Organizing an HMO by Contract: Some Transaction Cost Considerations

Thomas Palay
University of Wisconsin Law School, tmpalay@wisc.edu

Follow this and additional works at: https://digitalcommons.unl.edu/nlr

Recommended Citation
Available at: https://digitalcommons.unl.edu/nlr/vol65/iss4/4

This Article is brought to you for free and open access by the Law, College of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Nebraska Law Review by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Organizing an HMO by Contract: Some Transaction Cost Considerations

TABLE OF CONTENTS

I. Introduction .................................................. 728
II. General Purpose Governance Structures .................. 732
III. Transaction Cost Economics ............................... 735
IV. The Nature of the Transaction ............................ 738
V. The Required Governance Structure ....................... 742
VI. IPA Versus Group Practice Governance Structures ...... 746
VII. Conclusion .................................................. 747

INTRODUCTION

Responding to Derek Bok’s reservation about the extent of non-productive resources expended on and by lawyers,1 Ronald Gilson develops a novel view of attorneys as “transaction cost engineers.”2 He argues that in a world where transactions are costly to execute, an attorney adds value to economic enterprises by structuring them in a way that reduces transaction costs. In short, the lawyer “devise[s] efficient mechanisms which bridge the gap between capital asset pricing theory’s hypothetical world of perfect markets and the less-than-perfect reality of effecting transactions in this world.”3

* Professor of Law, University of Wisconsin-Madison. B.A., 1975, Tufts University; J.D., 1981, University of Pennsylvania; Ph.D., 1981, University of Pennsylvania. This Article and a companion study, Relational Contracting, Transaction Cost Economics, And The Governance of HMOs 59 TEMP. L.Q. ___ (1986), are part of an ongoing study of the organization of physician’s and other professional’s offices. This study has benefited from the generous support of the Graduate School Research Fund of the University of Wisconsin.

3. Id. at 255.
To illustrate his point, Gilson presents a model of a relatively discrete exchange, the acquisition of a company. Though the contracting is complex and often long-term, there are finite limits to the expected duration of the underlying relationship. In his framework, the lawyer adds value by devising contract terms that overcome problems related to the fact that parties have different expectations of a future none can foresee and that people cannot always be trusted to act in good faith.

Where the relationship between the contracting parties is no longer discrete, the parties will often require more than an “efficient” contract to govern their relationship. I argue that it will in fact require the development of private orderings. In this Article I extend Gilson's analysis to the continuing relations required to form a health maintenance organization (HMO) by contract, and I suggest why, in this context, transaction cost engineering demands more than clever drafting.

Traditionally, American medicine has been fee-for-service based; that is, the patient pays the individual physician for services rendered. If the patient carries health insurance he pays periodic premiums to the insurer who then either reimburses the insured or pays the physician for the services rendered. The insurance company and the care provider are independent entities brought together by the need to provide services to the same patient. By contrast, an HMO, in return for a set fee per time period, actually provides a broad spectrum of medical services directly to the patients. The fee prepaid by the patient is fixed, irrespective of actual services used, thereby providing health care providers an incentive to reduce costs. Thus the HMO is an insurance company that, unlike the conventional insurer that reimburses a patient for his medical expenses, provides the actual medical services.

This Article is concerned with the organization of the medical services provided by the HMO and how the lawyer can help the parties to understand the complexity of the contracting. In particular I introduce the reader to transaction cost economics and illustrate how an attorney can use it to assess the viability of using contract as the organizational structure for the delivery of medical services to an HMO.

This Article begins with the assumption that the decision to form an HMO has already been made. In a world of increasing complexity, information growth, and technological change, no individual physician is capable of providing the full range of services potentially required by a patient. To offer comprehensive service, the HMO must integrate both horizontally and vertically the skills of a wide range of physi-

---

5. Horizontal relationships involve doctors who share similar skills, training and expertise. For instance, a pediatrician might want to be associated with several
There are two principal methods of integrating the required physician skills. The more traditional method is to merge the physician services into a single firm. Both the group practice HMO and staff model HMO use this approach. In both instances physicians are combined financially, administratively, and physically into one vertically and horizontally integrated medical practice (firm); in most instances the doctors share a single building, support staff, laboratory, and diagnostic service. As I use the terms, the group practice and the staff model differ from each other in the physicians' relationship to the firm. With the group practice model, the doctors maintain (or expect in the future to have) an ownership interest in the medical practice. Thus they have a proprietary claim on the stream of profits or losses flowing from the firm. With the staff model, the physicians are employees of the HMO and do not possess the same ownership interests.

Alternatively, the integration can be accomplished through contract. An independent practitioners association (IPA), for instance, uses this technique. An IPA is an administrative entity to which independently affiliated physicians with geographically dispersed practices have agreed to provide services. The IPA then sells the collective product of the doctors to an HMO. Thus the IPA is a set of contracts between unaffiliated physicians who want to integrate some portion of their total product without physically merging their various practices into a centralized firm. The contractual relationship between the parties is necessarily long-term and complex. But at least in theory, the contracting approach permits the physician to maintain a high degree of professional autonomy and independence while still participating in the market for HMO benefits.

