

# CHAPTER 11

---



## Managed Care

---

**H**ealth care has evolved in recent years into an economic form that defies comparison with other sectors of the economy—the “managed care” sector. One cannot contemplate the modern U.S. health care system adequately without understanding the nature of managed care organizations (MCOs), what they attempt to do, what management tools they have at their disposal, and why consumers would bother to join them (instead of staying with a traditional insurance plan such as discussed in Chapters 4, 5, and 10). Managed care spans a wide variety of organizations, but the key idea is that MCOs are actively involved both in the delivery of health care and in the provision of health insurance. Sometimes they originate as health care delivery organizations, and more often as insurance companies, but as they evolve into the “middle ground” they can begin to look quite similar, no matter what their origins or history. They come in many flavors; Box 11.1 provides definitions of the most common types.

We begin this chapter with a discussion of the consumer’s decision—why would a consumer enroll in an MCO—and then turn to the incentives available to managed care organizations (both provider and consumer interventions) to accomplish their goals?

**BOX 11.1****THE ALPHABET SOUP OF MANAGED CARE**

The world of managed care organizations (MCOs) has evolved into a series of loose classifications that, it turns out, are almost invariably examples of TLAs.\*

**HMO:** The “classic” and most widespread phrase for these organizations is the Health Maintenance Organization, a phrase coined in the 1970s meant to describe a fairly specific organization in which the insurance plan, the doctors, and the hospital either were all the same organization or were closely affiliated. The doctors usually worked in the same (large) medical office building, often adjacent to or part of the hospital itself. The original Kaiser Permanente plans on the west coast were perhaps the prototypes. In such plans, the doctors are on salary, the plan owns the hospitals used by patients, and patients (except in emergency) must use the plan-affiliated providers for their care.

The essential feature of the pure “staff model” HMO is that the plan is paid on a “capitation” basis (“per head”), which means that they receive a fixed sum of money each year to provide all of the medical care the enrollee needs, no matter what. This offers the greatest incentive for cost control and cautious use of medical interventions (and to some observers, an incentive so strong that it might harm patients’ health).

*The term has evolved in common use to imply almost any kind of managed care organization and in some settings has almost the same meaning as MCO.*

**FFS:** Fee-for-Service, the antithesis of MCOs. Traditional insurance (including Medicare at its inception) offered defined coverage for billed charges for any service provided for the patient. FFS creates the least constraints on both provider and consumer behavior, and has higher costs for a given degree of its financial protection to consumers because of its unconstrained environment.

**IPA:** Independent Practice Association, a step away from the “pure” HMO. In an IPA, the doctors work as independent doctors in their own offices, and normally have a mix of patients enrolled in traditional FFS plans as well as capitated or “managed” patients.

**PPO:** Preferred Provider Organization. A PPO is a cousin of the IPA, but relies more on selecting a subset of all available providers and making contractual agreements with them that mostly focus on price. The “preferred providers” have patients steered to them by the PPO insurance plan in exchange for a lower price. The lower price allows the PPO to offer a lower total insurance premium, which attracts enrollees, which in turn creates a larger patient base which the PPO can use to negotiate

\*Three-letter acronyms.

**BOX 11.1** *Continued*

with providers. Doctors are paid on a fee-for-service system, usually with a set of negotiated fees for each service offered.

**POS:** Point of Service Plans. Often called HMOs without walls, POS plans have a capitation arrangement with providers for their enrollees, but the doctors are spread around in their own offices as in an IPA or PPO. The key difference is that in a POS plan, the compensation to the provider is per patient per year rather than FFS. Patients have strong financial incentives to use the POS plan doctors, since their insurance either covers nothing if they use out-of-plan doctors (the extreme form) or has a very high copayment rate.

## WHY MANAGED CARE?

Managed care organizations evolved as a response to a problem created by a solution to another problem. Health insurance (as we saw in Chapter 10) exists fundamentally to reduce individuals' financial risks. Variability in health outcomes provides the fundamental source of risk, and intelligent responses to those illnesses and injuries (buying appropriate medical care) create the derivative financial risk. In response to these risks created by illness, individuals chose to purchase health insurance. But the traditional forms of that insurance, including classic Blue Cross and Blue Shield "first dollar" coverage and then "major medical" insurance (first sold by commercial insurance companies) created another problem in return as they sought to reduce the financial risk—they altered people's behavior in choosing medical care in an ultimately undesirable way. Managed care evolved to help deal with that problem.

As we learned in Chapters 4 and 5, traditional health insurance creates a welfare loss by reducing the price of medical care at the time consumers (with their doctors) choose how much medical care (including what types of treatment, training of specialist, etc.). With an artificially low price, consumers are led to consume too much care, but (as we can infer through the voluntary purchase of the insurance) they gain more through the reduction in financial risk than they lose through excessive consumption of medical care. The discussion in the preceding chapter shows how a rational consumer balances off those gains and losses, and how the "optimal" insurance differs greatly, depending on the risks arising from the illness (and treatments thereof) and the elasticity of demand for the medical services themselves.

Managed care seeks to improve upon the problems created by traditional insurance by interfering in some way with the decisions about the amounts and types of health care chosen. Contracts between the consumer and the managed care organization limit the consumer's access to medical care in some ways in order to offset the incentives for overuse of medical care. The consumer who enrolls in an MCO in effect says "stop me before I kill again" (although perhaps with less extreme wording). The MCO inserts various incentives and constraints into the system to reduce the quantity of care and to shift to lower cost alternatives when more than one treatment choice is available.<sup>1</sup>

Managed care organizations are still evolving in their understanding of which interventions actually work, which seem to be acceptable to consumers, and how to price the insurance contract when including various combinations of "management." The mechanisms used to constrain health care consumers' choices affect both provider and consumer behavior. Because of the important "principal-agent" relationship between doctors and patients (whether involving demand inducement or not), influencing provider behavior has proven important in these arrangements. However (as we shall see below), those mechanisms that rely upon altering provider behavior are subject to potential legal attack, since they inevitably alter the principal-agent relationship between the patient and the doctor. Subsequent sections in this chapter explore the main types of these "management" tools and analyze their consequences.

At this point, it is worth noting that MCOs have not succeeded in the market "merely" because they have created insurance plans that cost less than standard fee-for-service (FFS) plans, although to the casual observer this may seem to be the primary reason for their success. However, lower premium costs *alone* would not necessarily attract rational consumers to MCOs. One can always achieve a lower insurance premium, of course, by accepting insurance with reduced coverage (higher copayments and deductibles, for example) or reduced scope of benefits (no coverage for prescription drugs, dental care, etc.). But that simply moves the consumer along a tradeoff of lower cost premiums in exchange for greater financial risk. The essential value of the

---

<sup>1</sup>Various alternatives exist, particularly in single-provider or single-insurer systems of countries such as Great Britain and Canada, to achieve the same goals. In those (and similar) countries, access to care is deliberately limited by availability of various services, and commonly rationed by waiting lists. (For an elegant discussion of the operation of the British National Health Service, see Aaron and Schwartz, 1984.) Various regulatory controls in the United States have attempted to deal with the same issue at the governmental level, as Chapter 15 will discuss in detail.

MCO comes in its ability to reduce insurance premiums to the consumer while at the same time retaining the elimination of financial risk that a comprehensive FFS insurance plan offers.

When consumers decide to enroll in an MCO (instead of a traditional FFS plan), they in effect precommit to a plan that they know in advance will constrain their choices, because such precommitment can actually improve their overall expected utility. Analysis of economic “games” commonly shows the value of precommitment to rational economic agents, both on the consumer and supplier side of the market.<sup>2</sup> Box 11.2 discusses an ancient and famous legend involving precommitment as well.

## Enrollment Patterns

The most recent information available about enrollment trends shows a rapid switch to managed care of some sort. In the early 1980s, most people had some sort of FFS plan. In a 1991 survey of the entire market, 13 percent of the population had some form of HMO insurance, the remainder holding FFS plans. Of those, IPAs, PPOs, and POS-like plans had by far the largest share, and “pure” HMOs (in the traditional sense) were down to 3.2 million enrollees, barely above 1 percent of the U.S. market.

