health insurance’ in a man’s ear and see whether his hand goes to his pocket” (Numbers, p. 89).

\textsuperscript{14} Not all physicians were anti-health-insurance. The split among members of the medical profession was especially pronounced in 1934, when the American College of Surgeons stated that they favored voluntary, prepayment health insurance for hospitalization, prompting some leaders of the American Medical Association to publicly condemn their statement. (See “Surgeons Back Health Insurance; Vote to Lead National Movement,” \textit{New York Times}, 6/11/1934, p. 1; “Doctors Resent Health Insurance,” \textit{New York Times}, 6/12/1934, p. 22 B).
To further these efforts, the AMA encouraged state and local medical societies to form their own prepayment plans. These physician-sponsored plans ultimately affiliated and became known as Blue Shield in 1946. Like the Blue Cross enabling laws, states that had enacted Blue Shield enabling legislation further facilitated the growth of health insurance.

Once it was apparent that Blue Cross and Blue Shield had successfully overcome problems associated with adverse selection and moral hazard, commercial, for-profit insurance companies began to move rapidly into the health insurance market. As the Superintendent of Insurance in New York, Louis H. Pink, noted in 1939,

\[ \ldots \text{There are twenty stock insurance companies which are today issuing in this state Individual Medical Reimbursement, Hospitalization, and Sickness Expense Policies. About half of these have only recently gone into this field. It is no doubt the interest aroused by the non-profit associations which has induced the regular insurance companies to extend their activities in this way (Pink, 1939).} \]

The success of commercial insurance companies was aided by the fact that they experience-rated their policies while Blue Cross and Blue Shield initially relied on a system of community rating. Under a system of community rating, insurance companies charge the same premium to sicker people as they do to healthy people. Since they were not considered to be nonprofit organizations, commercial insurance companies were not required to community-rate their policies. Instead, commercial insurance companies could engage in experience rating, whereby they charged sicker people higher premiums and healthier people lower premiums. Thus, commercial firms could often offer lower rates than the Blues to healthy employee groups. The market grew rapidly after 1940, as shown in Fig. 1.

Government policies in the 1940s and 1950s further encouraged the growth of health insurance and reinforced the efforts of insurance companies to link health insurance with employment. During World War II, wage and price controls prevented employers from using wages to compete for scarce labor. Under the 1942 Stabilization Act, Congress limited the wage increases that could be offered by firms, but permitted the adoption of employee insurance plans. In this way, health benefit packages offered one means of securing workers.\(^\text{15}\)

The tax treatment of employer-provided contributions to employee health insurance plans also spurred the employment-based system. First, employers did not have to pay payroll tax on their contributions to employee health plans.

\(^\text{15}\) Several other rulings in the 1940s further bolstered the tie between health insurance and employment. In 1945 the War Labor Board ruled that employers could not modify or cancel group insurance plans during the contract period (Scofea, p. 6). In 1949, the National Labor Relations Board ruled that the term “wages” included pension and insurance benefits. Therefore, when negotiating for wages, the union was allowed to negotiate benefit packages on behalf of workers as well (Inland Steel Co. v. NLRB (170 F. 2d 247 (7th Cir. 1948))). Eilers (p. 19) offers further discussion.
Further, under certain circumstances, employees did not have to pay income tax on their employer’s contributions to their health insurance plans. The first such exclusion occurred under an administrative ruling handed down in 1943 that stated that payments made by the employer directly to commercial insurance companies for group medical and hospitalization premiums of employees were not taxable as employee income (Yale Law Journal (1954), pp. 222–247). This ruling was highly restrictive and limited in its applicability since it only affected direct employer contributions to group plans issued by commercial insurance companies. However, the Internal Revenue Code of 1954 expanded the scope of this ruling and codified it, providing an actual statutory basis for the tax treatment of employer contributions to employee health plans. Between 1953 and 1958, the percentage of Americans covered by health insurance plans increased from 63 to 75%, in part due to the tax subsidy (Thomasson, 2000).

