

Chapter 2: LITERATURE REVIEW

You have to study a great deal to know a little.

Montesquieu²³

This chapter focuses upon the economic studies of HMOs that provide the background for the research performed here. Specifically, it discusses the nature of HMOs' cost advantage over other types of health insurance, the role of market power, the evolution of HMOs and previous studies examining differential growth across geographic areas.

Some of the literature used as background for this project was discussed in Chapter 1, Introduction. In particular, the diffusion of innovations academic tradition, which is mostly concerned with innovations outside of health care, was used to generate a framework for analyzing HMO growth and outlining the statistical design of this study. By contrast, this chapter focuses almost exclusively on health economic publications. However, there are some logical connections between these 2 diverse literatures, and those connections are discussed toward the end of Section A and in Section B of this chapter.

A. Nature of Cost Advantage

Much of the policy interest in HMOs rises from the belief that they offer a way to reduce health care costs or at least slow the rate of cost growth. Therefore the literature on the extent and nature of cost growth forms a critical backdrop to this project. This section discusses the cost advantage of HMOs. It begins with some basic observations and reviews concepts, methodology and major findings in that order.²⁴

A1. BASIC OBSERVATION

Basic economics of HMOs begins with the observation that HMOs charge lower premiums and out-of-pocket costs than seemingly comparable fee for service (FFS) plans. This was demonstrated by Luft (1978), which included a review of prior literature. Luft found that costs per capita for HMOs were usually 10-40% lower than for FFS plans. Most of this difference stemmed from a lower hospitalization rate by HMOs. Combining the insurance and provision functions in one organization means that HMOs avoid incentives to treat that providers paid by indemnity insurance have.

Miller and Luft (1994) updated and expanded the previous article with a review of the published record of 54 studies. Concerning utilization, the authors found some evidence that HMOs had lower hospital admission rates and strong evidence that hospital length of stay is lower for HMO enrollees. Evidence concerning physician office visits was mixed, but substantial data indicated that HMOs use fewer particularly expensive procedures, tests, and treatments.

²³ Thoughts and Unpublished Writings of Montesquieu, published in French, 1899.

²⁴ Different studies use different terms and assumptions, but their methods and findings are presented here using a common terminology. Issues of cost savings are purely quantitative, and they can only be resolved through mathematical exposition. Therefore, Thesis Appendix B, The Algebra of HMO Cost Estimation, addresses some of the major issues in a rigorous manner.

Miller and Luft could not reach a conclusion with regard to health care expenditures and health insurance premiums. More recent studies have found that HMOs generate cost savings in the 5-15% range (Baker et al. 2000; Polsky and Nicholson 2001). However, the overall level of premiums is much higher than it was 30 years ago, and most non-HMO plans now employ some managed care methods. Therefore the dollar savings from enrolling in most HMOs instead of a generous FFS plan may well be as much as or more than it was during earlier phases of the HMO movement.

A2. CONCEPTS

The initial hypothesis advanced by HMO advocates in the 1960s was that HMOs achieved their savings by reducing or eliminating unnecessary or counter-predictive health care. This concept is sometimes referred to as the utilization hypothesis. But selection is a major factor in the operation of health insurance markets (Akerlof 1970; Rothschild and Stiglitz 1976; Shain 1966). A contrasting hypothesis holds that HMOs have a natural advantage in invoking selection and that the superior health of HMO enrollees accounts for the lower costs.

HMOs can ensure favorable selection in a variety of ways. But a key concept is that HMOs start out offering some combination of restricted utilization and lower quality care for lower premiums. This approach attracts healthy individuals (or employers with healthy employees), who find the financial savings sufficient to justify the slight risk of being seriously harmed by inferior HMO care.²⁵ As these low-cost persons enroll in HMOs, they leave high-cost enrollees in FFS plans, and the difference in premiums grows. Possibly, the metropolitan areas in which HMOs succeeded in exploiting favorable selection are the ones in which they have grown most.

Besides possibly enjoying a selection effect, HMOs may also create a competitive effect. Quite simply, HMO competition may spur FFS plans to cut their costs. Unlike any effects from reduced utilization or favorable selection, the hypothesized competition effect would not change the value of HMO premiums. Therefore, studies that just focus on HMO premiums cannot test for the competition effect. However, any competition effect should be observed in both the premiums of other health insurers and in overall health care costs within an MSA or a state.

The selection effect may explain observed rates of HMO insurance premiums. The competition effect may explain observed rates of FFS insurance premiums. Logically, these two hypotheses are independent. In particular, it might be that HMOs have been able to undercut FFS insurance premiums by invoking favorable selection, *and* that FFS plans have responded by reducing their costs by greater efficiency or some other mechanism. It is also possible that HMOs and FFS plans have entered into intense competition for the least risky health insurance policyholders.²⁶

²⁵ This argument would most apply to healthy persons who are risk prone, risk neutral, or only moderately risk averse.

²⁶ And plausibly, because of this competition, both categories of insurers have become more selective in their enrollment and thus reduced the percentage of the population with private insurance coverage.