By 1996, the best data summarize plans obtained through employment groups, but since that represents the vast majority of all private health insurance, we can take these data as representative of the market as a whole. By 1996, only 27 percent of these enrollees persisted in an FFS plan, and by 1999, only 8 percent remained.

Thus it seems fair to say that at this point in the evolution of health insurance, traditional FFS insurance has almost evolved out of existence. During the evolution to managed care, the most rapid growth has come in PPOs, which increased from almost no market share in the early 1980s to 28 percent

---

<sup>2</sup>For example, on the provider side, precommitment can help organize an implicit collusion between sellers, where a “dominant firm” precommits to match the price of any lower price. This precommitment dissuades other firms in the market from cutting their prices, thus creating a tacit collusion to keep prices high. On the consumer side, a common form of precommitment can be seen in the purchase of U.S. Savings Bonds. The consumer makes a small investment up front in the bond, and its value grows through time, but the owner cannot “cash out” the bond until a specified number of years have passed. This precommits the bond owner to a plan of savings that could be achieved through smaller annual deposits into a savings account, but without the precommitment, the consumer runs a risk of failing to meet the annual savings goals, and hence not having as much money saved at the end.

**BOX 11.2****A CLASSIC MYTH INVOLVING PRECOMMITMENT**

An early Greek myth helps us understand the concept of precommitment—Ulysses (or Odysseus), the hero in Homer's *Odyssey*. Ulysses deliberately constrained his own behavior in advance in order to accomplish something else he greatly desired. His precommitment prevents him from taking an irrational (and in his case, fatal) course while still allowing him to achieve other goals. In most economic models involving precommitment, the same issues arise.

Ulysses, returning home from his travels, must sail by an island populated by some apparently irresistible women called Sirens, the songs of whom attracted sailors to hear them, but, alas, always fatally, as they crashed their ship upon the rocks.\* Ulysses wished to hear the famous song, but obviously did not wish to crash his ship, so he filled the ears of all his crew with wax so they could not hear, and then had himself carefully tied to the mast of his ship so he could not alter its course. The crew sailed by safely, and Ulysses was able to hear the famous Sirens' Song, but only by having arranged in advance that his own actions would be constrained. Thus he was able to accomplish his immediate goal of hearing the Sirens, and also to accomplish his long-run goal of maintaining a safe passage homeward. Ulysses' idea of "tying himself to the mast" has become symbolic of somebody who takes actions in advance to constrain his own later actions. Managed care has many of these same elements for consumers of medical care.

\*The motives for the Sirens to do this were never made clear by Homer, but perhaps they profited by scavenging the remains of the ships. Homer does allude to the scattered ruins of the victims' ships on the rocks.

in 1996 and 41 percent by 1999. POS plans, a relatively new entrant, moved from 14 percent in 1996 to 22 percent in 1999. Classic HMOs (combining pure HMOs and IPAs) remained about static between 1996 and 1999 at about 30 percent. (All 1996 and 1999 data are from Dudley and Luft, 2001.)

## TYPES OF INTERVENTIONS

To deliver on its promise to lower costs of health insurance while at the same time controlling risk to the consumer, MCOs must find acceptable ways of limiting access to medical care. They do so through altering the incentives

and “rules of the game” for both consumers and providers of health care. In this next section, we will learn about the mechanisms commonly employed by MCOs to achieve their goals.

## Consumer Side

A number of interventions in managed care deal directly with consumers’ behavior, seeking to modify their choice in seeking care and their decisions about what care to receive once their doctor has made a recommendation.

**Copayments** The most common and readily understood intervention used by almost every managed care plan is the simple imposition of copayments at the time medical care is used. Most plans now use a system in which patients pay a fixed amount on each occasion (for example) that they see a doctor. This amount typically ranges from \$5 to \$20. The consequences of such a copayment approach were explored in Chapter 4, and will not be discussed further here, except to note the issue about search incentives. As we learned in Chapter 4, a copayment model such as this obliterates the incentives for consumers to search for a lower price, since (once having paid the fixed amount per physician visit, prescription, etc.) the consumer has no financial advantage from seeking a lower priced provider, and hence presumably has minimal if any incentive to carry out any search. However, almost every managed care plan in operation already has negotiated prices with providers who are eligible for treating patients in the plan, in effect having done the searching for the consumer. Thus they really don’t care if the consumer searches or not, and are quite willing to have a copayment system that does not encourage search. The advantages to the insurance plan from using something like a \$10 copayment appear to center mostly on administrative costs—the doctor’s office collects the copayment at the time of the visit, and the insurance company has already agreed how much it will pay the doctor for each procedure or visit delivered to the patient.

**Second Opinion Programs** Another approach used in some managed care plans offers (and in a few occasions, requires) “second opinions” before they will pay for surgery. In a second-opinion arrangement, the insurance plan will pay for a second doctor’s opinion, but it is made clear to the second doctor that they will not provide the treatment. They are asked only whether they concur with the original recommendation.

The second opinion process focuses on the intrinsic financial incentive problem embedded in the doctor–patient relationship in a fee-for-service arrangement. The “agent” (the doctor) has financial incentives to recom-

mend treatment more aggressively than the “principal” (the patient) might prefer (Dranove, 1988a); because of asymmetric information, the patient cannot always tell when the advice has been altered in an important way. By offering to pay for a second opinion, the insurance plan not only signals to the patient the possibility of receiving “tilted” advice from the doctor, but also creates a way for the patient to get advice from a doctor who has no possible financial stake in the outcome.

Despite the intrinsic appeal of second opinion programs, they have not worked very well in altering patients’ choices, and relatively few managed care plans now require them, at least in widespread use. Three things work against the second opinion program: First, patients are often reluctant to use the second opinion when it seems as if they are challenging their doctor. Next, the second opinion almost invariably comes from another doctor in the same region (since for it to be otherwise, the patient would have to travel some distance for the second opinion). This means that the “styles” of doctors are likely to have some within-region similarity, as we learned in the analysis of regional variations in Chapter 3. Thus an initial recommendation for surgery will likely find a concurring second opinion even if the doctor giving the second opinion gives a completely honest opinion about the case, since we know that doctors within the same region have related styles of decision making.

Third, alas, we cannot fully expect that doctors will give completely honest opinions even if they have no direct financial benefit (e.g., doing a surgery) from the patient. The problem arises from the possibility of an “implicit collusion” that takes place in many settings in which participants play in a “repeated game” problem. To understand this, we must take a slight side-excursion into repeat-play game theory.

In repeat-play games, the agents interact repeatedly in similar circumstances, and often behave in such occasions differently than they would in a single-play game. People often observe, for example, that a “short timer” (somebody about to leave their job, their team, or the military) often behaves differently than others in the same setting. This occurs because the opportunities for others to take retribution are more limited. Similarly, your local mechanic is less likely to cheat you in any individual auto repair event than the mechanic in a distant town where you happen to need repair while passing through. Repeat-play games are the essential feature of many businesses’ (and doctors’) efforts to build a reputation—such efforts have no economic value except when the customer might return (a repeat-play situation).

Repeat-play games have formed the basis for a large number of economic studies. One of the most fascinating of these emerges from a computer-based competition in which entrants submit computer-program “agents” into a world in which they interact with other agents, but none know in advance

how each agent is programmed to respond when various forms of cooperation or noncooperation are possible. Without any participant announcing their strategy in these computer-based competitions, the agents learn others' strategies through time, and the winning strategy thrives and prospers through time. It turns out that a strategy that almost invariably survives best in such settings is a simple "tit for tat" strategy, namely "if you cooperate with me, I'll cooperate with you next time, and if you don't cooperate, I'll retaliate on the next interaction (or the next several)." Remarkably, the tit-for-tat strategy dominates others even when some of the entrants are programmed specifically to defeat a tit-for-tat strategy. And finally, the type of behavior that commonly emerges in such games is often a tacit agreement (never specifically stated) to cooperate.<sup>3</sup>

In second opinion programs, a similar repeat-play situation emerges. A patient from Doctor A goes to see Doctor B for a second opinion. But Doctor B knows that their roles might be reversed at some time in the future, with Doctor A rendering a second opinion about the value of Doctor B's recommendation. This will invariably alter the advice rendered by Doctor B, making the recommendation more likely to concur with the original recommendation. Thus, the commonality of practice styles within a region and the implicit cooperation that emerges in a repeat-play game both lead to the strong likelihood that a second opinion will match the recommendation of an initial opinion. Because of this, we actually have good economic reason to believe that second opinion programs will not likely work very well.