Clearly, several factors served to increase both the supply and demand for health insurance after the first modern Blue Cross plan began in Dallas in 1929. Among other factors, the demand for health insurance may have increased as a result of advancing medical technology, the improved quality of hospital and physician services, increasing income, and government policies that served to increase the demand for employment-based insurance. The supply of health insurance may have increased as hospitals developed the Blue Cross plans to reduce interhospital competition and as physicians followed suit with Blue Shield. State-level enabling legislation that freed the Blue Cross and Blue Shield plans from traditional insurance regulatory requirements and granted them tax-exempt status may also have increased the supply of health insurance. Further, the success of the Blues and their ability to surmount the problems of adverse selection by marketing to employee groups led commercial insurance companies into the market in the 1940s. In the next section, this paper analyzes the demand and supply of insurance in order to measure the effect of each of these factors on the development of health insurance in the United States from 1930, a year after its inception, to 1955, a year after the implementation of the tax subsidy.

III. MODEL AND ESTIMATION

To measure how these factors influenced the development of the health insurance market over the period 1931–1955, state-level data are used to estimate a system of supply and demand equations. The demand for health insurance may be influenced by several of the factors discussed in the previous section, such as the availability and efficacy of medical services, the price of insurance and the availability of group insurance, and socioeconomic factors such as income, race, and urbanization. Factors that may determine the supply of health insurance include the price of health insurance, the presence of Blue Cross and Blue Shield enabling laws that made it easier for the Blues to operate, the availability of hospitals and physicians to act as providers of insured services, and socioeconomic factors such as race and urbanization. Since there may be differences
across states that do not vary over time, as well as changes over time that are uniform for all states, the model is estimated using two-way fixed effects (see Greene, 1993). The system is modeled as

\[
Q_{it}^d = \alpha_i + \gamma_t + P_{it}\beta_1^d + X_{it}\beta_2^d + \varepsilon_{it}^d \quad (1)
\]

\[
Q_{it}^s = \alpha_i + \gamma_t + P_{it}\beta_1^s + X_{it}\beta_2^s + \varepsilon_{it}^s \quad (2)
\]

\[
Q_{it}^d = Q_{it}^s = Q^*, \quad (3)
\]

where (1) represents the demand for health insurance in state \(i\) in time \(t\), (2) represents the supply, and (3) is an equilibrium condition requiring that the quantity demanded equal the quantity supplied. To control for differences in states that affect the supply of health insurance but do not vary over time, the \(\alpha_i\) are included as state dummy variables. In addition, the \(\gamma_t\) are included as year dummy variables to control for factors that do not vary across states but that may vary over time, such as advances in medical technology and changes in government policy.

The dependent variable is the natural logarithm of the real value of accident and health insurance premiums written per capita in state \(i\) in year \(t\). Information on accident and health insurance premiums is from the Spectator Insurance Year Books. Unfortunately, actual “health” insurance data are difficult to obtain. The Spectator data contain information about the total premiums written and total losses incurred on accident and health insurance data by state for each year. However, as reported in the year books, insurance against medical expenditures and indemnity sickness insurance are uniformly reported as “health” insurance. Further, some companies did not distinguish between accident and health insurance. As a result, information on accident insurance often contains some premiums sold as health insurance, and vice versa. In some states, insurance companies always reported accident and health business together; in these cases, it is impossible to distinguish between health insurance and accident insurance. Given the difficulty in distinguishing between them, both categories are summed and the total is used as the measure of real total accident and health insurance premiums written per capita within a state. The Consumer Price Index (1967 = 100) for all items is used to deflate the series.

Included in both the demand and supply equations is the natural logarithm of the price of insurance \((P_{it})\), or more technically, the load. The load is the percentage increase in an actuarially fair premium that is included to pay for administrative and other overhead expenses. It represents the expense to the consumer of choosing to purchase an insurance policy instead of self-insuring—the total premium of a policy is the actuarially fair premium multiplied by the

---

16 This is the measure used by Farley and Wilensky (1985). While measures such as total enrollment in health plans or the value of coverage may be preferred, this information is not available at the state level for most of years used in the analysis.
load. A “no-load” policy has a loading value of one. A policy with a 40% load to cover costs would have a loading value of 1.40. The loading fee can be calculated using the Spectator data as the ratio of real insurance premiums written in a state to expected real total losses, where expected real total losses are the one-period lagged value of real total losses. Given that the quantity of health insurance and the load are endogenous variables, the system is estimated using three-stage least squares. The error terms in each equation are assumed to be normally distributed and uncorrelated both across states and over time. Identification comes from the exclusion restrictions as discussed below.