If we concentrate upon explaining HMO premiums *per se*, the utilization and selection hypotheses are normally associated with different value judgments concerning HMOs. Reducing unnecessary utilization is tautologically good, while exploiting favorable selection is almost always viewed as a social evil.²⁷ A third hypothesis is that HMOs do not reduce utilization, but they do reduce reimbursement to providers. This hypothesis implies that HMOs are able to use market power to prevail upon providers to dispense equivalent treatments for reduced reimbursement. If the reimbursement hypothesis is true, the associated value judgment is less clear. It depends upon what group of individuals ultimately secures the resources “taken away” from providers and a subjective evaluation of the merits of providing that group with those resources.²⁸

Even though utilization, selection, and now reimbursement are conceived of as competing hypotheses to explain observed HMO premium rates, all three may be true. After a brief discussion of methodology, section A4 presents findings concerning these three hypotheses. A number of individual studies have sought to estimate the extent of the selection effect, and, until recently, the absence of a selection effect was taken imply that cost differences were generated by the utilization effect. Only one recent article, still unpublished, estimates the magnitude of the reimbursement effect (Polsky and Nicholson 2001).

A3. METHODOLOGY

Because extensive data are available on HMO enrollment and on health insurance premiums or health care costs, in principle, it is possible to test whether HMO reduced costs result from favorable selection or other causes. The empirical analysis can be conducted on three levels: individual, provider, or market.²⁹ Each has advantages and drawbacks. In the next section, we consider outstanding examples of studies on all three levels.

Table 2-1 summarizes three levels of analysis and their qualities. Analyses at the provider level and the market level are similar in that they both use HMO market share in a geographic area as the independent variable and a cost related measure as the response variable.³⁰ As noted in the table, these two levels of analysis both incur the risk of endogeneity; endogeneity occurs when the response variable contributes to variability in the independent variables. However, the two types of studies differ fundamentally in their control variables.

²⁷ Possibly a necessary evil if other benefits, associated with a free market, are to be enjoyed.

²⁸ For example, some (but not all) of the funds transferred from providers could go to HMO executives' salaries and stock options. Or the funds might be appropriated to HMO enrollees in the form of lower premium payments and higher wages. Another issue concerns how providers adjust their behavior in response to the reduced reimbursement. According to a recent press report, an unpublished study finds that the uninsured in high managed care areas “have greater difficulty accessing care than the uninsured in areas with lower managed care penetration.” American Health Line, 5/21/01, #8.

²⁹ A 4th possible level would be by employer. Few studies have been conducted on this level probably because of difficulties in securing employer level cost records from insurance companies or because of distrust of the concept of matching comparable employers. One such study was Feldman et al, 1993.

³⁰ Usually HMO costs aren't distinguished from FFS costs because it is very difficult to do so on either the provider or market levels. One study, Wholey et al 1995, was able to estimate HMO costs on the market level. With the right combination of cost data and a credible model of how the selection and competition effects work, it is possible to determine if HMOs reduce global costs and to estimate how much.

Table 2-1: Alternative Levels of Analysis

Comparing Costs for HMO and FFS Enrollees

Methods that control for Selection effects

Method	Advantages	Drawbacks
<p><u>1. Individual</u> 2 variants: a) Randomly Assign b) Compare Switchers with non-Switchers</p>	<p>1) Possibility of analyzing costs in detail - i.e. Hospital vs Pharmaceutical, etc. Primarily applying to Variant b³¹: 2) Possibility of knowing sample in detail, and, if necessary, controlling for numerous individual attributes. 3) Possibility of surveying subjects for reasons switched, health status, satisfaction, etc.</p>	<p>1) Samples are usually small and non-representative of Universe (US Population). 2) Difficulty of following the same individuals over a period of years. 3) Will not naturally measure the competition effect, though it may be possible to do so.</p>
<p><u>2. Provider</u> Regress a provider-specific measure or correlate of costs on HMO Market Share or growth thereof</p>	<p>1) Controls for characteristics of individual providers. 2) Will show competition effect. 3) May be possible to keep better track of units than with individual consumers. 4) Potential for greater n, but see drawback 1). 5) Many components and correlates of costs could be analyzed this way to get the most complete picture.</p>	<p>1) Phony inflation of n. Since providers in an MSA tend to be highly correlated, residuals are probably not independent. 2) Any provider-specific measure will have only a weak relationship with total health costs. 3) Endogeneity 4) Ignorance of proper control variables. 5) Inconsistency of HMOs across geo units.</p>
<p><u>3. Regional</u> (Usually MSA or County) Regress total hospital costs³² by geographic area on HMO Mkt Share/Growth</p>	<p>1) Study population can include large numbers of persons and is perceived to be representative of US universe. 2) Will capture competition effect. 3) Units can be consistently defined over a period of years, but see drawback 6.</p>	<p>1) Small number of observations until recently. 2) Problems with comprehensiveness and accuracy of cost data. 3) Endogeneity 4) Ignorance of proper control variables. 5) Inconsistency of HMOs across geo units. 6) Consistently defined geo units do not necessarily imply populations that are congruent over time.</p>

³¹ Special advantages of random assignment are well known. In brief, the singular advantage is the likely elimination of *all* differences in confounding variables, whether those variables are recognized and measured or not.

³² Total health care costs for MSAs, Counties or smaller units are generally not available.

With provider level data, it is possible to control for provider specific characteristics that may influence health costs. Market studies can include provider characteristics aggregated to the market level, but controlling for each provider's attributes facilitates more accurate estimation of the true effects of the control variables, subject to the required assumptions.³³ However, the use of a large number of providers for purposes of testing the statistical significance of HMO generated cost effects may be misleading, since the residuals for providers within geographical units may be highly correlated.