The evidence on second opinion programs generally supports this conclusion. Private insurance second opinion programs have achieved only moderate success, if any, in controlling medical care costs. One reason may be that the insurance coverage for second opinion plans changed only a few people's use of the idea (although there is no evidence on this point). Despite some widely publicized successes, e.g., in Massachusetts Medicaid programs (Gertman et al., 1980), the analysis done on second opinion programs has been filled with methodological problems. Perhaps most importantly, none of the studies published to date uses the most important scientific model, having a well-defined and appropriate "control group" to estimate what happens without the "treatment" of a second opinion program. (Almost all of the studies rely on historical controls, which are particularly problematic in a world in which other features of health care delivery and compensation change frequently and rapidly.) An analysis of this literature (Lindsey and Newhouse, 1990) concludes that there is no evidence of any effects of volun-

---

<sup>3</sup>For one of the landmark works in this area, see Robert Axelrod's *The Evolution of Cooperation* (1984).

tary programs, and only weak evidence of poor scientific quality for effects of mandatory programs.

**Gatekeeper Models** Another approach, widely used in managed care programs, is called a “Gatekeeper” model. In these managed care programs, the patient *must* see a primary care doctor (family practice, internal medicine, pediatrics, or in the case of pregnancy, obstetrics) before seeing any specialist, and a specific referral from the primary care provider (PCP) is necessary before the insurance plan will pay for specialist treatment. The PCP visit obviously costs the insurance plan something, so they must expect some reduction in treatment to occur to justify that expense.

The belief that overall treatment costs will fall in a Gatekeeper model comes from several sources. The key situation is one in which either the PCP or a specialist might treat the patient (and many illnesses and injuries fall into this category). The specialist normally receives a higher fee than the generalist (the PCP), and at least on some occasions, we know that specialists tend to use other medical resources (laboratory tests, hospitalization, etc.) more intensively than generalists.<sup>4</sup> Thus the insurance plan can rely on the normal economic incentives for the generalist to “hold onto” the patient rather than referring the patient to a specialist, and can anticipate that the total costs of treatment will be lower than if the patient were treated by a specialist. The basis of the gatekeeper model, then, is to prevent the patient from bypassing the PCP and going directly for specialty treatment, a specific example of “tying oneself to the mast.”

Data supporting the efficacy (or not) of gatekeeper models are extremely sparse. One early study of the first gatekeeper model in Seattle, WA (Moore, Martin, and Richardson, 1983) provides a case study of a failed plan (which offered incentives to providers that did not support the gatekeeper concept).

## Provider Side

Traditional health insurance sought to control health care costs exclusively through the use of consumer-side incentives (copayments and deductibles). Modern managed care plans use such strategies, but also deal with the incentives faced by providers. This section discusses the major utilization and cost-control strategies that focus on providers.

---

<sup>4</sup>One classic study of this looked at doctors’ recommendations for terminal cancer patients. General internists recommended within-home and hospice care more often, whereas cancer specialists (oncologists) recommended hospitalization and intensive care more often (Kissick et al., 1984).

## Payment Strategies for Doctors

Perhaps the most traditional and widely studied method of altering provider behavior in managed care involves compensation of the doctor, who (by training and law) sits in the position of prescribing and managing all of the patient's care. The key issues in this matter have already been discussed in Chapter 7.<sup>5</sup>

**Salary Versus Fee-for-Service** The idea of paying doctors a fixed annual salary (instead of paying on a fee-for-service basis) arose most prominently in the "classic" Health Maintenance Organizations (such as Kaiser Permanente), where all of the medical treatment for enrolled patients is provided by a closed panel of physicians who all receive fixed salaries. This HMO literature provides a good source of information about the effects of physician compensation, although we cannot treat it as definitive, since the HMOs do other things besides change physician compensation. For example, they often tout their use of preventive services as a way of reducing health care costs, and emphasize the value of continuity of care provided through a single integrated organization. That having been said, we can turn to the HMO versus FFS literature for guidance about the effects of salary-based compensation.

The concept originated with the Kaiser Permanente Corporation in the 1940s in Portland, Honolulu, and Los Angeles, as a way to provide health care benefits to corporation employees. The insurance plan hires the doctors (in the pure form, on a straight salary) or contracts with a specific group of doctors to provide care, and either builds its own hospital or contracts for the services of a hospital within the community. The original Kaiser Permanente group plans also all had their own hospitals associated with them. The key feature of such programs is that the providers of care (doctors, hospitals, etc.) have no financial incentive to provide extra care to the patients enrolled with the insurance plan, because the insurance plan, the hospital, and the doctors all share common financial interests.

This form of payment is often called "capitation payment," since the payment for caring for patients comes on a "per person" or "per head" basis (hence the name "capitation"). Capitation payment is really the essential part of a pure HMO, because of the financial incentives contained therein. In standard fee-for-service insurance plans, providers always do better financially by providing more care to patients. The insurance company is "at risk"

---

<sup>5</sup>The assiduous student should here reread the summary of the very important study by Hickson, Altmeier, and Perrin (1987) discussed in the final section of Chapter 7. Or better, find the original article from *Pediatrics* in a medical library and read it directly.

in a large way, the providers are not “at risk” at all, and the patients are “at risk” only to the extent of their deductibles, coinsurance, etc. In this type of world, the two parties closest to the decisions about using medical care—providers (notably, doctors) and patients—have no financial interest in conserving scarce resources. The “third party”—the insurance company—has the financial risk, but the weakest ability to do anything about expenditure control. In fact, the only interest the insurance company has in controlling costs is the extent to which mechanisms can be devised to control costs that help make the insurer’s premium costs to insurance buyers more attractive, and the value of accomplishing that is blunted by the tax subsidy to health insurance.

By contrast, in the pure capitation model, doctors and hospitals are paid either on a flat salary (in the case of doctors) on the basis of agreements about how many “covered lives” they will take care of. The providers agree to provide all “necessary” care for the patients in exchange for a specific amount of payment per patient per year.

This arrangement greatly changes the incentives for doctors and hospitals. They no longer make money on volume business. They make money by finding ways to prevent the use of care, and they have strong financial incentives to do just that. They have legal and market incentives to provide “adequate” amounts of care, which can help ensure that patients are not undertreated, but they most certainly do have incentives to conserve on resources.

In many of the “pure HMO” plans, the doctors operate as a separate legal entity (commonly, a large multispecialty group practice) that enters a contract with the HMO-insurance company to provide care for the enrolled patients, agreeing to use the HMO’s hospital exclusively. The patients pay a fixed premium for the year, for which all care is provided (sometimes, with small copayments for physician visits). The physician group shares in any extra money remaining at the end of the year. Thus, the group has strong financial incentives to avoid expending resources unnecessarily. They can avoid some of the coordination problems of large groups (see Chapter 6) through rules and supervision, and—importantly—by being able to fire doctors who do not cooperate.

These plans create a supply-side incentive to restrict the amount of medical care rendered. Not surprisingly, the major cost savings generated by this form of care come from reduced hospitalization of enrolled patients, the one very costly activity over which physicians have almost exclusive control. Extensive studies of HMOs (see Luft, 1981, for a summary) have repeatedly shown their reduced propensity to use the hospital, compared with fee-for-service arrangements with extensive insurance coverage. (Note that the relevant comparison is the rate at which medical resources are used in a full-

coverage insurance plan, since HMO patients have no financial risk and an out-of-pocket price at or very near zero.)