Variables included in the supply equation that are not included in the demand equation are dummy variables controlling for whether or not a state had implemented the Blue Cross and/or Blue Shield enabling laws that allowed these plans to bypass traditional insurance regulatory requirements. Variables included in the demand equation that are not included in the supply equation include the natural logarithm of real state per capita income, a measure of educational attainment, the percentage of nonagricultural workers in a state who are unionized, and the natural logarithm of manufacturing establishments per capita in a state. If insurance is a normal good, then the amount of insurance purchased should increases as income increases, ceteris paribus. To control for this effect, the natural logarithm of real state income per capita from the Bureau of Economic Analysis is included in the demand equation. Given that unions may have negotiated on behalf of employees for fringe benefits such as health insurance, more health insurance may have been sold in states with high levels of union membership. Information on union membership at the state level is available from Fishback and Kantor (1996) for 1929 and from Troy in 1939 and 1953.

Education is also expected to positively influence the demand for health insurance (Dewar, 1998). Calculating educational attainment is complicated because there is not a standard measure of adult education across censuses. The U.S. Censuses in 1920 and 1930 provide numbers on adult literacy rates, while the 1940, 1950, and 1960 censuses report years of school completed. However, a rough approximation of how years of school completed translates into illiteracy is provided by a Special Report of the U.S. Census published in 1948 that lists the proportion of the population that is illiterate, based on the number of years of school they completed (also see Smith, 1984). This information is used to calculate an estimate of the percentage of adults in a state over age 25 that are illiterate.

17 Bradford and Logue (1996) also use the ratio of real premiums to expected losses. The decision to specify expected losses as the one-period lag of real total losses comes from the fact that Dickey–Fuller tests on the dollar value of real total losses showed that they can best be forecast as a random walk.

18 The census does not disaggregate the number of school years a person has completed; it reports years of schooling as zero years, 1–4 years, 4–6 years, etc. Thus, I obtain the percentage of persons who have completed 1–4 years of schooling who are illiterate by taking the ratio of the average of people who are illiterate with one year of schooling and with four years of schooling, to the average
Other variables that are included in both the supply and demand equations are the natural logarithm of the number of hospitals and physicians per capita, the percentage of a state’s population living in an urban area, the percentage of firms employing over 20 workers, and the percentage of a state’s population that is black. People in states with greater numbers of hospitals and physicians will be more likely to demand health insurance. On the supply side, states with higher numbers of hospitals and physicians may have had a greater supply of health insurance given that hospitals and physicians organized health insurance plans as a means to ensure that patients paid their medical bills. Data on the number of hospitals in a state are published in *Hospital Service in the United States and Hospitals: The Journal of the American Hospital Association*. Data on the number of physicians in each state come from the U.S. Census.

Urbanization may also affect both demand and supply by lowering the costs of obtaining medical care and health insurance (on the demand side) and the cost of offering health insurance (on the supply side). As one insurance agent noted, “Southwestern Virginia is a largely rural community . . . We should have to place our contracts individually . . . such individual soliciting would entail a large overhead expense” (American Foundation, p. 1024). To further control for the fact that group contracts (particularly for large employers) lowered the cost of health insurance, the percentage of firms employing over 20 workers is also included in both equations.

Race may be an important determinant of demand if blacks had a more difficult time obtaining care in hospitals (Smith, 1999), thus dampening their demand for health insurance, or if insurance companies were reluctant to insure blacks. There is some evidence that insurance companies may have used race as a factor when underwriting policies. As Faulkner notes, “the underwriting problems involved in attempting to insure members of more than one race would be considerable owing to differences in physique, temperament, habit and environment” (p. 118). There is also evidence that even when policies were available to blacks, the premiums were actuarially more expensive, since the premium may have been “. . . graded in proportion to the percentage of non-Caucasian persons in the group” (Faulkner, p. 308). To control for these differences, U.S. Census

---

19 A study conducted by the *Journal of the American Medical Association* found that over 20 percent of general hospitals excluded blacks in 1940 (*JAMA* (1941), p. 1066). In addition, few physicians were black, and it may have been the case that some blacks felt uncomfortable seeking treatment from nonblack physicians.
data on the percentage of a state’s population that was black is included in both equations.