My specific interest is with this second method of integration. In particular I use transaction cost economics to examine the efficacy of using an IPA to provide medical services under contract to an HMO. I argue below that an IPA is a viable method of organizing the delivery of HMO medical services, but only if the parties employ a "proper" governance structure.

other pediatricians. This type of association might be used to provide coverage over certain periods of time (like every third weekend), or alternatively, it might entail consulting on cases, providing second opinions, exchanging information, and sharing expertise.

6. Vertical relationships are those between physicians with different expertise. The motivation to enter into these relationships might include anything from the need to consult with a specialist, to a recommendation that a patient see a particular doctor, to a network of referrals from primary care physicians. For example, an internist whose patient has cancer might need to consult with or recommend a particular oncologist.

7. Increasingly, however, group practice physicians are finding an attraction to satellite clinics.
By governance structure I am referring to the mechanism that parties develop or otherwise acquire to monitor, adapt and enforce their contractual relations. There are a wide variety of governance mechanisms available to parties who desire to project exchanges into the future. Internal organization, courts, regulatory bodies, markets, and private orderings all can provide the necessary help under the correct circumstances. A “proper” governance structure is one that is correctly matched to a given set of underlying transactions. Transaction cost economics, which I discuss below, suggests that in this instance a correct matching requires a careful, microanalytic examination of the physician transactions that make up the IPA.

I argue below that a proper governance structure for the complex physician contracting relationships that comprise an IPA requires the parties to use a “private ordering.” That is, the governance arrangements used to handle the wide range of (potential) contracting problems must be internally derived and created by the parties themselves for the particular set of transactions at hand. More generally applicable governance methods, markets or courts for example, will not work effectively. One can predict that if an IPA relies on these governance systems, i.e., markets plus courts, it will encounter serious economic difficulties.

Ironically, what I term a “proper” governance structure requires that the participating physicians give up at least some of the autonomy they anticipated retaining. The result is that a well functioning IPA will differ from the group practice more in degree than in kind. An attorney advising a client or drafting an agreement must recognize this in order to best serve the interests of his client.

I begin the next section with a brief description of the governance structure that doctors and lawyers often believe will be adequate to handle IPA contracting. I refer to this arrangement as a “general purpose” governance structure and I use it to establish a point of contrast. In section III I then introduce the reader to transaction cost economics and identify the very limited set of conditions under which a general purpose governance structure will be successful. Section IV argues that the contracting world of physicians fails to meet the assumptions required for general purpose governance and therefore will prove inadequate. Instead the parties will require a more customized institutional arrangement—a private ordering. Section V briefly describes some of the attributes required of an IPA governance structure. In

---

8. I take this term from Galanter, though I operationalize it in a somewhat different manner than he does. Galanter, Justice in Many Rooms, Courts, Private Orderings, and Indigenous Law, 19 J. LEGAL PLURALISM 1 (1981). Elsewhere I have referred to what I call “private orderings” as “specialized governance.” As far as I am concerned, the words are interchangeable, but the former has a little more intuitive appeal.
section VI I note that a “proper” IPA governance structure will not differ as much as many imagine from the arrangements supporting a group practice. This in turn implies that while the IPA, if governed properly, is a viable method of organizing an HMO, the member physicians must be prepared to give up a significant degree of autonomy.

II. GENERAL PURPOSE GOVERNANCE STRUCTURES

Many physicians, and often their attorneys, venture into the formation of an IPA with the same objectives and expectations they would possess in any arms-length contracting. Their primary objective is to devise a comprehensive agreement. In this instance that contract, at a minimum, must specify (1) the services to be provided by each physician; (2) a minimum expected quality level; (3) a budget constraint; (4) a compensation formula; and (5) an arrangement for acquiring, paying for and sharing long-lived capital assets. An attempt might be made at the outset to anticipate and handle potential problems, but often little consideration is given to how the parties will monitor, adapt, and enforce the agreement once executed. Under these circumstances the physicians implicitly have assumed that financial considerations will align incentives to everyone’s mutual interest. If somehow financial concerns fail in their intended role the doctors expect to end up in court. In other words, the parties implicitly anticipate that their agreement will be regulated by some combination of markets and courts.

I refer to governance structures of this nature as “general purpose” arrangements. They are general in the sense that they are relatively accessible to any who want to employ them and are always external to the basic transaction in question. That is, the parties to the transaction do not expend resources to develop or refine the governance arrangements, though of course they might make expenditures to use the structures. Similarly, the parties to the transaction do not control the processes, deliberation or, to a large extent, the outcomes of a general-purpose governance structure.

United Healthcare (UHC), an HMO organized by the SAFECO Insurance Company, attempted to employ a governance structure that I classify as generally applicable.10 UHC was organized during the


10. To be sure, UHC in its last year, attempted to transform its governance structure. THE SAFECO PLAN, supra note 9, at 115-52 and 222-27.
1970s as an independent practitioners association. Unlike a more traditional IPA, UHC contracted only with primary care providers (PCPs)—general practitioners, family physicians, internists, pediatricians, and some obstetrician/gynecologists. The primary care providers were to act as gatekeepers who, by managing and regulating the patient's access to specialists, emergency department use, and hospitalization, were to control the health care costs of their patients. An account was set up for each physician and a percentage of the premiums paid by her patients was deposited to it each month. The account was used to compensate the physician for the services she provided the patient and to reimburse hospitals and specialists. The primary care provider's compensation was based on either a discounted fee-for-service or a per capita basis, depending upon how many HMO members were in the physician's panel. Hospitals and specialists, neither of whom had contracts with UHC, were reimbursed 100% of their charges.