These nonexperimental findings were confirmed in the RAND Health Insurance Study (HIS), which enrolled some patients (randomly assigned) into an HMO in Seattle. These HMO patients were compared with other HIS enrollees with full-coverage insurance, and also with a set of patients previously enrolled within the HMO. The first comparison (with full-coverage fee-for-service patients) tests the effect of the HMO-incentive system, holding constant the coverage from the patient's point of view (row 1 versus row 3 of Table 11.1). The second comparison (with previously enrolled HMO patients) tests a separate issue—namely, whether the HMOs have been getting an unusually favorable or unfavorable set of patients in their usual enrollment (row 1 versus row 2). This is the “self-selection” problem previously discussed. It has loomed important in the long debate about HMOs, because the previous comparison studies, like those undertaken by Luft and others, always contained the risk that the results achieved by the HMO were due to a favorable patient mix.

Two important results emerge from this study. First (in parallel with the previous nonexperimental results of Luft and others), the HMO plan used considerably less resources than the comparable  $C = 0$  fee-for-service plan in the same city (Seattle). (The importance of using the same city as a comparison is heightened by the cross-regional variations in hospitalization shown in the work of Wennberg and others—see Chapter 3.)

Second, the RAND HIS studied the health outcomes of people enrolled in the experimental HMO, and compared those with people on the full-coverage fee-for-service plan. As reported by Sloss et al. (1987), at least on a battery of 20 physiological measures of health status that encompass every major organ system (vision, musculoskeletal, digestive, etc.), the HMO enrollees fared at least as well as the corresponding fee-for-service enrollees with full coverage ( $C = 0$ ). Thus, at least in this HMO, the reduced use of resources did not lead to reduced health outcomes for the enrollees.

**TABLE 11.1** HMO HOSPITAL USE AND TOTAL COSTS IN THE RAND HIS

	Percent Using Hospital	Percent Ambulatory Any Care	Total Cost (\$)
HMO experiment group	7.1	87	439
HMO controls <sup>a</sup>	6.4	91	469
Fee for service with $C = 0$	11.1	85	609

<sup>a</sup>Previously enrolled in HMO, not as part of RAND HIS.

**The Proper Balance of Incentives** Economists always worry about incentives, and when one looks at the problems associated with FFS payment (incentives to overprescribe treatment) and those of a flat salary-based compensation (shirking and incentives to withhold treatment if savings are shared by the doctors), then it comes as no surprise that people have begun to think about a blend of payment schemes that provides the best possible outcome. Ma and McGuire (1997) did just this, modeling the physician's behavior with a compensation system that provided both a flat salary (of smaller size, of course, than if the total compensation were salary) and a fee-for-service (again, smaller in size than if the FFS were the only payment). The best blend between the two depends on the strength of the adverse effects generated by each payment system (all of which can, in concept, be measured). If demand inducement effects are relatively large, then the salary component rises and the FFS component shrinks. If shirking and withholding matter more, then the reverse holds (Ma and McGuire, 1997).

The proper balancing between salary and FFS depends in part on how hard doctors work under different compensation arrangements. As reviewed in Chapter 6 (see discussion in the section entitled "The Size of the Firm—Group Practice of Medicine") an important analysis of physician productivity under various compensation arrangements (Gaynor and Gertler, 1995) showed that as physicians are paid more and more on salary (versus by the treatment), their work effort shrinks, and the effect gets worse in larger firms as the ability of managers to monitor the work effort of the physicians deteriorates. In small groups, salary-paid physicians worked about two-thirds as much as doctors paid in an FFS model. In the largest groups, the work effort by salaried doctors dropped to 40 percent of that found in comparably sized FFS groups. This effect is precisely captured in the Ma and McGuire approach.

Implementing the Ma and McGuire approach in a real-world setting involves conjectures (data to support a carefully designed plan simply do not exist) on the part of insurance plan managers about the strength of the effects of demand inducement and shirking. Gaynor and Gertler's work provides reasonably strong evidence on the relationship to work effort and compensation. However, detailed understanding of the importance of demand inducement is missing. Most likely, the extent of demand inducement will vary considerably from setting to setting, perhaps by treatment, perhaps by the educational level of the consumer, perhaps by the degree of competition and information in the market. Such a level of detail about demand inducement simply does not exist now, so the balancing act implied by Ma and McGuire must rest more on conjecture than on firm analysis at this point.

**Holdbacks** A separate but common method for controlling costs in MCOs occurs when the insurance plan "holds back" a part of the payments due to

physicians (say, 15 percent) until the end of the year, to see if total treatment costs within the plan have come within targets. If the plan comes in at or under target costs, the holdback is dispersed to all the doctors (in proportion to the payments that they have already received). If the MCO has unusually large costs, then the “holdback” is kept to help the MCO cover total costs. The logic is that since the doctors control the use of medical care, they should have an incentive to control the costs by sharing in any losses that the MCO might encounter.

This approach certainly helps the MCO deal with cost overruns when they occur, but its ability to actually lower costs remains in doubt. The central problem from an analytic point of view is that the doctors within the MCO have little economic incentive to engage in lower cost medical treatments because of the holdback, and indeed, if they do “cooperate” with the MCO, they are likely to suffer relative to their peers. The concept of a global holdback is best analyzed using noncooperative game theory, and this fits into the class of games known as a Prisoner’s Dilemma. In such games, cooperation (here, to save money) almost always falls apart, and hence can have no effects on total medical care use. The problem is directly analogous to the global spending cap problem analyzed in Box 12.1 which discusses a similar feature in Medicare; you may wish to jump ahead and read that material now.

## Payment Strategies for Hospitals and Related Organizations

Hospitals have distinct characteristics about their services that make them eligible for a particular form of payment that lies in between FFS and capitation—the so-called payment by “the case.” Since a hospital “event” has clearly defined boundaries—the admission and discharge of the patient—insurance plans can pay hospitals for single hospitalization events in a lump-sum. This is not the same as FFS payment (in which the hospital bills and is paid for each distinctive event that happens to each patient) nor capitation (in which the hospital is at risk not only for the cost of each hospitalization but also for the number of such events during the year).

The quintessential “case-based” payment system emerged from the federal government’s desires to control hospital costs under the Medicare program, creating a “prospective payment system” (PPS) involving nearly 500 “diagnostically related groups” (DRGs).<sup>6</sup> This system of payment has been adopted by a number of private insurance plans as well. Discussion of the details of the DRG system occurs in Chapter 12 (Medicare), where the system arose, and thus will not be discussed further here.

---

<sup>6</sup>Yet another pair of TLAs for the student of MCOs to learn!

## Provider Selection

Many MCOs rely on careful selection of providers to control costs of care. The benefit comes from two sources. First, if there are actually intrinsic differences between the costs of various providers, then finding the low-cost providers makes good sense for an MCO. Second, whatever the actual cost structure of providers, if they have been pricing (in the “old” market structure) above their actual costs of treatment, then negotiating the price downward has salutary effects for the MCO.

Perhaps the most important evidence on the efficacy of PPO plans is their rapid growth in the market (from 28 to 41 percent of the market in the short span between 1996 and 1999, for example). Price-conscious employers and employees choosing lower cost health insurance would find PPOs increasingly attractive (as these data show) if they developed a relatively low premium structure through their primary cost-containment activity—bargaining with providers. Glied (2000) surveys the literature on managed care and finds no unambiguous result with respect to PPOs compared with alternative (FFS) plans.

One reason why PPOs might not achieve their full possible benefit is that they seldom actually use available information to select among providers carefully. One study of the primary care physicians in Rochester, NY (encompassing almost all doctors in the region) shows large and systematic differences in the total costs of treating patients among doctors in a very “open” IPA panel of physicians. The top 10 percent of doctors had charges per patient per year that were twice those of the lowest 10 percent. Further, these differences controlled carefully for the underlying illness conditions of patients, and they replicated well from year to year. Thus, the authors concluded that doctors in this study have identifiable and reproducible practice styles that allow their grouping into relatively lower and relatively higher cost providers (Phelps, Mooney, Mushlin et al., 1994). Yet neither of the competing IPA insurance plans in the community opted to limit their panel of physicians to the relatively lower cost providers. Their choices invariably embraced the philosophy of bringing into the “panel” of allowed providers all those who wished to participate, apparently on the hypothesis that more doctors meant more patients and hence a larger market share. (Both plans are not for profit, and may prefer market share as a goal rather than net plan profits.)