Random assignment of individuals to HMO and FFS plans is the easiest way to compare costs with the assurance that selection does not influence the results. However, this has only been done once and is unlikely to be repeated any time soon.³⁴ Thus studies on the individual level mostly use retrospective data. When HMOs were growing rapidly, it was easier to compare persons who switched from FFS to HMO with those who have not.³⁵

Individual studies promise greater precision in the analysis of control variables.³⁶ Finally, individual level data permit comparison between a population exclusively enrolled in HMOs and one exclusively enrolled in FFS plans.

However, individual studies also suffer from drawbacks relative to market-wide studies. Individual studies usually involve individuals enrolled in a small number of plans in one or two markets. The internal validity, though strong, only applies to a small population, and the external validity is weak. Additionally, individual studies suffer from attrition, which most likely is not random, and any competition effect that HMOs have in reducing FFS costs may be difficult to isolate in an individual study.

A4. MAJOR FINDINGS

Individual Studies

The RAND Health Insurance Experiment (HIE) found that favorable selection did not play a role in reduced costs.³⁷ Enrollees randomly assigned to Group Health Cooperative (GHC) of Seattle incurred expenses about 30% lower than enrollees randomly assigned to generous Fee For Service plans. The GHC experimental group incurred about the same average expenses as a control group of persons with ongoing GHC enrollment, suggesting there was no favorable selection of membership by the cooperative.³⁸

³³ Particularly that the provider attributes are truly structural, and not merely correlated with unobserved structural variables.

³⁴ Randomized assignment to health plans usually requires that the experiment pay the costs of health insurance for the subjects in return for their surrender of their freedom of choice; this makes it an extremely expensive research method.

³⁵ In 1988, Luft and Miller analyzed a number of studies of HMO disenrollees, and there have been more studies since. But the large enrollment in PPOs complicates the analysis.

³⁶ In social science theory, there is only 1 set of "true" control variables. Therefore individual patient and individual provider studies can't both be right. The easy way out conceptually, which is the hard way out in practice, is to conclude that both sets of control variables are needed.

³⁷ Manning et al, 1987

³⁸ However, GHC may not have been typical of all HMOs at the time HIE was conducted. The GHC of that time certainly is not typical, in either its structure or its operation, of HMOs today.

In contrast to the HIE findings, a study of Blue Cross/Blue Shield enrollees who switched to HMOs in the Twin Cities from 1978 to 1981 found considerable favorable selection (Jackson-Beeck and Kleinman 1983). Those who chose to enroll in HMOs when their employers starting offering this option had recorded less than half the hospital days of the majority who remained with Blue Cross/Blue Shield; similarly, defectors to HMOs had incurred only about 60% of the total health care expenses of the those who did not switch. However, the circumstances of the Twin Cities study may limit the validity of the findings. At that time, it was understandable that sickly persons were reluctant to leave a prestigious Blue Cross plan to enroll in a form of insurance that was still novel in the Midwest.

Luft and Miller (1988) attempted to summarize the results of about 50 individual-level studies published in the 1970s and 1980s. A small portion of the results found that HMOs suffer from adverse selection. However, the bulk of the results divided about evenly between findings of statistically significant favorable selection and no findings in either direction. Based on this summary, the authors concluded, “biased selection exists.” They were unable to estimate, or even bracket, the magnitude of favorable selection effects.

One very recent study finds no evidence for either selection or utilization effects. Polsky and Nicholson (2001) report that HMO enrollees on the whole are about as healthy and have about the same degree of utilization as non-HMO enrollees. They find that lower HMO premiums are due to the reimbursement effect, to the lower payments that HMOs make to providers. These findings are consistent with previous findings that favorable selection is likely to decline as HMOs grow and enroll more diverse populations (Miller and Luft 1994).³⁹

Effects On Overall Expenditures and Its Components:

In principle, when the selection effect dominates, HMOs have little or no effect on overall health care expenditures. However, if lower HMO premiums are the result of reduced utilization, than overall expenditures are lower. Independently, a competition effect that lowers costs for fee-for service insurers would also lower total costs. If lower HMO premiums are the result of a reimbursement effect, then total costs may be lowered, depending upon whether providers are able to make up their reduced HMO reimbursement with higher reimbursement from other insurers.

Numerous studies have attempted to estimate or otherwise understand the effect of HMO growth on health care expenses from retrospective data on an aggregate level. A recent study found that the higher HMO penetration was at the MSA level, the lower both HMO and non-HMO premiums tended to be (Baker et al. 2000). This finding suggests that HMOs do not achieve lower premiums solely by invoking favorable selection. It also rebuts the hypothesis that provider price concessions to HMOs result in higher payments to other insurers. However, a comparable study found that increased HMO market share only leads to lower premiums under certain conditions (Wholey, Feldman and Christianson 1995).⁴⁰

³⁹ In the early 1990s favorable selection may have been most significant in the Medicare population, where overall market share was lower than in the commercially insured population and the average level of health care expenditures was close to four times that of the rest of the population. However, subsequent events suggest that favorable selection declined in that market.

⁴⁰ Please see Section B, The Role of Market Power, for further discussion of this article.