Thus, supply is identified through the exclusion of illiteracy, unionization, income, and the number of manufacturing establishments. Demand is identified through the exclusion of the dummy variables that control for the presence of state-level Blue Cross and Blue Shield identifying laws. Since identifying a supply and demand system using exclusion restrictions can be complicated when there is more than one instrument available, tests of the overidentifying assumptions were performed to verify the validity of the exclusion restrictions (Woolridge, 2002). The tests failed to reject the exclusion restrictions, suggesting that the specified model is appropriate.\(^{20}\)

Summary statistics for the data are reported in Table 1. While the health and accident insurance premium data, income data, and hospital data and the dummies for Blue Cross and Blue Shield enabling legislation are reported yearly, the other variables in the equation are reported less frequently. The data on establishment size are reported in odd years from 1929 to 1939, and then in 1947 and 1954. With the exception of the data on unionization (reported in 1929, 1939, and 1953), the other variables (urbanization, illiteracy, physicians, and population) are reported only in census years. Thus, the census-year-only variables and the

\(^{20}\) As an additional check, a reduced form equation was estimated instead of the system described in Eqs. (1)–(3). Results were not substantively different from those that are described in Table 2. Since it is not clear that the gain from estimating a reduced form outweigh the cost of losing information on the role of the load in demand and supply, only the results from the system estimation are reported here.
establishment size variable are linearly interpolated, while the state-level union-
ization variable is interpolated using the annual national unionization rate. The
equations are then estimated using three-stage least squares over the entire
1931–1955 period, and the results are presented in Table 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Supply equation</th>
<th>Demand equation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln(phys/capita)</td>
<td>0.723</td>
<td>0.494</td>
<td>0.073</td>
</tr>
<tr>
<td>Ln(hosp/capita)</td>
<td>0.357</td>
<td>-0.246</td>
<td>0.145</td>
</tr>
<tr>
<td>Load</td>
<td>1.324</td>
<td>-1.609</td>
<td>0.007</td>
</tr>
<tr>
<td>Percent firms ≥20 employees</td>
<td>-0.745</td>
<td>1.978</td>
<td>0.029</td>
</tr>
<tr>
<td>Percent black</td>
<td>1.813</td>
<td>-7.142</td>
<td>0.003</td>
</tr>
<tr>
<td>Percent urban</td>
<td>-1.079</td>
<td>-0.741</td>
<td>0.181</td>
</tr>
<tr>
<td>Ln(manufacturing est./capita)</td>
<td></td>
<td>0.580</td>
<td>0.001</td>
</tr>
<tr>
<td>Ln(real income/capita)</td>
<td></td>
<td>0.751</td>
<td>0.000</td>
</tr>
<tr>
<td>Percent illiterate</td>
<td></td>
<td>0.418</td>
<td>0.625</td>
</tr>
<tr>
<td>Percent unionized</td>
<td></td>
<td>0.001</td>
<td>0.754</td>
</tr>
<tr>
<td>Blue Cross enabling law</td>
<td>0.159</td>
<td>0.014</td>
<td>0.018</td>
</tr>
<tr>
<td>Blue Shield enabling law</td>
<td>0.106</td>
<td>0.014</td>
<td>0.007</td>
</tr>
<tr>
<td>1932</td>
<td>0.109</td>
<td>0.041</td>
<td>0.626</td>
</tr>
<tr>
<td>1933</td>
<td>0.266</td>
<td>-0.315</td>
<td>0.036</td>
</tr>
<tr>
<td>1934</td>
<td>0.098</td>
<td>-0.217</td>
<td>0.037</td>
</tr>
<tr>
<td>1935</td>
<td>-0.073</td>
<td>-0.049</td>
<td>0.539</td>
</tr>
<tr>
<td>1936</td>
<td>-0.049</td>
<td>-0.074</td>
<td>0.373</td>
</tr>
<tr>
<td>1937</td>
<td>-0.092</td>
<td>0.245</td>
<td>0.018</td>
</tr>
<tr>
<td>1938</td>
<td>0.032</td>
<td>0.210</td>
<td>0.017</td>
</tr>
<tr>
<td>1939</td>
<td>-0.014</td>
<td>0.336</td>
<td>0.001</td>
</tr>
<tr>
<td>1940</td>
<td>0.071</td>
<td>0.492</td>
<td>0.000</td>
</tr>
<tr>
<td>1941</td>
<td>0.293</td>
<td>0.164</td>
<td>0.170</td>
</tr>
<tr>
<td>1942</td>
<td>0.287</td>
<td>0.172</td>
<td>0.191</td>
</tr>
<tr>
<td>1943</td>
<td>0.327</td>
<td>0.088</td>
<td>0.581</td>
</tr>
<tr>
<td>1944</td>
<td>0.320</td>
<td>0.434</td>
<td>0.002</td>
</tr>
<tr>
<td>1945</td>
<td>0.600</td>
<td>0.240</td>
<td>0.206</td>
</tr>
<tr>
<td>1946</td>
<td>0.690</td>
<td>0.312</td>
<td>0.106</td>
</tr>
<tr>
<td>1947</td>
<td>0.717</td>
<td>0.449</td>
<td>0.012</td>
</tr>
<tr>
<td>1948</td>
<td>0.773</td>
<td>0.705</td>
<td>0.000</td>
</tr>
<tr>
<td>1949</td>
<td>0.938</td>
<td>0.855</td>
<td>0.000</td>
</tr>
<tr>
<td>1950</td>
<td>1.105</td>
<td>1.433</td>
<td>0.000</td>
</tr>
<tr>
<td>1951</td>
<td>1.698</td>
<td>1.008</td>
<td>0.000</td>
</tr>
<tr>
<td>1952</td>
<td>1.920</td>
<td>1.147</td>
<td>0.000</td>
</tr>
<tr>
<td>1953</td>
<td>2.124</td>
<td>1.282</td>
<td>0.000</td>
</tr>
<tr>
<td>1954</td>
<td>2.332</td>
<td>1.260</td>
<td>0.000</td>
</tr>
<tr>
<td>1955</td>
<td>2.447</td>
<td>0.964</td>
<td>0.016</td>
</tr>
</tbody>
</table>