Do MCOs select doctors on the basis of quality of care as well as cost of treatment? A new study on this issue uses New York State’s periodic report on the quality of care (measured by risk-adjusted mortality rates, or RAMR) of surgeons and hospitals for coronary bypass artery grafts (CABG). Mukamel and Mushlin (1998) studied the probability that an MCO would contract with a cardiac surgeon, and found that MCOs preferred to contract with

doctors with better RAMR and they also tended to avoid low-volume providers (holding RAMR constant). The preference for high-quality “outliers” was greater in the New York City metropolitan area where (presumably) more choices were available to the MCOs. This one study, at least, suggests that quality of care also affects provider choice (when it can be measured) in addition to cost of treatment.

In this light, a rather remarkable change occurred in a large insurance carrier in California in 2001; Blue Cross of California, first among its peers, abandoned rewards linked to cost control in its HMO MCO plans, and structured a new set of bonus payments to physicians directly tied to patient satisfaction with the plan.<sup>7</sup> The new system, applying to the 1.5 million HMO enrollees among the 5.5 million overall enrollees of Blue Cross in California, will offer incentive payments of up to 10 percent bonuses for the highest satisfaction scores, as measured by random surveys of patients enrolled in the plan.

MCOs also select hospitals as well as individual doctors. In a study utilizing California data from 1983 to 1997, Zwanziger, Melnick, and Bamezai (2000) studied the consequences of California introducing selective contracting both through the state’s Medicaid program (“MediCal”) and by introducing legislation allowing selective contracting in private insurance markets. They developed measures of competition among hospitals and measured the rate of cost growth and how it changed due to selective contracting entering the market. They found a strong negative relationship between the degree of competition (the market concentration) and the rate of cost growth after selective contracting emerged—more competitive markets had less cost growth.

Still another form of “selection” occurs when MCOs limit the choices of consumers (and their doctors) regarding prescription drugs. Many MCOs now have a “formulary” of drugs that receive the highest coverage (typically generic drugs when available) and one or two more expensive (to the patient) tiers of drugs where the patient’s out-of-pocket costs can be considerably higher. Chapter 15 discusses insurance coverage of prescription drugs in more detail.

## Prices and Fee Schedules

The essential feature of PPOs (as noted in Box 11.1) is to bargain with providers about price. The circularity of success in such programs is compelling, particularly in markets in which providers have had the opportunity to price above marginal production costs. The cycle looks like this: A PPO enrolls a series of people with the promise of a low premium (based on the

---

<sup>7</sup>Reported on the AP wire service, July 10, 2001.

expectation of low prices from providers). It negotiates with providers (doctors, hospitals, nursing homes, etc.) by promising to bring them a high volume of patients (those they have enrolled) in exchange for a favorable price.

To the extent that a PPO is successful in recruiting more patients, it enhances its ability to convince more providers to join the organization, and it gives it more leverage in bargaining for a better price. Similarly, as it adds more providers to its panel, it becomes more attractive to more consumers (particularly those who do not wish to change providers as they change insurance plans). Hence, successful growth in either the patient or provider market can enhance the ability of a PPO to grow in the other market. The only limiting factor eventually may be the proportion of patients who find managed care objectionable and are willing to pay more for more generous alternatives. (Recall that PPOs usually come with other restrictions on utilization as well in the traditional “managed care” environment.)

If setting prices did not affect providers’ behavior, then the only “achievement” coming from price negotiations between MCOs and providers would be to lower the overall premium, since the plan would pay less for each unit of service provided, and hence have lower benefit payments and thus can charge lower premiums for the insurance coverage. But price negotiations do more than that—they affect the quality of service that will be provided, and they can alter the mix of services provided by altering relative prices of different procedures.

On the question of quality of service, we can understand readily the relationship between quality and price in many standard consumer purchases, ranging from restaurant meals to guitars to automobiles and homes. We should expect nothing different in health care. Chapter 9 details mechanisms by which not-for-profit hospitals will select output, price, and quality when facing consumers with standard insurance plans. MCOs alter that arrangement by bargaining on price with the hospital but leaving the hospital able to alter quality and quantity within the constraints provided by the price bargaining. A review of Figures 9.1 through 9.3 (with the altered perspective of a hospital negotiating with an MCO) would be a good idea at this point.

The same idea will hold also for physician firms and other similar providers; for example, if the PPO bargains for a lower price, one ready response for the physician firm would be to shorten the time of each patient visit. Other characteristics of quality will also change in response to a lower fee from the insurance plan, including (in the longer run) the upgrading of medical equipment in the office, the inventory of available supplies, the quality of furniture in the waiting room, and, indeed, even the number of magazines (and the degree to which they are current) in the waiting room.

PPOs must also attend to the question of *relative* prices among services (unless they are a POS model where providers are paid on a capitation basis).

Rogerson (1994) develops a formal mathematical model of the problem in a structure almost identical to that posed in the discussion of quality, quantity, and hospital preferences in Chapter 9. He concludes that when a payer with significant market power (a large PPO or Medicare, as the next chapter discusses) sets prices for individual services, the not-for-profit hospital will respond by adjusting quality in a specific way: For those goods for which demand is not quality sensitive (one might think perhaps of urgent care where the consumer had little time to contemplate alternative choices), the hospital will lower quality and hence cost (relative to the price offered by the PPO or Medicare) and capture a greater “margin” to help support other hospital goals. In those services for which demand is very sensitive to quality (one might think of highly predictable events such as childbirth) the hospital will maintain high quality.

## Intervention in Specific Treatment Choices

Another class of interventions occurs at the time of specific medical decisions. In general, the interventions discussed previously for patients (copayments, gatekeeper, etc.) and providers (selection of providers through profiling, holdbacks, price setting) all take place “in the large.” Many MCOs also use two types of interventions for decisions “in the small” to help control medical costs.

**Prior Authorization** Particularly for expensive interventions (surgery, hospitalization, major diagnostic tests) many MCOs require that the physician receive prior authorization from the insurance plan. Often approval is routinely granted, but sometimes with limits attached. For example, a hospitalization may be authorized for a length of stay of  $X$  days, at which point if the doctor wishes to extend the stay, further authorization is required. Although often granted by the MCO, the demand for the authorization sometimes tips the balance in the doctor’s decision making about discharging the patient to home or not. The “hassle” of getting the authorization offsets the gains from extending the stay, hence sometimes leading to reduced length of stay. These protocols can also help MCOs establish “norms” for treatment that help control costs.

**Denial of Payment** Another powerful device for MCOs to control costs is denial of payment, probably the most controversial of the various cost-control devices. Unlike prior authorization mechanisms (about which the patient and provider know in advance of treatment), denial of payment takes place *after* treatment has been rendered. In that case, either the doctor (and

hospital) or the patient (or some mix of them) has to “eat” the costs of treatment. The spectre of denial of payment leads physicians to be very cautious about venturing into treatment protocols where the risk of denial of payment looms as a possibility. Because they materialize after treatment has already been rendered, the conflicts they cause between MCOs and the provider and patient can be considerable.

There exists, of course, a fuzzy boundary between denial of payment and denial of treatment through prior authorization. In many cases, refusal of an MCO to authorize a treatment gets described as a denial of payment. Sometimes even the process of making a decision about whether to deny authorization or not can lead to major disputes. One of the landmark cases in the area of patients suing MCOs over denial of treatment centered on the time lapse between a request for treatment and the actual decision (see the discussion of MCO liability in Chapter 13 for details).