Since reliable estimates of total health care costs on the county or MSA level are generally not available, most of these studies have examined a particular aspect or component of costs. For example, expenses not paid through insurance premiums must be covered by out-of-pocket payments. Another recent study examined the estimates of US out-of-pocket health care spending for the years 1990, 1993, 1995 and 1997 (Gabel et al. 2001). Contrary to widespread perceptions, this study found that out-of-pocket payments went down for people with employer-based insurance; the growth of HMO enrollment was a major cause of the decline.

Some of the most frequently studied effects of HMO growth are effects on hospital costs and other hospital attributes. Several works find HMO enrollment had little effect on hospital costs per capita at the MSA level (McLaughlin et al 1984; McLaughlin 1987; McLaughlin 1988a; McLaughlin 1988b). A more recent study by Gaskin and Hadley (1997) found considerable evidence that high levels of HMO enrollment in MSAs reduced hospital cost inflation between 1985 and 1993.⁴¹ Other researchers have found in several studies that increased Preferred Provider Organization (PPO) and HMO enrollment can lead to lower costs under certain conditions (Melnick and Zwanziger 1988; Melnick et al. 1992; Zwanziger, Melnick and Bamezai 2000). Section B discusses these studies. Chernew (1995) studied 1982 and 1987 data and found that growth in HMO enrollment led to a decline in number of hospital beds per MSA.

One study from the 1990s found that high HMO enrollment had ambiguous effects on physician fees and incomes (Baker 1994). The effect of HMO market share on other components of costs deserves further study. The most important components include pharmaceutical prices and costs, the compensation of nurses and other allied health providers and administrative costs. Several recent studies provide additional perspectives on overall costs, but much remains to be resolved (Sheils and Haught 1997; Cutler and Zeckhauser 1997; Sullivan 2000).

Towards Resolution of Conflicting Studies:

The contradictory evidence with respect to the selection, utilization, competition and reimbursement effects may be beyond complete resolution. However, the diffusion of innovation literature, and in particular the categorization scheme for innovation adopters, both presented in Chapter 1, can serve as an analogy to help resolve the conflicting findings concerning favorable selection by HMOs. It has been established that farmers, for example, that adopt innovations earlier than average have certain characteristics that distinguish them from the bulk of their peers (Rogers 1995, 261-266).

Likewise, it is reasonable to expect that persons who enrolled in HMOs early may have had some distinctive characteristics of their own. If these characteristics are primarily the result of favorable selection, this is secondary. The main point is that these findings may only be applicable for a limited time. Just as all the eventual adopters of hybrid corn in Iowa were different than those who started hybridizing early, it is reasonable to hypothesize that, once HMO enrollments reached 20% to 30% of the population, those distinctive characteristics of

⁴¹ Unfortunately, this study was sponsored by the largest national HMO organization, the American Association of Health Plans (AAHP). However, almost all of the data are publicly available, and it should be possible to replicate it.

HMO enrollees would become less pronounced. The later studies, such as Polsky and Nicholson (2001) that find no evidence of favorable selection are consistent with this hypothesis. Likewise, these findings are consistent with the numerous studies, cited above, that find that higher HMO enrollment tends to reduce overall health care costs or some component thereof.

However, the selection process can work both ways, i.e. enrollees have some ability to select insurers even as insurers seek to select enrollees. This process of joint selection may help resolve some of the contradictions in the literature.

Kemper et al. (1999/2000) also found that HMO enrollees had similar health status to subscribers in other insurance plans. However, HMO enrollees differed in income and demographics, being poorer, younger, more likely to be single, and more likely to be a member of a minority group. Most significantly, HMO enrollees differ in *attitudes*. The authors write, "Many more HMO enrollees report that they are willing to trade off provider choice than are people with non-HMO insurance." Much of the rest of the population desires wide choice of providers and convenient access to a broad array of health services, and they are willing to pay for those advantages. Possibly related to attitudes concerning the choice/premium tradeoff, survey data show that about a quarter of the population thinks that HMOs provide bad service to health care consumers (Blendon et al 1998).

This implies that the selection effect does not result exclusively from HMOs enrolling healthier people. Rather, it is at least partly the result of *people who wish to economize on health care costs* enrolling in HMOs for that purpose.⁴² This distinction is important because it shows how HMOs may increase total welfare. If HMO premiums were and are lower simply because they enroll healthier people, then HMOs' effect on overall costs and total welfare would be small.⁴³

But if selection operates on the basis of consumer preferences as well as health status, then HMOs frequently enroll people who want to save money on health care costs and help them to satisfy their legitimate preferences.⁴⁴ If there were no HMOs, these people would only have the options of more expensive insurance or no health insurance at all. If we assume, as a first approximation, that all these people would enroll in more expensive plans, then the effect of HMOs is to reduce health care costs. By contrast, we can assume that many would enroll in

⁴² The specific finding is that 65% of HMO enrollees, and only 50% of non-HMO enrollees, are willing to trade off provider choice for cost savings. This increased willingness is the most distinct attribute of HMO enrollees. For example, 65% of HMO enrollees and 69% of non-HMO enrollees are married.

Since employers mediate this mechanism, it can only function imperfectly. But more people who are willing to give up provider choice enroll in HMOs one way or another. (Reschovsky 1999/2000).