Note. Dependent variable is Ln(real premiums/capita). Equations estimated using 3SLS. Demand equation $R^2 = 0.77$. Supply equation $R^2 = 0.88$. For both equations, the omitted state dummy is CT and the omitted year dummy is 1931.
The results strongly support the underlying theory that the price of insurance is a determinant of both supply and demand and demonstrate the importance of other key factors in the development of the health insurance market. Overall, the supply and demand system is well behaved. Results from the 3SLS estimation of the supply equation show that special enabling legislation allowing Blue Cross and Blue Shield plans to form as nonprofit organizations had a positive, statistically significant effect on the amount of health and accident insurance sold within a state. Insurance companies in states with a Blue Cross enabling law sold roughly 17% more insurance per capita than in states without such legislation. In states with a Blue Shield enabling law, insurance companies sold nearly 11% more insurance per capita than in states that did not enact special enabling legislation. Both the number of hospitals and the number of physicians (per capita) in a state had a positive, statistically significant effect on the quantity of accident and health insurance supplied. Increasing the number of hospitals per capita by 1% increases the amount of insurance supplied by 0.357% while increasing the number of physicians per capita by 1% increases the amount of insurance supplied by 0.645%.

As expected, the estimated coefficient on the loading fee is a positive, statistically significant determinant of the supply of health and accident insurance in a state. A 1% increase in the load causes a 1.32% increase in the quantity of health and accident insurance supplied. Urbanization has an unexpected, statistically significant negative effect on the supply of health insurance. This unexpected result may indicate that the urbanization variable is capturing more than just the advantages of selling insurance in an urban area, such as health disadvantages associated with living in an urban environment, or that the urban measure is capturing the inability of price level adjustments to nominal income to differentiate between more and less urban states (Collins and Thomasson, 2002).

Results from the estimation of the demand equation also support economic theory. The estimated coefficient on the loading fee is negative as expected and statistically significant, suggesting that a 1% increase in the price of insurance generates a 1.6% decrease in the quantity of accident and health insurance demanded. Income has a large, positive effect on the demand for accident and health insurance. A 1% increase in real per capita state income generates a 0.75% increase in the quantity of health insurance demanded. The availability of physicians is a positive, statistically significant determinant of the demand for health insurance, while the number of hospitals does not affect the demand for accident and health insurance in a statistically significant way. This result may reflect the fact that patients are usually admitted to hospitals on the advice of physicians. Both the number of manufacturing establishments per capita and the percentage of firms with over 20 employees are directly related with the amount of accident and health insurance in a state, suggesting that health insurance was more prevalent in more industrialized states, and in states with large firms. This result supports the hypothesis that the availability of group insurance was key
determinant of insurance demand. Results also indicate that states with high
technology, ceteris paribus. A 1% increase in a state’s black population generates a 7% decrease in
the quantity of accident and health insurance demanded.