Specifically, UHC used three mechanisms to enforce the primary care provider's obligation to minimize costs. First, the PCP was required to review and approve each bill submitted by the hospital or specialist. Second, the primary care physicians were placed at financial risk for any deficit in their account at the end of the year.11 Finally, UHC periodically exercised its right to remove a PCP from the plan.

Of the three enforcement mechanisms, SAFECO believed that the second was the most important.12 Thus, the plan's creators believed that the market oriented, financial incentives built into the initial contract would provide the necessary enforcement mechanism. Internally generated procedures or norms were given little consideration until 1981, one year prior to UHC's closing. Until that time, the HMO employed neither concurrent nor retrospective utilization review and did not mandate second opinions or preadmission authorization. As the study stated:

United Healthcare by intent had no monitoring or utilization review system during most years of its operation. It was the opinion of UH management that the gatekeeper or case management approach, if used properly, would negate the use of utilization review. Concurrent hospitalization review was begun after high hospital costs were noted in 1979.13

11. The physician shared 50% of the deficit or surplus with UHC, up to a maximum of 10% of her own fee-for-service revenues for UHC enrollees. In a panel of 200 or more patients where the PCP was paid a fixed sum per capita for her services, the PCP's maximum share of the surplus was increased from 10 to 50% of her fee-for-service revenue from UHC and the share of any deficit was decreased from 10 to 5% of that revenue. COST CONTAINMENT, supra note 9, at 48.

12. "Safeco was convinced that the incentives offered to physicians were the reason for lower costs in closed-panel HMOs." COST CONTAINMENT, supra note 9, at 49.

13. THE SAFECO PLAN, supra note 9, at 110.
Thus, the penalty and reward structure used to enforce the UHC contract was derived from sources outside the intra-physician transactions. Individual self-interest, market substitutes, rules of professional conduct, and traditional tort or contract remedies provided the primary sources of enforcement incentives. Hierarchy played no important role. To the extent incentives were intrinsic to the original contracting, they tended to be financially oriented and derived during the initial contracting stage. Where incentives are financially oriented the parties are relying fundamentally upon the existence of markets to guarantee their agreement.

The method of monitoring the agreements between UHC physicians revealed similar characteristics. Principal responsibility for monitoring was placed on the physician’s financial incentives to comply with the plan’s cost objectives. That is, parties were assumed to self-monitor because it would not be in their financial best interest to do otherwise. Thus monitoring was handled extrinsic to the basic physician transaction. The parties either relied upon outside institutions, like markets, to monitor the agreement or they did no monitoring. The organizers assumed that the terms of the agreement would be in the best interests of the physician. They believed that the doctor would realize that through her failure to comply with quality or cost standards the IPA might fail, she might lose patients and income, or the IPA might ask her to leave.

Further evidence of extrinsic sources of monitoring can be found in the heavy reliance that was placed on hospital and professional standards review organization (PSRO)\textsuperscript{14} oversight of the quantity and the quality of a physician’s services. In addition, little prospective monitoring was done. That is, there was virtually no attempt to screen prospective physicians for desirable characteristics. United Healthcare allowed any physician who wanted to join to do so. In a sense, all prospective monitoring was left to external sources—in this case, medical schools, residency programs, and licensing boards.

Of course, intrinsic monitoring was never totally absent. One study of United Healthcare reveals that the physician relations staff made sporadic checks of hospitalization rates and while the quality of care was not officially checked, the medical director did examine claims to identify flagrant abuses.\textsuperscript{15} In addition information feedback systems and physician education was minimal. Monthly newsletters

\textsuperscript{14.} Havighurst and Blumstein define PSROs as “[S]elf-regulatory organizations of physicians which are charged with monitoring individual physicians’ decisions affecting the use of health care resources under federal health programs.” Havighurst and Blumstein, \textit{Coping With Quality/Cost Trade-Offs in Medical Care: The Role of PSROs}, 70 NW. U.L. REV. 6, 8 (1975). Self-regulatory organizations were first organized pursuant to provisions of the Social Security Amendments of 1972, 42 U.S.C. § 1320 (Supp. II, 1984).

\textsuperscript{15.} \textit{THE SAFECO PLAN}, \textit{supra} note 9, at 111.
or cost statements were sometimes used to send information to physicians, but it is unclear how much these affected physician practice habits or whether the doctors even understood them.\textsuperscript{16} As the study of UHC found:

Thirty-eight percent of the physicians in the Richardson study said they were more aware of costs after joining United Healthcare, but sixty-two percent of the physicians said their awareness of costs had not changed. Furthermore awareness of costs was not related to length of time in the plan, number of patients in the plan, or account balance.\textsuperscript{17}

III. TRANSACTION COST ECONOMICS

In October, 1981, the SAFECO Insurance Company decided to terminate its United Healthcare operations. One of the reasons for the decision was the inability of the HMO to achieve the anticipated cost savings and utilization control. Though various reasons have been put forth for the demise of UHC\textsuperscript{18} and other IPAs,\textsuperscript{19} none explicitly accounts for transaction cost considerations. Transaction cost theory suggests that a partial explanation for UHC's problems in effectuating utilization controls can be found in their use of generally applicable governance structures. I argue that governance structures of this nature are incompatible with the underlying physician transaction.