⁴³ In fact, if this were true, HMOs might even increase total costs, as some studies have implied. But even if HMOs' lower premiums were entirely attributable to favorable selection, it would still be possible to argue for a competitive effect. It is possible that HMOs, by enrolling many of the healthy, force other plans to be more efficient in the treatment of their enrollees, who are on average less healthy. However, if HMOs truly have no incentive to enroll persons other than the very healthy, the competitive effect on the treatment of the less than very healthy could not be very great.

⁴⁴ In the abstract, it is possible for HMOs to reduce policyholders' co-payments while increasing the premiums paid by employers. However, as noted above, Baker et al (2000) indicates otherwise. Because employers defend their own interests, HMOs have had to offer lower premiums hand in hand with lower co-payments to build market share. Since the current trend is for employees to pay a larger share of premiums, it seems unlikely this pattern will change.

more expensive plans but some would forego insurance. Then overall health care costs might be lower, the same as, or higher than we currently observe. However, all those current HMO enrollees who would prefer HMO insurance would be unambiguously worse off.⁴⁵ Additionally, a higher percentage of people would be uninsured, so it seems likely that society as a whole would be worse off without HMOs.

Therefore, the evidence indicates that HMOs provide value to many of their enrollees, but HMO insurance is inherently unattractive to much of the rest of the population. This suggests that the natural market for HMO insurance is not the entire population, but a subset whose tastes and circumstances predispose them to enroll. When this inference is placed within the product life cycle framework (LA Brown 1981), it raises the possibility that the HMO industry can reach maturity well before 100% of the population is enrolled. In fact, the size of the natural HMO market is an empirical question, and there is no strong prior reason to believe one particular percentage is “ordained” rather than any other percentage.

In Chapter 1, it was suggested that the HMO industry might have shifted from the growth stage to the maturity stage during the 1990s. The key finding of Kemper et al offers a possible explanation of why the industry would reach maturity with enrollment well below 100% of the population. The research results of this study, presented in Chapter 4 and evaluated in Chapter 5, offer further evidence concerning this key issue.

⁴⁵ In this scenario, some persons currently enrolled in HMOs against their preference for more generous coverage would be better off if HMOs were eliminated altogether. However, some with similar preferences would be forced to go without insurance altogether. It seems likely the welfare loss of the latter would at least equal the welfare gain of the former.

B. The Relationship Between HMOs and Physicians

Given the conflicts between HMOs and physician organizations early in HMO history, it is not surprising that many researchers have focused on the HMO-physician relationship (Luft 1987). The severe legal and other conflicts of the 1940s and 1950s gave way to grudging acceptance of HMOs by organized medicine by the 1970s. However, a degree of tension has continued to characterize the relationship. HMOs need physicians to provide and supervise care, but many physicians, often the majority, have had unfavorable attitudes toward HMOs. Since the 1970s, the relationship between HMOs and the physician profession has been complex, and a variety of factors have influenced the decisions of individual physicians to affiliate or not to affiliate with an HMO.

During the 1970s, entrepreneurs trying to introduce HMO insurance in many cities of the U.S. had considerable difficulty attracting a sufficient core of physicians to make new HMOs functional. In some cases, only under-employed physicians were willing to affiliate. Even if a new HMO was operating and growing, physician recruitment issues often constrained expansion. If newcomer physicians lacked the enthusiasm and commitment of the founding physicians, rapid growth could lead to large losses and financial disaster (LD Brown 1983).

By the 1990s, large numbers of physicians were affiliated with HMOs, and physician relationship issues were somewhat different. In order to keep expanding, HMOs felt pressure to contract with virtually all physicians in their geographic area. Additionally, different physician groups and organizations held vastly divergent attitudes toward capitation risk, utilization management and related operational issues. Therefore, HMOs frequently were asked to cater their contractual arrangements to the evolving norms of different metropolitan areas and the particular tastes of specific physician groups (Robinson 1999).

During the rapid HMO growth of the 1990s, a number of surveys indicated that many physicians had reservations about, or were completely disenchanted with, their relationships with HMOs. Some of the areas where HMOs were most heavily criticized included physician autonomy, authorization policies and evaluation of physician performance (Borowsky et al 1997). There was some indication that physicians who were heavily dependent on HMOs for incomes rated HMOs less favorably than those who received little or no income from HMOs (Baker and Cantor 1993). Certain HMO policies, such as gag clauses, provoked widespread protest and legislative action in a number of states (LA Times 1996). A statistical study found that HMO growth is associated with reductions in the numbers of medical and surgical specialists on the MSA level (Escarce et al 2000).

In conclusion, HMOs' relationship with physicians is a critical area affecting the success and evolution of this industry. There is evidence that the recruitment of physicians was a limiting factor for the growth of HMOs in the 1970s. The relationship issues were somewhat different, but still critical, during the 1990s. During this period of widespread HMO success, many physicians displayed dissatisfaction with, or avoidance of, HMO affiliation. Thus physician relationship issues may still have constrained HMO growth.

C. Role of Market Power

Today HMOs have a presence in virtually every state, and they enroll over one quarter of the U.S. population. Traditional fee-for-service plans have shrunk to a small percentage of the commercial insurance market, and most people are insured by some type of managed care plan.⁴⁶ A strong case has been made that only HMOs and PPOs in competition with each other fully exploit their market power to leverage lower prices from hospitals and doctors (Melnick et al. 1992; Zwanziger et al. 2000).