While these results provide an indication of what factors were important in market development, they offer little information about the relative contribution of each variable to the overall growth of the market from 1931 to 1955. The mean value of real accident and health insurance premiums per capita increased from $2.76 in 1930 to $21.70 in 1955. The results in Table 2 can be used to calculate how much of the total increase in the mean value of the dependent variable over the period can be explained by each control variable in the demand and supply equations. These computations are presented in Table 3. In the demand equation, rising income accounts for 35% of the increase in the quantity demanded of health insurance from 1932 to 1955. Increases in the number of manufacturing establishments and of large firms together account for nearly 15% of the increase in the quantity of health insurance demanded, while falling prices for insurance account for an additional 7%. An increase in the number of hospitals per capita comprises 2.6% of the increase in the quantity of health insurance demanded. On

21 Table 3 shows the change from 1932 to 1955, since 1931 is the omitted dummy variable in the regression.
the supply side, the presence of Blue Cross and Blue Shield enabling laws together account for 9.4% of the increase in the quantity of health insurance supplied, while a decrease in loading fees over time tends to decrease the quantity supplied by 6%.

However, the single largest factor in explaining the rise in health insurance from 1932 to 1955 in either equation is time. The year effects in Table 3 account for 43% of the increase in demand and over 108% of the change in supply. These effects are more fully reflected in the estimated coefficients on the year indicator variables in Table 2, which capture changes that were constant across states but that varied over time. Figure 2 depicts these estimated coefficients graphically. As previously mentioned, such changes might include the effect of federal government policy changes that promoted the growth of group insurance, and changes in medical and pharmaceutical technologies. These effects undoubtedly tell a significant part of the story in explaining the rise in health insurance from 1931 to 1955. By examining the magnitude of the estimated coefficients of these variables, it may be possible to gain a sense of the impact that these changes had on the health insurance market. During the Great Depression, the estimated coefficients on the year dummies in the supply equation increase dramatically from 1932 to 1933, perhaps reflecting the fact that more hospitals were beginning to offer Blue Cross plans to smooth revenue. In contrast, the estimated coeffi-

22 That is, the increase in quantity supplied due to time more than offsets the decrease due to factors such as falling prices.
cients in the demand equation fall during the Great Depression, and do not turn upward until 1937. Developments in medical technology (such as sulfa drugs in 1935, or the increasing role of hospitals as places of birth) may be reflected in this upward trend.\textsuperscript{23} As medical treatment became more effective and as people increasingly received medical care from physicians and hospitals, the demand for health insurance may have also increased as families sought ways to pay their medical bills. There is a large increase in the magnitude of the estimated coefficient for 1946 in the demand equation, perhaps reflecting the new availability of penicillin after World War II.

Government policies that fostered the growth of employment-based health insurance may have also led to an increase in both the supply and the demand for health insurance in the 1940s and 1950s. These policies include a 1942 War Labor Board ruling that enabled firms to offer health insurance instead of wage increases during a period of wage controls, and tax policies (in 1943 and 1954) that allowed workers to exempt employer contributions to their health insurance plans from taxable income. Insurance companies benefited from employment-based insurance because it reduced adverse selection and administrative costs. Workers benefited because of the lower price of group health insurance and because of the tax benefits. The estimated coefficient on the demand year indicator variable rises sharply after 1943, possibly in response to the first change in tax policy. The estimated coefficients continue to grow in magnitude in the late 1940s, possibly because health insurance became more closely linked with employment. For example, a National Labor Relations Board ruling in 1948 enabled unions to negotiate for benefits on behalf of workers, affirming the relationship between employment and health insurance. Overall, Fig. 2 indicates that changes that are hard to measure, such as changes in government policy and medical technology, play a large role in the development of the insurance market.