An introduction to some of the basic tenets of transaction cost economics may be helpful to those unfamiliar with the subject. Transaction cost theory is concerned with the organization of economic behavior. The theory argues that, all other elements held constant, economic efficiency requires parties to organize their contracting to minimize transaction costs. In other words, they will either devise or choose the least costly governance arrangement that best facilitates their exchange.

Transaction costs, quite literally, are the "costs of running the economic system."\textsuperscript{20} They are "the economic equivalent of the friction in

\textsuperscript{16} Id. at 108.

\textsuperscript{17} Id. (The Richardson study referred to is the COST CONTAINMENT study, supra note 9).

\textsuperscript{18} COST CONTAINMENT, supra note 9; THE SAFECO PLAN, supra note 9.

\textsuperscript{19} Principle among the reasons given for IPA and HMO performance have been the structure, management, and financial incentives of the organization, see H. Luft, HEALTH MAINTENANCE ORGANIZATIONS: DIMENSIONS OF PERFORMANCE 351-80 (1981), mistakes in marketing the plan, lack of widespread exposure of panel physicians to fiscal risk in the plan, lack of review mechanism for regulating hospital utilization, adverse selection, and high hospital utilization, see Sorenson, Saward, & Wersinger, The Demise of an Individual Practice Association: A Case Study of Health Watch, 17 INQUIRY 244, 250-52 (1980).

\textsuperscript{20} JOINT ECON. COMM., 91ST CONG., 1ST SESS., THE ANALYSIS AND EVALUATION OF PUBLIC EXPENDITURES: THE PPB SYSTEM 47, 48 (Comm. Print 1969) [hereinafter cited as THE PPB SYSTEM].
physical systems”21 and refer to the human and environmental conditions that make contracting potentially costly. It is the existence of transaction costs that makes governance structures necessary.

Transaction costs of three types are usefully distinguished. First, human decisionmakers are only limitedly rational. They seek their own self interest, and generally prefer more of most things to less of them. However, they have only a restricted capacity for gathering, evaluating and storing information. They cannot see the future and they have only qualified problem solving capabilities. In short, they possess what Simon terms “bounded rationality.”22

Second, human decisionmakers can and do act opportunistically. On the one hand they attempt to exploit the advantages which may be attained from the making of “false or empty, that is self-disbelieved threats and promises”23 concerning future conduct. On the other hand, they distort or selectively disclose information.24

Third, the environment in which these actors interact is often uncertain and complex, thereby causing bounded rationality constraints to be met. In addition, the world of contract often is populated by only a small number of potential buyers and sellers. Where the participants are few, either markets cannot be relied upon to work correctly or situations akin to monopolistic competition develop. Of special interest to transaction cost economists are those circumstances in which a large number of buyers and sellers exist at the outset but are reduced substantially once initial bargains are set. What appears initially to have the requisite buyers and sellers for a well functioning market devolves into some form of monopolistic or “small numbers” bargaining environment.25

In a world without transaction costs, any governance structure would be more than adequate for the needs of the parties. More precisely, in a world where transaction costs are zero, the choice of governance structures is irrelevant because, by definition, enforcement, monitoring, and adaptations are effected costlessly. Hence, a generally applicable governance structure, consisting of some combination of markets and courts, would have neither a comparative advantage nor disadvantage over other imaginable governance structures.26

However, if nontrivial levels of transaction costs are present, gen-

24. O. Williamson, supra note 21 at 47.
eral purpose governance will usually prove inadequate. The combination of bounded rationality and complexity/uncertainty severely limits the ability of parties to draft self-enforcing, once-and-for-all, contingent claims contracts. The alternative is for the parties to use "adaptive sequential contracting"—crossing bridges as they are reached and renegotiating agreements as needed. But contracting of this nature requires that the parties be able to rely upon their partners' promises to act in good faith at the contract renegotiation or renewal stage. The presence of opportunistic conduct means that there is no a priori reason to believe in the other parties' promises. If markets work correctly, the problems would be minimized because one could always take one's business elsewhere. However, to the extent that the parties face a world in which small numbers bargaining conditions prevail, markets will fail to exist or will work improperly.

Thus, in a world with transaction costs, parties need to give thought to the types of governance structures they use. The precise nature of the governance mechanism is dependent upon those characteristics of the transaction that give rise to the contracting frictions that, in turn, make contract execution potentially costly. Williamson argues that the principal determinants of governance choice are (1) the extent to which investment is transaction-specific, (2) the frequency of exchange, and (3) the level of uncertainty involved. Of the three, he considers the first factor to be the most important. Simply put, he believes that parties with significant investments in non-redeployable assets will attempt to govern their exchanges with "private orderings"—that is, transaction-specific or customized governance structures—like firms or relational contracts. Where the capital investments can be transferred to other uses, the parties are less likely to incur the costs of private orderings and will tend to use more general purpose institutions like courts or markets. The reasoning is straightforward. The more specialized the investment, the lower its value in its next best use. Consequently, even if there were a large number of buyers and sellers at the outset, the number will be greatly reduced once an actual agreement is reached. Small numbers conditions can raise problems during the performance or renegotiation stage by heightening the risk of opportunistic behavior. Both sides, therefore, are willing to make increased expenditures to protect themselves against the possibility that the other party will attempt to exploit her ex post bargaining advantage. These expenditures take the form of governance mechanisms that are bilateral and unique to the parties. Conversely, as the investments become more standardized the parties have less incentive to make extra expenditures on governance. After all, by definition they have alternative sources of sale or

supply. Consequently, they are more willing to use the generally applicable alternatives that already exist.28

IV. THE NATURE OF THE TRANSACTION

The last section used transaction cost economics to argue that a general purpose governance structure requires either that bounded rationality constraints not be met, that opportunism not be prevalent, or that capital investments be redeployable. This section demonstrates that the contracting world of physicians fails to meet these assumptions. In particular, physicians face three types of transaction costs.