Melnick et al reached this conclusion by considering HMO and PPO relationships with hospitals. Managed care plans' selective contracting with hospitals, permitted in California since 1982, enables plans to take advantage of excess capacity in competitive hospital markets. Particularly if a small number of plans constitute an oligopsony, the plans tend to extract substantial concessions from hospitals and pass much of the savings onto their policyholders.

The authors analyzed market power factors affecting hospital payments of the California Blue Cross PPO, a large insurer. They found that hospitals with strong market power within their geographic area were able to collect higher fees from the Blue Cross PPO. But if Blue Cross had market power over the hospital (i.e. if Blue Cross provided a high percentage of the hospital's patients), then Blue Cross was able to make lower payments. These results reinforce the authors' hypothesis that hospital charges to insurers, which account for the largest fraction of private health care expenditures, are partly determined by the relative market power of the hospital and of the insurer. They maintain that the same factors would influence HMO payments. Based on this, the authors imply that merger or consolidation of insurance plans would reduce hospital costs. However, mergers or consolidation of hospitals would lead to large hospital price increases. A later study reinforced these conclusions (Zwanziger et al, 2000).

Wholey et al. (1995) took a complementary approach. They found, with some caveats, that the greater the index of competition among HMOs at the county level, the lower HMO premiums are. These results indicated that market forces worked through HMOs to lower overall premiums. However, these authors also found that a higher level of HMO market share without greater competition leads to higher HMO premiums. They attribute this result to adverse selection stemming from the movement of high-risk persons from FFS to HMO insurance as HMO enrollment grows in a given market.

⁴⁶ On the role of preferred provider organizations (PPOs), please see "(Loosely) Managed Care Is in Demand", NYT, 9/29/98.

D. Evolution of HMOs

In the past 30 years, HMOs have undergone three major changes: 1) the shift from Staff/Group models to predominantly IPA models, 2) reduction in ownership by local non-profit bodies, and 3) increasing reimbursement of M.D.s by capitation and FFS instead of salaries. Recent evidence suggests that capitation has peaked as a means of reimbursing physicians (Dudley and Luft 2001).⁴⁷ HMOs currently rely upon careful selection of provider networks, negotiated fee for service schedules and utilization review as preferred methods of influencing physicians to help control medical costs.

These changes can be seen as part of the overall evolution of HMOs toward a more business oriented organization, as opposed to the community oriented and paternalistic arrangements that characterized their early years. Some HMOs continue to require referrals by the patient's primary care provider (PCP or gatekeeper), but recent pressure from patients for more convenient and rapid access to care has jeopardized reliance on PCPs (Dudley and Luft 2001). HMOs continue to evolve, and the future may see different aspects and potentials of HMOs coming into prominence.

InterStudy (1994) reported that over two-thirds of HMO enrollees belonged to IPAs or mixed model HMOs that operated predominantly on an IPA basis. Many of these IPAs are owned by national HMO organizations, including those controlled by the Prudential, CIGNA and Aetna insurance companies.

Robinson (1999) analyzed the future evolution of HMOs. He predicted that most HMOs will cease to be separate insurers and that they will become part of large national insurance companies that also offer PPO and perhaps other insurance products. He refers to these companies as multi-market, multi-product health plans. Robinson predicted that continuing efforts of HMOs to own exclusive provider networks would fail. He argued that vertical disintegration will prevail, with hospitals, physicians and HMOs all remaining independent. Robinson also forecast an "upside down" version of managed competition, with employers only contracting with one insurer but giving employees the choice of several plans from that insurer. Robinson's arguments, based primarily on trends observed in the 1990s, seemed compelling at the time, and many multi-product firms, including Blue Cross affiliated insurers, have prospered. However, other evidence, including the difficulties faced by Aetna, the most aggressive multi-product firm, suggests that his scenario may be too simplistic.

⁴⁷ Table 2 of this article shows that capitation accounted for 8.4% of total physician revenues in 1996 but only 7.4% in 1999. This article cites of increasing rejection of capitated contracts by large medical groups in California and elsewhere.

E. Causes of Enrollment Growth

As indicated above, the overwhelming majority of academic studies have focused upon the *effects* of HMO enrollment growth, particularly the effect on health care costs. There is a small body of literature (McNeil and Schlenker 1975, Cromley and Shannon 1981, Wholey et al 1990, and Christianson et al 1991) that addresses the entry and exit of HMOs on the MSA level. This literature address only one aspect of the HMO enrollment growth process, but it provides useful background for the more comprehensive studies.

Cromley and Shannon used discriminant analysis to estimate the establishment of new HMOs during the year 1980 in 243 metropolitan areas. It should be noted that they analyzed new HMO establishment regardless of pre-existing HMO operation in each MSA. Their discriminant model shows that a high population to physician ratio was negatively associated with HMO establishment. This is equivalent to a positive association between physician concentration and HMO establishment. Wholey et al analyzed the effect of state regulations on the entry and exit of HMOs from MSA markets. They found that state regulations could affect the entry and exit rates of HMOs without profoundly affecting the number of HMOs in a market.