## **INTEGRATED DELIVERY: THE NEXT PHASE OF MANAGED CARE?**

The next phase of changes arriving on the U.S. health care scene are intimately related to, and in at least partly in response to, changes in ways insurance companies pay doctors, hospitals, nursing homes, visiting nurses, pharmacies, therapists, and the like: All of these types of providers, which once stood independently of each other with vigorous force, are now merging into a new form of health care provider organization called integrated delivery systems (IDS). This shift in organization represents a change in the U.S. health care system that could be extended further in its consequences than even the formation of Medicare and Medicaid and the subsequent revisions in those government programs (as subsequent chapters will discuss).

To begin this discussion, it is useful to recall first how the providers of various health care services relate to one another in the traditional U.S. market. Hospitals stand organizationally at the top of the pyramid, at least in terms of expense per person per year, and in terms of where the most sophisticated medical care takes place. In most cases, physicians in private practice, working on a fee-for-service basis, treat patients, and (when required by their medical judgement) refer those patients to specialists for further diagnostic analysis or specialized treatment. These specialists would in turn practice their speciality independently, on a fee-for-service basis, either in solo practice or in small groups of related specialists. Either the initial (primary care) provider or the specialist might admit the patient to the hospital for more

complicated treatment or surgery. The doctors also might prescribe additional treatment for the patients—pharmaceuticals (which would be purchased from an independent drug store or mail order pharmacist) or physical therapy (which might be delivered by somebody working for a specialist or by somebody in private fee-for-service practice)—or decide the patient should be admitted to a nursing home, in which case the patient (or family) would make arrangements with an independent nursing home for admitting the patient. The key words in all of this description repeat themselves almost endlessly: independent and fee-for-service.

In these settings, there is little to no automatic exchange of information. A history and physical examination obtained by the primary care doctor will likely be repeated by a specialist seeing the same patient (and similarly laboratory and imaging studies), and possibly again by a resident doctor in the hospital to which the patient might be admitted. Copies of the medical records might be sent from one doctor to another, but the “outpatient” records of the various doctors would almost never find their way into the hospital record, and the insurer would have no cognizance of any of this except as recorded in the patient’s bill and perhaps an abstract of the patient’s care upon discharge from the hospital. The pharmacist prescribing the drug would commonly know little or nothing of the diagnosis, let alone other illnesses the patient might have or other prescription drugs the patient might be taking.

All of this is changing rapidly, through a process of vertical integration. In the usual economics jargon, “horizontal integration” implies the merger of two organizations that produce or sell the same thing. Thus, for example, a merger of GM and Ford Motor Company would constitute a “horizontal integration” (although one that would be impossible under current U.S. legal standards for mergers). “Vertical integration,” by contrast, implies the merger of organizations involved in different steps in the chain of production and sales. Thus, for example, the merger of Ford Motor Company with Firestone Tires would constitute a vertical integration, since GM now uses tires bought from other corporations to put on their new vehicles. A further layer of vertical integration might occur if (to continue the hypothetical example) Firestone Tires bought or merged with a company producing raw rubber.

Horizontal integration normally occurs either to take advantages of economies of scale or to create market power, by eliminating rivalry between two (or more) former competitors who merge. For this reason, the U.S. Department of Justice and the Federal Trade Commission look carefully at mergers of any sort (including those between not-for-profit hospitals) to determine if current standards of competition would be violated by the mergers. Vertical mergers take place for other reasons. Sometimes they increase the control over the supply of a crucial input. In other occasions, vertical

integration has occurred to bypass a monopoly in the input market. (Continuing our hypothetical example, if Firestone Tires faced a monopoly in the purchase of raw rubber, they might try to develop their own supply of rubber, either by purchase or developing new rubber plantations, as a way of bypassing the monopoly.)

In the case of health care delivery, vertical integration can serve several purposes. Key among these, however, is likely to be the development of more efficient and effective treatment of patients through a carefully designed information system that spans the entire range of organizations and providers who treat patients. The second outcome, already occurring in many places, is the seamless transfer of patients from one setting to another to minimize costs of treatment. (Cross-organizational availability of medical records on each patient assists in this process, of course.)

The most important outcome is that vertically integrated health care delivery systems can more readily adapt to changing payment mechanisms, particularly those that use capitation payment (one annual fee per patient in exchange for all necessary medical treatment, wherever rendered). Thus, changes in the private sector that lead to more common use of capitation as a payment system will likely encourage more development of integrated delivery systems, although there are many apparent advantages to integrated delivery with or without capitated payments.

IDS means many things to many people. At one extreme, it represents a rather loose affiliation of various groups of providers, who contractually agree to send patients to each other and to cooperate in the exchange of information about those patients. For example, an IDS might arise with an insurance company at the center, contracting with a tertiary care hospital (for the most serious illnesses), several less-intense community hospitals, and perhaps several small hospitals in surrounding rural areas. At the other extreme, the IDS would consist of a closely held group of provider physicians, possibly all under salary arrangements, working for a single group. The classic HMO fits this model closely.

Any modern IDS would probably also have contracts or agreements with several nursing home and extended care facilities, some of which emphasize posthospital rehabilitation and some of which (with lower costs per day of care) emphasize mostly custodial care of patients no longer able to care for themselves. The arrangements would also commonly include a home health service, visiting nurse service, and perhaps even a group of pharmacies.

The IDS will also invariably include groups of physicians within the system, both primary care and specialty care. Indeed, the development of a "primary care network" stands as the central task of any would-be IDS. This network must include doctors affiliated with the hospitals in the IDS, so they

can treat patients within those hospitals. Getting the right number of primary care doctors, and a sufficient geographic dispersion of those doctors to bring in patients from all around the region, is a key to success here. There will also be arrangements or contracts with specialists, to whom the primary care doctors refer patients for more complicated care. Typically, the more “loose” IDS programs will encourage, but not require, referral to the affiliated specialists. In more tightly controlled IDS organizations, such “within-plan” referrals are usually mandatory, unless there simply is no specialist of the requisite type in the portfolio of IDS specialists. If the IDS insurance plan includes dental services, the plan would likely also have agreements with a series of dentists to deliver primary dental care.

Ultimately, IDS organizations seem likely to become a prominent way for providing health care delivery and the normal functions of insurance (risk spreading). IDS organizations have many of the advantages of both HMOs (and the incentives built into capitation systems) with fewer of their disadvantages. Pure “staff model” HMOs, by their very nature, have more limited geographic dispersion (and hence more difficult access for patients) than a standard fee-for-service market model, with widely dispersed doctor offices and hospitals. Traditional HMOs tend to operate out of a few (sometimes only one) fixed sites. Thus the IDS has more geographic convenience. Second, because they are organizations built from affiliations of other organizations, it is probably easier to expand and contract the scale of an IDS organization than a pure staff model HMO. This inherent flexibility also has advantages in today’s health care market. Third, because an IDS can encompass a wide array of primary care providers, they can be designed to have much less impact on established doctor-patient relationships, particularly if the IDS can become large relative to its geographic area. It is easy to conceive of an IDS organization serving half a million or more people, so in all but the largest cities in the United States, a small handful of IDS organizations could easily serve almost the entire metropolitan area of most cities. As IDS organizations grow in scope and size, it may well be that many individuals within the United States have the opportunity to join an IDS without having to change their primary care provider, which should (in turn) enhance their attractiveness to consumers.

## LONG-RUN ISSUES

As one might expect, when a major change such as managed care sweeps across a market, a number of subsequent changes will reverberate through the market, particularly in “input markets” that may require longer time pe-

riods for adjustment. Hospitals experience lower demand for their services, and either convert to other uses such as long-term care or actually close (see Chapter 9). The new insurance organizations will also experience growing pains as they experiment with different cost-control strategies to find out which are more effective, which annoy provider and/or patients too much, and which cost more to implement than they return in value.

As managed care organizations grow and solidify their hold on the insurance market, we can look ahead to various long-run consequences in health care markets. These include effects on physician location and retirement, the behavior of insurance markets themselves, and possible changes arising from both litigation and new legislation that could change the legal background in which MCOs operate. This section discusses these issues in more detail.