IV. CONCLUSIONS

This paper discusses and analyzes the forces that contributed to the initial development and growth of the U.S. market for health insurance. Results suggest that although the demand for health insurance rose during the first half of the century, supply innovations had a major impact on the initial development of the market. During the 1930s, hospitals developed an institutional innovation in the form of prepayment plans as they attempted to reduce interhospital competition and to smooth revenue uncertainty during the Great Depression. They were later joined by the doctors’ organizations that established the Blue Shield prepayment plans to provide coverage for nonhospital treatment. Once the Blues signaled the presence of a market for health insurance, commercial insurers joined the

\textsuperscript{23} Birth shifted from homes to hospitals in the late 1920s and 1930s. In 1935, 36\% of all births occurred in hospitals. By 1945, this number was 79\%. There are several factors accounting for the shift, but increasing use of anesthesia during labor and the availability of sulfa drugs may help to explain part of the trend (see Leavitt, 1986).
competition and found ways to lower prices through experience rating and selling group insurance.

The supply-side innovations were aided greatly by specific government policies at the state level. The passage of enabling legislation that freed the Blues from insurance regulations led to significant increases in the sale of total health insurance even though the Blues became a smaller part of the market. This legislation had its cost, however, because it required the Blues to follow a community rating policy in setting their insurance premium rates. The requirement of community rating by the Blues opened the door for commercial insurers to compete by lowering their rates for lower-risk insurance purchasers.

While supply-side forces played a significant role in shaping the early health insurance market, changes in demand also contributed to the expansion of the health market. Health insurance clearly was a normal good, as rising incomes helped to fuel the expansion of insurance. Government policy also influenced the demand side of the market. The wage controls imposed during World War II and the early favorable tax treatment of employer contributions to health insurance premiums (which was firmly established in 1954) contributed to the expansion of employment-based, group insurance (see Thomasson, 2000).

Finally, there were also significant changes in factors that were not easily measured. In the empirical estimation the coefficients of the year dummies show a clear upward growth in the demand and supply for health insurance, even after controlling for income growth and other demand-side factors. The most plausible effect that the year dummies capture is that of the growth in medical technology and of the increasing awareness of medical technology among Americans. While this effect is not measurable, it no doubt contributed to an increase in the demand for health insurance during the late 1930s and 1940s. Considering both the measurable effects reflected in the supply and demand equations, as well as the effects captured in the individual year dummies, the growth of the health insurance market between 1930 and 1950 resulted from no single force, but rather the simultaneous occurrence of a number of factors. Starting from the prepayment schemes fostered by the hospitals in the early 1930s, health insurance grew into its own as improvements in medical technology stimulated the demand for health insurance, and insurance companies began offering insurance to employee groups. Government policy in the 1940s and 1950s reinforced this trend, and cemented the employer-based system of private health insurance that the United States has today.

REFERENCES

The American Foundation (1937), American Medicine, Volume II. New York: The American Foundation.


Anderson, Odin W. (1944), State Enabling Legislation for Non-Profit Hospital and Medical Plans,


Council on Medical Education of the American Medical Association (various years), Hospital Service in the United States. Chicago: American Medical Association.


Flexner, Abraham (1910), Medical Education in the United States and Canada. Carnegie Foundation for the Advancement of Teaching.


Health Insurance Institute (various years), Source Book of Health Insurance Data. New York: Health Insurance Institute.

Hedinger, Fredric R. (1966), The Social Role of Blue Cross as a Device for Financing the Costs of Hospital Care: An Evaluation. Iowa City: The University of Iowa.

Inter-University Consortium for Political and Social Research (no date), Historical, Demographic, Economic, And Social Data: The United States, 1790–1970 [Computer file]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor].

Journal of the American Medical Association, various issues.


Michigan State Bar Journal (1940), 19.


Ohio Health and Old Age Insurance Commission (1919), Health, Health Insurance, Old Age Pensions. Columbus, F. J. Heer Printing Co.

Pink, Louis H. (1939), “Voluntary Hospital and Medical Associations and the State,” address to the Meeting of the Medical Society of the County of Queens, Forest Hills, NY, February 28, 1939. [Published in Journal of American Insurance, March 1939 16(3)].


Spectator Company (various years), Insurance Year Book, Casualty, Surety, and Miscellaneous. New York: Spectator Company.