Extensive search and review of the literature has only yielded 4 prior studies of the *causes* of growth (Goldberg and Greenberg 1981; Morrissey and Ashby 1982; Welch 1984; Dranove et al 1998). The first 3 cover time periods ending from 1976 to 1980, and their analyses are broadly comparable with this study's analysis of the period from 1973 to 1978. The most recent published study is concerned with explaining total managed care enrollment as of 1994 and 1995. This makes it roughly comparable with the analysis of HMO growth from 1988 to 1993 featured in this report. These 4 articles are summarized and compared in terms of data, methods and results.

Data:

Goldberg and Greenberg (1981) use state level data for 1966 and 1976. The key outcome variable was the percentage of insured persons enrolled in HMOs in 1976. The authors tested several standard economic and demographic measures as explanatory variables. They also used an estimate of hospital costs and estimates of the percentages of physicians in group practice and in state medical societies. Finally, they tested six dummy variables relating to state regulations; these included certificate of need (CON) requirements, requirements that HMOs be non-profit and prohibition against HMO advertising.⁴⁸

Morrissey and Ashby (1982) use MSA-level data from 1975 and 1978 for over 250 MSAs; about 70 of these had an HMO presence at the time. Percentage of population enrolled in all HMOs and the presence of an HMO in an MSA were the dependent variables. This article did not use the legal variables used by Goldberg and Greenberg, but, in some runs, it used 1975 HMO market share to predict 1978 values. Two key explanatory variables were the age of the oldest HMO in the MSA and a measure of medium and large employers. It also used

⁴⁸ The 1973 Federal HMO Act exempted HMOs from state restrictions on advertising, so such prohibitions ceased to be a factor after that.

data on the supply of doctors and hospital beds, the ages of doctors, the percentage of doctors and the percentage of employers with 250 or more employees.

Welch (1984) also uses MSA level data, but only considers enrollment in Prepaid Group Practices (PGPs), those HMOs that essentially provided full-time employment to their affiliated physicians. This study analyzed enrollment data for each of the years from 1977 to 1980; during those years PGPs accounted for over 80 percent of HMO enrollment. It used a one-year lag of MSA PGP market share as an independent variable. It used several of the same economic and legal predictor variables as Goldberg and Greenberg.

The more recent study, (Dranove et al 1998) considers combined HMO and PPO penetration (i.e. managed care enrollment and expenditures), using 2 different measures to estimate this value. Using physician survey data, these authors estimate the percent of physician revenues from managed care for 276 MSAs. They also use market research survey data to estimate managed care enrollment for 99 of the larger MSAs. The market research data show a much higher extent of managed care penetration than do the physician revenue data. The authors suggest that this difference is due to managed care enrollees having, on average, lower utilization than enrollees in other types of plans. They regress these 2 outcome variables on 3 variables relating to health care supply and on 10 general economic and demographic variables.

Methods:

Goldberg and Greenberg applied Tobit analysis to the data, arguing that it was appropriate because of the large number of zero values in the dependent variable (limited dependent variable). The final model (Equation 6) excludes several variables that were statistically insignificant and judged unpromising based on previous runs. Morrissey and Ashby basically deal with the limited dependent variable problem by running two separate models, the first for presence of an HMO and the second for HMO market share conditional on presence; both models are constructed on a simple ordinary least squares (OLS) basis. The article presents both models with and without the three-year lag variable.

Welch argues that the only "correct" models include lag variables because adjustment to changes in the environment takes time and therefore HMO market share at any point in time is a product of its prior state and changes in exogenous factors. However, this argument can be disputed on both practical and conceptual grounds. On practical grounds, the lagged values will usually absorb much of the variance in the current values, making it difficult to spot relationships that may be more interesting and specific to health economics.

Conceptually, one would want to use *changes* in the predictor variables if one thinks that they are what drive changes in HMO market share, which Welch did not do. In fact, MSA level changes in variables such as education level or net immigration tend to be incremental and their true effects on HMO market share are probably not measurable. It is robust differences *between* MSAs, if anything, exerting themselves over time that account for the observed differences in HMO market share. Lags in effects occur, but that does not necessarily mean that all models must include lag terms. Welch also uses a double log functional form and Heckman and Lee's estimator for a censored sample; this last is a generalization of Tobit analysis that allows for different effects on probability of existence and on magnitude of enrollment.

Dranove et al address managed care enrollment and growth in the 1990s. Since sampling error is a major issue with their Physician Survey, they apply Weighted Least Squares (WLS) regression.⁴⁹ To avoid potential endogeneity, they use supply variable data from 1985 or 1980.⁵⁰ They also separately estimated parameters for markets with high and low hospital concentration.

Results:

Most included variables did not pass the test of statistical significance in each of the four studies. Goldberg and Greenberg found four variables were marginally statistically significant (i.e., significant at the 10 percent level); these were unionization of the metro labor force, net migration as a measure of population mobility, hospital costs and the prevalence of physician group practice. Each was positively related to HMO enrollment. The article's Regression 6, which included these four variables and two non-significant state regulation variables, achieved an R^2 of 33 percent. The results were slightly better when change in HMO market share from 1966 to 1976 was the dependent variable. The lack of solid statistical significance and the relatively low R^2 must have been disappointing, but this article did make a start toward understanding this phenomenon.

Morrissey and Ashby used several of the same variables as Goldberg and Greenberg. Of these, Morrissey and Ashby found a marginally significant *negative* effect for per capita income (Goldberg and Greenberg found no hint of any effect at all). On the other hand, Morrissey and Ashby did not find remotely significant effects for unionization or the prevalence of group practice. Net migration had a statistically significant positive effect on conditional market share.