### **Physician Incomes, Geographic Distribution, Mobility, and Retirement**

The market for physician services has been particularly stressed by the arrival of managed care. In part because of general incentives and in part because of the specific choices of MCOs to emphasize primary care (versus specialty care), the relative financial status of primary and specialty providers has been turned somewhat topsy turvey, particularly in markets with large MCO penetration. This in turn should ultimately affect specialty choice for physicians in training, retirement decisions, and the like. We can observe some of this happening already.

One way to look at the long-run outcomes is to compare the staffing patterns of traditional group-practice (salary, staff model) HMOs with the U.S. health care system as a whole. Although the pure staff model HMO is far from the favorite of consumers in the U.S. market, it sets a benchmark of personnel utilization that is worth understanding. Table 11.2 shows the staffing patterns of various HMOs and compares it with the U.S. physician supply in 1992. (The U.S. mix is quite similar now since the stock of active physicians changes relatively slowly even if the flow of new graduates were to change abruptly in its mix, which has not occurred.)

These figures approximate the equilibrium demand for physician services in a strongly managed care environment. They show an important decline in the utilization of specialists of all types. If we extrapolate these data to the United States (a risky enterprise, since these are staff model HMO data, not more popular PPO and IPA staffing arrangements), as the demand for specialists declined, those with multiple \$100,000 earnings would find their annual incomes plummeting and perhaps even reach the stage of unemployment. We must view these data with some caution since the staffing

**TABLE 11.2** COMPARISON OF HMO STAFFING PATTERNS WITH NATIONAL MD SUPPLY

	HMO Staffing <sup>a</sup>				U.S./HMO Ratio <sup>f</sup>
	Seven Kaiser Plans <sup>b</sup>	Kaiser Portland <sup>c</sup>	GHC Seattle <sup>d</sup>	1992 U.S. Supply <sup>e</sup>	
Total	111.2	136.8	121.7	180.1	1.5
Primary care	53.6	56.3	57.1	65.7	1.2
Medical subspecialties	11.8	14.7	11.5	17.8	1.4
Surgical subspecialties	27.8	33.0	34.3	43.8	1.4
Hospital based	9.7	16.5	—	22.0	1.7
Other	10.2	12.3	—	20.3	1.8

<sup>a</sup>All figures represent full-time equivalent physicians per 100,000 population. Totals and subtotals may include some subspecialties not listed.

<sup>b</sup>1983 data from four to seven Kaiser Permanente sites (depending on specialty).

<sup>c</sup>1992 data from Kaiser Permanente Portland HMO, staff model HMO with 378,000 members. Primary care includes "urgent care" physicians.

<sup>d</sup>1992 data from Group Health Cooperative of Puget Sound, staff model HMO with 475,000 members.

<sup>e</sup>Nonfederal patient care M.D. and D.O. physicians, excluding residents and fellows in training.

<sup>f</sup>Ratio of the fourth column to unweighted average of HMO data from the first three (or two) columns.

Source: Weiner (1994, p. 224).

mixes reported in Table 11.2 are over a decade old, but the general conclusion about physician supply probably stands, even if future patterns of physician demand do not perfectly match those appearing in this table.

Medicare's payment schemes instituted in 1992 will enhance these trends, since they lower the earnings of specialists emphasizing procedures (surgery, etc.) and increase the earnings of doctors emphasizing cognitive skills (internists, family practice doctors, etc.). Chapter 12 discusses the Medicare payment reform known as Resource-Based Relative Value System. However, here we can study separately the effect of managed care growth on physician earnings.

In one study, for example, Simon, Dranove, and White (1998) used sophisticated statistical models to assess the changes in physician earnings using data collected by the American Medical Association (AMA) on physician practices (a 1 percent sample using telephone interviews with a 60–70 percent response rate annually). They conclude that managed care penetration into a market enhances *primary care* physician earnings significantly, and in parallel significantly cuts into the earnings of a group of hospital-based spe-

cialists (radiologists, anesthesiologists, and pathologists, or RAPs) by comparable amounts. Specifically, each 1 percentage point increase in MCO market share in the insurance market adds \$2263 to the annual earnings of primary care doctors in their study. Since the markets saw MCO market share rise by over 15 percentage points in the time period they studied (1985–1993), they estimate that these primary care doctors experienced an average increase of about \$34,000 per year due to MCO growth, a 50 percent growth from their base income average of \$69,000. By contrast, RAPs experienced a \$1993 decline per percentage point increase in MCO market share, for an overall “hit” of about \$30,000 a year.

Economic theory (and the evidence we saw in Chapter 7 in the discussion surrounding Table 7.1) tells us that physicians tend to locate in ways that equalize earning opportunities across regions. Greater MCO penetration into markets did affect the mobility of some physicians, but not all: Younger physicians were more likely to move in the face of an influx of MCOs than older doctors (as you would expect; they have more time to “amortize” the costs of the move than older physicians). (Escarce, Polsky, Wozniak, Pauly, and Kletke, 1998.) Specialists were more likely to move than generalists (in accord with the notion that generalists, if anything, experience positive income outcomes from MCO growth). Other types of physicians seemed unaffected in location decisions by MCO growth in the market (Polsky, Kletke, Wozniak, and Escarce, 2000). As one might expect, those physicians who moved generally headed toward markets with the same or lower MCO penetration, with few moving to a market more intensively populated by MCOs.

The same set of authors report that large MCO penetration into a market caused many physicians to retire early, with results varying by specialty (Kletke, Polsky, Wozniak, and Escarce, 2000). Not unexpectedly, older physicians were most likely to move to retirement, and among that group, both generalists and medical/surgical specialists opted for earlier retirement. As an example of the magnitude of this effect, they compared the retirement probabilities of generalists in a high-HMO penetration market (45%) with a low-penetration market (5%) and found that retirement probabilities were 13 percent higher in the high-penetration market than in the low-penetration market. Medical/surgical specialists’ retirements differed by 17 percent in the same comparison.

## Market Segmentation

In Chapter 10 we explored the issues originally posed by Rothschild and Stiglitz (1976) regarding the possibility that health insurance markets might “separate” into submarkets for low- and high-risk consumers. The restrictions on access to care that are the cornerstone of cost management in MCOs

may do exactly that. To begin this discussion, return to Chapter 10 and the discussion surrounding Figures 10.5 and 10.6. The key idea there (reworked slightly from Rothschild and Stiglitz' original presentation) is that insurance companies have difficulty in identifying relatively healthy and relatively sick individuals. One mechanism they have to accomplish this (and hence regain the ability to charge a premium to the sickly individuals that will reflect their higher health care use) is to limit the degree of insurance coverage to the point at which sickly people find the more-generous plan desirable, even at a higher premium. The cost of this arrangement (in social welfare terms) is that the relatively healthy people in the population find their coverage opportunities limited (at prices reflecting their generally lower cost health outcomes).

Could the restrictions on managed care plans serve the same role? The essential feature of all MCOs is to restrict (or deny) access to care in myriad ways (second opinions, gatekeeper arrangements, denial of payment to providers in ambiguous cases, etc.). These are precisely the sorts of restrictions that a very sick person would find particularly undesirable, and a relatively healthy person would find relatively less obnoxious (because they get sick less often). Does the selection of health plans by people follow this sort of model?

Glied (2000) reviewed two dozen published studies on the issue of selection in MCOs. The primary method of identifying "sickly" and "healthy" individuals in these studies was to look at prior-year medical spending as a measure of the propensity to get sick. Her summary of this literature states that "Overall, the results of selection studies suggest that managed care plans in the private sector tend to enjoy a 20–30% prior utilization advantage over conventional indemnity plans while Medicare plans enjoy a similar advantage over traditional Medicare."

These results suggest that the primary outcome of MCOs may not be to control medical cost so much as to create a viable mechanism by which insurance carriers can identify relatively healthy people, and attract them into an insurance plan that sickly people will find relatively unappealing. If so, then they provide a crisp mechanism for achieving the sort of market segmentation that was discussed in Chapter 10 using the example of "copayment" as the selection device.