First, bounded rationality constraints are met because the world of physicians is complex and fraught with uncertainty. Initial agreements concerning compensation, work loads, patient allocation, risk pooling, and capital expenditure require some assessment of the future. Yet at the outset of the agreement, physicians are unable to determine whether variables like the demographic characteristics and tastes of their partners29 or patients will vary over time, and if they do whether the changes will be significant. Similar difficulties arise in determining how research and technical change will influence the demand for, and usefulness of, the equipment and physician skills existing at the time of the initial contracting. Finally, there is the continuing and very real possibility of unexpected changes in exogenous conditions such as the regulatory, political, or economic atmosphere.

The uncertainty of the future and the complexity of the agreement make it impossible at the outset to derive a complete, once-and-for-all

28. Uncertainty increases the importance of devising a mechanism for working out problems “[s]ince contractual gaps will be larger and the occasions for sequential adaptations will increase in number and importance as the degree of uncertainty increases.” Id. at 254. Consequently, as the degree of uncertainty rises, the use of elaborate, transaction-specific governance structures increases likewise.

Expected frequency of interaction also influences the institutional choice. Because governance involves significant set-up costs, the lower the frequency of exchange—and consequently the fewer the expected opportunities to avail oneself of the institutional structure—the more likely the parties are to want to cost-share. The result will be governance structures usable by a larger number of parties. Markets and courts, for instance, are excellent examples of generally applicable institutions. For a given degree of investment specificity, as the frequency of exchange increases—and consequently, the expectation for use of the governance facilities—the more likely the parties are to be the only users of their governance structure. Williamson, supra note 27.

29. There is also uncertainty about whether a partner’s skills, philosophy, work habits, and personal incentives—such as the physician’s “taste” for money—will remain constant over the life of the contract. While a doctor might know her partners’ characteristics at the outset, it is not until the execution stage that she will be able to tell just how skilled, reliable, hard working, avaricious, and easy to work with the physicians are.
contract. The alternative is to leave the contract incomplete and fill in blanks during the execution stage by renegotiating the agreement as events unfold. For adaptive sequential contracting of this nature to be effective, a physician must be able to rely upon her partners’ promises to act in “good faith” during renegotiation.

However, there is always that possibility that one’s contracting partners will cheat, break the agreement, or fail to act in good faith at the renegotiation stage. Professionalism and collegiality notwithstanding, there is no a priori reason for a physician to rely more heavily on another doctor’s promise than on the word of anyone else. In the physician context, contract breaches might include the failure to deliver promised services, quality levels, hours of work, or coverage. Compensation incentives can help to alleviate these problems, but only if the parties can specify the conditions that will trigger inducements and then agree that they have occurred. But compensation agreements of this nature can lead to cheating in the form of patient “stealing.” This can be especially prevalent in a competitive environment where compensation is based at least in part on a per capita or fee-for-service basis. Along similar lines, physicians might refuse to provide services in return for the agreed upon compensation. Physicians in high income specialties such as surgery or radiology often come to believe that they are subsidizing the primary care doctors. Whether this is true is irrelevant to my analysis. As long as surgeons or other specialists perceive it to be true, they are likely to become unhappy with their existing compensation. They might respond by demanding more money, unilaterally reducing their case load, or adhering strictly to the letter of their contract. Finally, a physician might simply impose on her partners more costs in caring for her patients than she promised. She might spend more time, require more tests, refer to outside specialists more often, admit patients more readily, or keep them hospitalized longer than she promised.

Of course, to the extent that alternative sources of supply or demand exist, the parties are protected. If a surgeon stops doing what she promised, her partners can simply replace her. If her colleagues do not act in good faith when an unexpected contingency arises, she can leave and find work elsewhere.

Dissolving a partnership is never costless, and this is all the more true if the parties have made significant investments in nonredeployable, nonrecoverable assets. If for some reason the partnership were to dissolve, these assets would be difficult to sell. In other words, the value of these assets in their next best use would be greatly reduced. I am interested in two types of physician investment. On the one hand, doctors purchase costly, long-lived capital assets and technology. Even though there may be buyers for the physical capital, rapid technological change, imperfect depreciation, inflation or deflation, valuation
problems and contentious bargaining can lead to nontrivial transfer costs.\textsuperscript{30}

On the other hand, physicians also make significant transaction-specific investments in human capital. First, they invest in specialized skills and reputations.\textsuperscript{31} A physician's reputation provides information on her temperament, philosophy and the quality of her work. It also conveys information to consulting physicians about the quality of care a patient has received in the past.\textsuperscript{32} In addition, reputation can provide a ready means of marketing one's assets and differentiating one's product. They tell other physicians something about how communicative, responsive, and sensitive to preexisting doctor-patient relations a colleague is.\textsuperscript{33} A physician must expend considerable resources in order to develop and communicate any changes in her reputation. In this sense, then, the investment in reputation is not costlessly redeployable.\textsuperscript{34}

\textsuperscript{30} Williamson makes a similar point in the context of the transfer of the right to provide cable television. Williamson, Franchise Bidding for Natural Monopolies—In General and with Respect to CATV, 7 BELL J. ECON. 73 (1976).