For other variables, Morrissey and Ashby found solidly statistically significant effects for age, gender and race. Those equations with a three-year lag term found it to have a very significant positive coefficient, with its inclusion raising R^2 by .20 to .72. The lag term would probably have had even more effect had not an "Age of Oldest HMO" variable, reflecting the age of the oldest HMO in the MSA, been included in both equations. The inclusion of both of these variables accounted for most of the improvement in R^2 over Goldberg and Greenberg.

The lag term is even more significant in Welch, which is to be expected since it is a lag of only one year and the "Age of Oldest HMO" variable is not included. Here the lag explains roughly 50 percent of the variance and has a t-value of about 25. This study shows per capita income to have a negative effect, which attains solid statistical significance in one specification. Significant effects were found for hospital cost and education level (both positive). Three of the included state regulation variables had no significant effects, and restrictions on the employment of physicians were found to have a significant negative effect, as might be surmised. However, CON laws were found to have a significant positive effect, which was

⁴⁹ They use number of physician respondents to weight their physician survey data. They also applied WLS regression to their market research survey data. It appears that the market research data for some cities are based on sample sizes of 200 or lower.

⁵⁰ As explained in Chapter 3 of this report, HMO penetration may influence health care supply variables simultaneous with those variables influencing enrollment. Dranove et al address this problem by using supply data from long before the date of their HMO penetration data. Their method is comparable to the method used in this study.

surprising at first.⁵¹ The author tried running the model without the CON variable, and the other coefficients changed only slightly. R²s for both models were between .80 and .85.

Taken together, the three studies indicate that income has a negative effect on HMO Market Share conditional to HMO existence. The two articles that looked at average hospital cost both found a positive effect. Two studies found substantial evidence of a positive migration effect, while the other (Welch) found a statistically insignificant negative migration effect. The two that looked at group practice found materially positive effects that were not statistically significant in either case. Possibly a meta-analysis could quantify the extent of agreement among these articles and this study.

The results are muddier for the other variables. Population mobility, unionization, lags (discussed above), restrictions on M.D. employment and CONs were each examined by two of the 3 articles. The positive unionization effect was close to significance in Goldberg and Greenberg, but very weak in Morrissey and Ashby. Restrictions on physician employment had virtually no effect in one study, but a significant negative effect in the other.⁵²

For the 1990s, Dranove et al find that health care supply variables, particularly hospital concentration, had a substantial effect on managed care penetration. MSAs with high hospital concentration tend to have lower managed care penetration, and there is some indication that hospital concentration may have an effect on the coefficients of the other supply variables (hospital occupancy and extent of solo practice by physicians). However, the authors also find that lowering hospital concentration beyond a certain point may have little effect on managed care penetration. Their results provide moderate evidence that increasing hospital occupancy rates and a greater degree of solo practice by physicians tend to decrease managed care penetration.

Concerning economic and demographic variables, Dranove et al find a negative relationship between managed care penetration and a large non-white population but a positive relationship with college education and urbanization at the MSA level. They also found a non-linear relationship between managed care and firm size, with higher managed care penetration associated with higher concentrations of large firms (over 500 employees) or moderately small firms (20-99 employees).

After my findings are presented in Chapter 4 of this report, Chapter 5 offers a comparison of them with the published findings for both the 1970s and early 1990s periods.

⁵¹ Possibly institutional providers were more favorable to CON regulation in states where HMOs were emerging. In such a case, the causality ran from HMO growth to CON regulation rather than vice versa.

⁵² Please see Chapter 5, Section B2, Comparison With Prior Studies, for an analysis of how the results generated here compare with those of these three prior studies.

F. Chapter Summary and Significance

This chapter covers the literature on the economics of HMOs. This literature includes extensive analysis and debate concerning the nature of HMOs' cost advantage. The preponderance of the recent literature indicates that the cost advantage is not solely due to a selection effect. It may be largely due to lower reimbursement to providers. Other major topics include the relationship between HMOs and physicians, the role of market power and the evolution of HMOs. Section E of this chapter covers a small number of studies that analyze the causes of HMO enrollment growth.

The published literature contributes to the main analysis because prior studies and findings were used to generate hypotheses that are tested by this study. The specific hypotheses are discussed in Chapter 3. One important hypothesis is that an abundant supply of physicians was conducive to HMO growth. This hypothesis relates to the literature showing that physician recruitment is a critical prerequisite for HMO success. Another physician variable, prevalence of physician group practice has been found to have some influence on HMO growth in one prior study.

Another major hypothesis is that an abundant supply of hospital beds was conducive to HMO growth. It is well established that HMOs achieve a cost advantage by reducing hospital costs. Additionally, three prior studies found that hospital related variables have an effect on HMO growth.

Because of their lower costs, HMOs have often been thought of as a viable health insurance alternative for working class and lower middle income families and individuals. Two prior studies have provided some indication that per capita income has a negative relationship with HMO growth. Therefore, this hypothesis is emphasized in Chapter 3.

In conclusion, review of the literature reinforces the point that HMOs are an important and unique component of the health economy; they have produced substantial effects on health care costs and other aspects of the health care sector. The literature also suggests several specific hypotheses concerning the causes of HMO growth, and these hypotheses are developed and tested in the next 2 chapters.