To date, we have no strong randomized controlled trial data to show the effects of various MCO strategies on utilization and cost control, and, indeed, it seems unlikely that such studies will appear in the near future. The problems from conducting a "definitive" study, either using a prospective randomized controlled trial (RCT) format or careful econometric analysis, seem formidable, since the array of possible "management tools" is so large, and the combinations of them employed by MCOs will make it almost impossible to define the effects of any one of these strategies cleanly.

## Legal Considerations

Managed care, at least in its modern manifestations, has created several new legal issues that traditional insurance and health care mechanisms never raised. Early versions of what are now called HMOs—notably the Kaiser Permanente plans that arose before 1950 and the Ross-Loos group in the Los Angeles area—drew considerable fire from both organized medicine and (on some occasions) the legal establishment for what was then widely known as “the corporate practice of medicine.” The key issue in that era was organized medicine’s fight against doctors working on salary for a corporation. The modern manifestations of MCOs have widened the realm of legal complexity considerably.

**Antitrust and Related Issues** Even the creation of the most popular model of MCOs—the PPO—required special legislation. The essential feature of the PPO is selective contracting, and attempts to engage in such selection (before the California legal reform) ran afoul of antitrust regulation, the argument essentially being made that exclusion of a provider from the plan was an illegal boycott. The government of California swept all that aside in a dual series of reforms in 1982, first passing laws that enabled selective contracting (by providing immunity from other laws that might have hindered them) and at the same time establishing a major customer base for this new form of insurance plan—California’s Medicaid services. Thus began the PPO movement, which has subsequently swept the nation.

Even with that change, major legal considerations remain. One major antitrust hurdle remaining was the issue of whether physician-operated networks could organize a full MCO service to compete with corporate-run organizations. Under previous antitrust rulings and considerations, such activity would likely have been considered as collusion among the physicians to fix prices. The basic considerations here stem from the 1890 Sherman Antitrust Act, which outlaws (among other things) coordinated efforts to set market prices. Under the Sherman Act, just talking about setting prices constitutes a *per se* violation of the law, even if the discussions do not lead to a successful price-fixing arrangement. The Sherman Act also outlaws boycotts (an issue with selective contracting), market division, and tying of the purchase of one product to that of another (a potential issue in insurance plans that require the use of specific sources of medical care).

Two landmark cases in the U.S. Supreme Court held sway here. First, in *Goldfarb v. Virginia State Bar* (1975) the Court concluded that professions such as medicine (and, presumably, the law) were not immune from antitrust considerations. Next, in *Arizona v. Maricopa County Medical Society* (1982) the Court concluded that the Medical Society’s fee schedules—set through a “foundation” that contracted to provide medical services for health

insurers—constituted illegal price fixing. The Foundation enrolled over two-thirds of the doctors in Maricopa County (the county in which Phoenix resides), established *maximum* prices (rather than minimum prices, the usual choice of price-fixing conspiracies), yet the Court still ruled them in violation of the Sherman Act.<sup>8</sup> The Court ruled that the maximum fee schedules could “discourage entry into the market and may deter experimentation and new developments by individual entrepreneurs.” Thus *Maricopa* became the touchstone for later rulings by the Federal Trade Commission (FTC, which enforces the Sherman Act and related federal law), and put a considerable damper on the development of physician-owned or physician-led MCOs.

The organization of physician-directed PPOs was particularly onerous to the FTC because the very purpose of PPOs is to establish prices. Thus the market converged on corporate ownership of PPOs, and by AMA testimony, 93 percent of PPOs had corporate ownership (the remaining being hospital based). Organized medicine, led by the AMA, began to seek legislative relief, including even having provisions to allow doctor-owned PPOs included in a Medicare reform bill in 1996 that was vetoed by President Clinton.

Not until 1997 did the legal landscape change, not with legislation but rather in the published guidelines from the FTC and the Justice Department. Rather than using the strict *per se* violation test created in *Maricopa*, the new guidelines use a “rule of reason” test, the key issue being whether, on balance, the creation of physician-owned networks is procompetitive. Where these new guidelines will lead remains to be seen, but the legal landscape has changed, and may now allow significant physician-owned competition with the corporate owners of PPOs that now dominate the scene (Kuttner, 1997).

**Liability and Patients’ Bill of Rights** The remaining legal issue confronting MCOs is whether they can be sued for withholding payment for medical services. This issue—in the forefront of health care politics in the 2000 presidential election—remains unresolved at this writing. A more complete discussion of this issue appears in Chapter 13 (“Medical Malpractice”), after providing the context of the legal background of tort law and its applications in medical malpractice. Readers familiar with the basic concepts of tort law may wish to skip ahead and read the concluding section in Chapter 13 at this point, but others can wait until after the proper background of legal issues is established.

---

<sup>8</sup>The Court’s decision is not without economic logic. For example, it is possible for a series of independent operators to take a “maximum” price as a target to which all of their prices move. Announcing the maximum price facilitates the coordination.

## SUMMARY

Managed care organizations in their various flavors have emerged as a potent force in the U.S. health care market. The evolution should not come as unexpected. Traditional insurance plans help protect consumers from the financial risk imposed by illness, but in turn create incentives for overconsumption of medical care that ultimately drive up medical spending (both quantity and quality) beyond a desirable point. It should come as no surprise that in a market-driven economy such as the United States, inventive entrepreneurs create alternative organizations that deal with the risk and at the same time provide mechanisms to control the spending.

Numerous types of MCOs have emerged during the past two decades to compete with traditional insurance and with the “pure” staff-model HMO that was the only viable alternative to FFS insurance for decades. We now have an alphabet soup of IPAs, PPOs, and POS plans, and now IDS as various models. All seek to accomplish the same goals—protect the consumer against financial risk while at the same time controlling the incentives for overuse of medical care. The tools available to accomplish this include incentives placed on the patient (copayment, gatekeeper requirements, second opinion programs), on the provider (holdback mechanisms, provider selection), and on individual medical events (prior authorization, denial of payment).

Each of these mechanisms to control cost comes with some inconvenience to patients, providers, or both. How much inconvenience U.S. consumers of health care are willing to tolerate has yet to be determined, and will certainly change through time. (Avoidance of inconvenience is likely a luxury good, and will grow as U.S. per capita income grows.) In addition, legal challenges to some of the mechanisms used by MCOs to control costs may further change the landscape. We should expect continuing evolution of these plans for years to come.

## RELATED CHAPTERS IN *HANDBOOK OF HEALTH ECONOMICS*

Chapter 13, “Managed Care” by Sherry Glied

Chapter 27, “Antitrust and Competition in Health Care Markets” by Martin Gaynor and William B. Vogt

## PROBLEMS

1. Identify the key distinctions between a staff-model HMO and an IPA model HMO. Which of these models do you think will have a greater ability to reduce utilization, and by what mechanism(s) do you think this will occur?

2. Suppose you were advising a PPO about selecting doctors to join the “panel” of approved doctors. What characteristics of physicians might you seek out? (*Hint:* Think of the physician-specific variations from Chapter 3; think of malpractice propensities of some doctors; think of patient satisfaction with their care and how the doctors’ behavior could affect that.)
3. Would you expect enrollment in managed care to increase or decrease in the future, presuming that future trends include (a) increasing technology available to treat illnesses, which would of course bring with it (b) increasing cost of medical care, which would be accentuated by (c) increasing per capita income over time (as a general trend)?
4. Discuss the potential value of second-opinion programs in light of what you know (see Chapter 3 for a refresher) about the variability of physician practice patterns across regions. If a second-opinion “consultation” comes from a doctor in the same geographic region as the original opinion, is it more or less likely to agree with the first doctor, compared to having the second opinion drawn from a wholly separate region?
5. Gatekeeper models of MCOs require patients to see a primary care doctor before going to see a specialist. What two key economic principles would be involved in a successful gatekeeper program? (*Hint:* One of these involves “technical” matters, the other involves incentives to providers.)
6. If managed care organizations continue to grow in market share (compared with traditional fee-for-service insurance), what do you think the consequences will be for physician incomes (in general)?
7. If managed care organizations continue to grow in market share, what do you think the consequences will be for specialty mix of physicians?