\textsuperscript{31} Formally, the exchanges between doctors, have all the characteristics of what Satterthwaite defines as a good reputation:
1. Each seller's product is differentiated from every other seller's product.
2. Product quality is consumer-specific, i.e., one perfectly informed consumer may prefer seller i's product over seller j's product, while a second perfectly informed consumer may prefer seller j's product over seller i's. This results from the fact that different consumers value each seller's product's attributes differently, rather than that different consumers perceive the attributes of a seller's product differently.
3. The attributes of each seller's product can only be fully evaluated by experience with the product over a significant length of time.
4. The product is important to consumers, i.e., each consumer is willing to expend significant effort to find a seller offering a product that is, according to his particular preferences, of high quality and reasonable price.


\textsuperscript{32} In choosing a covering physician or a specialist, doctors are particularly concerned about the quality of the service that will be provided. The quality of their associates' services reflects upon their judgment, affects the care their patients receive, has an impact on their work load, and can influence the patient's desire to remain with the primary care physician or clinic. Thus the primary care physician will want some assurance that the doctor with whom she has entrusted her patient has standards, skills, a demeanor, and a philosophy compatible with her own. She will also want some assurance that she can have confidence in the advice, information, second opinion or services she is obtaining.

\textsuperscript{33} Similarly, the covering or referral physician needs to know that her patient has had an adequate medical work-up and exam. Without this assurance, a second examination or set of tests might be ordered. Besides the obvious monetary costs involved in duplication, a reputation for redoing initial work-ups can lead to bruised egos, legitimate disagreements, and, ultimately, to fewer referrals.

\textsuperscript{34} Of course, to the extent other physicians are looking for the identical mix of skills, philosophies, standards, etc., the investment is not perfectly idiosyncratic.
Second, doctors make nonredeployable human capital investments in their relationships with their patients. This relationship is commonly referred to as the “doctor-patient relationship” and it requires an investment of both time and resources. In an HMO setting this relationship is not necessarily redeployable if the doctor were forced to leave the practice. Quite often, the patient’s membership in the plan will run for a minimum fixed period thus limiting his ability to change plans and follow the physician.

Physicians also make transaction-specific human capital investments in administrative functions. For instance, devising a coverage schedule takes considerable time and effort. Someone has to create, coordinate, administer, and adjust the schedule. This will require meetings and the resolution of disputes, especially over complicated questions of billing. For instance, an understanding will have to be reached concerning payment for patients admitted to the hospital by a covering physician. These investments in coordination, communications, and administration would also be lost if the partnership were dissolved.

But the number of alternatives is limited, because not every doctor will want to practice with a physician who has a particular mix of attributes. In other words, investing in a reputation is inevitable and important because it conveys easily processed information to other physicians. But by the same token, that reputation locks its owner into a particular designation.

35. Learning about, understanding, and treating a patient’s health problems requires the patient’s cooperation. A careful and informed diagnosis of current medical problems requires an understanding of a patient’s history. The doctor makes investments not only in taking that history, but in reducing it to words and recording it in a manner understandable to herself and to other practitioners. Patients provide information to doctors imprecisely and in code. The doctor, to be effective, must learn to read that code. Similarly, the physician must learn whether she can rely on the patient’s information. For instance, how likely is it that the patient complaining of a particular problem actually has it. She also needs to determine just how much information the patient can handle. The doctor also makes investments in the doctor-patient relationship to ensure desirable reputation effects. Finally, whether a patient continues on a prescribed regimen of medication, diet, exercise, or the like, depends in part on the degree to which she trusts the doctor’s diagnosis and judgment. As Mechanic points out,

36. The general rule is that the patient is the patient of the admitting physician. But what happens when the admitting physician does little more than simply sign someone else’s pre-existing patient into the hospital? Who retains responsibility for the care and receives the primary billing?
V. THE REQUIRED GOVERNANCE STRUCTURE

Section III argued that transaction costs of the nature described in section IV preclude supporting the physician contract with generally applicable governance structures and instead require a private ordering. This section illustrates what is meant by the term "private ordering." The required governance structure is not described in complete detail here; rather, the concept is illustrated to help an attorney in his role as a transaction cost engineer.

By the term "private ordering" I am referring to a specialized governance structure that is internal and is customized to the transaction. The primary purpose of the arrangement is to foster cooperative and adaptive behavior. Internal organization, vertical integration, and hierarchy—all characteristics of the group practice—are typical examples of specialized governance structures. Properly conceived, the governance structure of the IPA can achieve similar results. Where the group practice relies upon the geographic proximity of the participating physicians to foster social cohesion, a team-orientation, informal interaction, interdependence, and hierarchy, the IPA must find specialized governance through "quasi-integration"—that is through a contractual relationship where nominally independent parties are so closely related that they approximate a vertically integrated enterprise.

---

37. This is not to say that a governance structure must be internally generated to be transaction-specific. One could imagine an external agency of some sort that specialized in enforcing only a certain class of transactions. Williamson, supra note 30, offers the infrequency of interaction as one explanation for why parties to transactions involving idiosyncratic capital might want specialized, external agencies to develop. Regulatory agencies like the Federal Communications Commission or the Interstate Commerce Commission might be classified as specialized. However, because a large number of similar, yet different, transactions must be handled by an agency of this sort, they are relatively less transaction-specific than governance structures that are derived by the parties themselves. In fact, parties to transactions involving idiosyncratic investments find that even specialized agencies are inadequate and, therefore, look for means to avoid their auspices. Palay, Avoiding Regulatory Constraints: Contracting Safeguards and the Role of Informal Agreements, 1 J.L. ECON. & ORGANIZATION 155 (1985). My point, then, is that the existence of an internally generated governance structure indicates transaction-specificity. The reverse need not be true.

38. Not only does this indicate the degree to which the transaction remains autonomous from outside governance, but internally generated incentive systems are not easily transferred to other parties or exchanges. Thus they are indicative of a transaction-specific governance structure.


Take for instance the IPA governance structure required to enforce and monitor physician agreements on hospital, laboratory, and medical services utilization.\textsuperscript{41} Rather than rely upon markets and courts to enforce the agreement on utilization, the parties will need to develop and monitor a penalty and reward structure that combines formalized peer interactions with expressed norms. The former refers to the joint decision-making, consultation, meetings, review activities, and overall reinforcement and scrutiny that constitutes the environment within which the physician works.\textsuperscript{42} Geographic dispersion, infrequent physician contact, and a strong countervailing physician desire to maintain autonomy results in a less cohesive and homogeneous underlying social structure than in the group practice. Thus, where peer interaction in the group practice can be carried out “in the halls” or while “sharing charts,”\textsuperscript{43} the IPA requires relatively regularized and scheduled meetings, peer review proceedings, and administrative procedures. Because meetings take up relatively large blocks of time investments in the peer incentive structure of the IPA might appear to be greater than in the group setting. However, once one adds up all the time spent on informal social control in the group setting, the differences in the size of the investments in peer interactions is likely to prove illusory.

To develop an internalized enforcement mechanism, the parties will also need to emphasize the development of administratively devised rules and procedures. For instance, the IPA will need to employ some combination of administrative controls on admissions by outside specialists, referral controls, hospital utilization review, concurrent review, retrospective review, discharge planning, and mandatory second opinions. The objective will be to modify physician practice patterns by reminding the doctor of the need to be cost conscious. For the most part these administratively devised norms are a means of communicating standards and establishing benchmarks of behavior. For instance, outside referrals to specialists not affiliated with the HMO can be particularly costly. Both the direct fees charged by the specialists and the potential lack of control on hospital admissions can constitute a signif-

\textsuperscript{41} In the past I have found it quite useful to also examine how parties adapt their contracts to changed circumstances. Palay, \textit{Comparative Institutional Economics: The Governance of Rail Freight Contracting}, 13 J.L. STUD. 265 (1984)). However, the present data is such that meaningful descriptions of adaptations are impossible.

\textsuperscript{42} G. MEIER \& J. TILLOTSON, PHYSICIAN REIMBURSEMENT AND HOSPITAL USE IN HMOs (1978) 73, 75 \& app. B.

\textsuperscript{43} In the group practice, peer interaction is much more informal than that required by IPAs. Interactions can occur naturally in the course of watching what others do, through contacts in the hall, over lunch, through conversations in the doctors' lounge, or similar unstructured activities. Real resources are expended both in developing the trust relationships required by an informal process and in the numerous small blocks of time used for consultations and the like.
icant drain on revenues. HMOs often use a variety of controls to cause the physician to think carefully about whether the referral is warranted. The most common method is to require a physician to get the medical director’s approval before making an outside referral. Similarly, HMOs use a variety of controls to limit hospital admissions by outside physicians. For instance, the department director might first screen referral admission for urgency. Those that are classified as elective would then have to be approved by the medical director.44

Of course, often the real incentive to adhere to utilization standards comes not from the administrative procedures, but from the sanctions—also internally derived—associated with noncompliance. For instance, InterStudy reports:

The New Mexico Health Care Corporation, with 375 participating physicians, could not rely on social structure to maintain cost-effective orientation. The HMO retained a comprehensive assemblage of control mechanisms and did not hesitate to enforce them, regardless of the severity, whenever it appeared warranted.45

The incentives can take one of several forms. Some involve monetary penalties especially, for example, where the physicians are put at personal risk for a percentage of the HMO’s revenue shortfall. Other penalties include denying payments for care considered unnecessary or inappropriate. Often the mere questioning of an occasional bill, without actually denying payment, is enough to remind physicians to be more cost conscious in the future.46 Others sanctions are related to the social stigma of being the only one in an association who is not meeting some standard. Still others involve terminating physicians who consistently fail to adhere to agreements on utilization.

Using formal norms as an incentive for enforcement is particularly important for IPAs. Unlike the group practice, IPAs cannot rely upon close geographic proximity to foster social cohesiveness and frequent communication. Thus, there is a greater emphasis on carefully delineating rules and codifying accepted practices as a means of communicating the norms and modifying utilization and admissions behavior. After studying nine HMOs, InterStudy concluded:

As is true with preadmission certification, formal concurrent review programs are likely to be more effective in an IPA setting, where practice patterns vary widely among physicians and the physician/HMO interaction may be less frequent. Concurrent review can occur informally in group practice settings, where physicians commonly discuss hospital cases; a formal concurrent review program may add little. The larger and less interactive the group becomes, the more effective a formal program may be. Of the five HMOs employing concurrent review, Health Maintenance Plan/Cincinnati was the only group practice; here the review portion was viewed as a program of major

44. G. MEIER & J. TILLOTSON, supra note 42, at 56.
45. Id. at 76.
46. Id. at B-7.
This is not to say that group practices do not establish utilization norms, but that they are likely to be less important than those found in an IPA.

To effectively support a long-term contractual relation, a governance structure must provide information concerning the status of transactions and the behavior of contracting parties. The data is required both to promote planning and ensure compliance. I refer to this data gathering and processing function as monitoring. The degree to which the parties to the transaction gather, process and act upon the relevant information, rather than rely upon others, helps to determine whether the monitoring is indicative of a private ordering. As with enforcement, intrinsic monitoring processes are not easily transferred to other settings and, therefore, indicate transaction-specific governance.

To generate the requisite information, an IPA will need a formalized, sophisticated, and technically complex feedback system. The Franklin Area Health Plan (FAHP)—a Maine based IPA—provides an example of one such system. First, a committee conducted an ex post review of the admissions practices and ambulatory care patterns of the participating physicians. A series of HMO and individual physician performance indicators were generated from the review. This information was both transmitted directly to the doctors and used in internal enforcement proceedings. Besides the formal retrospective review, FAHP also generated and processed informal information on concurrent review. For instance, the medical director received a continuous flow of updated data on current hospitalizations. He would review and discuss the cases with the admitting physician in an effort to minimize unnecessary hospital days.

As might be expected, the continuous interaction among group practice members reduces the value to them of sophisticated information systems. Consequently, one would expect the group practice to spend less resources on formal feedback and education. Instead, monitoring resources would be expended in the form of time spent on such activities as peer supervision and “chatting in the corridor.” Hard data would be used both to “validate impressions” and to identify trends and problem areas.

Some monitoring can be achieved prospectively by choosing carefully the physicians with whom to contract, thereby reducing the future need to do compliance monitoring. Historically, the group practices have been more successful at screening and selecting physi-

---

47. Id. at 76.
48. Id. at B-7.
49. Id. at 58.
50. Id. at B-7.
cians than have IPAs. The latter have faced competing objectives: containing costs by carefully selecting of physicians while broadening the patient base by maximizing the number of participating physicians. For marketing reasons, they have tended to reconcile these objectives in favor of broader participation. Thus IPAs tend to be less selective in choosing physicians and are often forced by their organizers (especially where these are the local medical society) to accept any physician who applies. This can mean having to hire physicians who are high utilizers or are not cost effective. There are exceptions to this tendency and there are IPAs with successful physician selection programs. For instance, at the New Mexico Health Care Corporation, there is an annual review of each physician's cost and quality of care record. The IPA requires the review as part of its reappointment procedures and providers have been refused reappointment in the past. But in general, the IPAs have required more sophisticated information feedback systems to compensate for inadequate prospective monitoring.

VI. IPA VERSUS GROUP PRACTICE GOVERNANCE STRUCTURES

Throughout the previous section I implicitly argued that transaction-specific governance could be defined in terms of the extent to which governance tasks are handled internally by the parties to the transaction. Ironically, this makes the successful IPA and the group practice HIO quite similar. Both must develop their own transaction-specific governance structures. To be sure, there are differences between the two types of HMOs. But those differences tend to be a product of the variations in the communications and social structure developed by the two organizations. The group practice can take advantage of close physician proximity to foster a strong social structure and communication network. This in turn permits interactions to be more informal and norms to be established by example rather than by fiat, though the impact of hierarchy and a powerful medical director must not be discounted. The IPA, however, must rely more heavily upon formal processes, meetings, and rules. But the differences witnessed between the IPA and the group tend to be more in degree than in kind. Despite the evident differences in organizational form, the IPA and the group practice are both capable of providing similar governance structures for physicians practicing in an HMO setting. Both governance structures are transaction-specific and permit the parties to maintain fundamental control over their governance arrangements.

51. Id. at 58.
52. Palay, supra note 37.
53. Id.
This would not be true if a general purpose governance structure were used.

To put the point a different way, it is possible to develop an internalized governance structure even if the parties choose not to organize themselves as a firm. Parties that are more loosely associated—as in an IPA—can effectively develop the necessary governance mechanism, but only if they are prepared to become quasi-integrated. Of course, one of the attributes of quasi-integration is that the contracting parties, in varying degrees, give up their autonomy of action. They agree, more or less, to subordinate their individual interests to their joint interests. This will certainly have an impact on a physician's HMO practice patterns. But in addition, to the extent a physician is unprepared to maintain two practice patterns—one for the HMO and one for her other patients—her non-HMO health care services will also be influenced. Thus, physicians who want to participate in an HMO must be prepared to give up some of the autonomy they have traditionally cherished.

VII. CONCLUSION

Gilson hypothesizes that attorneys add value to their client's transactions by devising efficient governance structures. In the context of his analysis the attorney's role is to devise optimal contract terms. But if an attorney is to truly add value to his client's transaction then he must recognize that clever contract drafting is but one alternative for solving governance problems. Effective transaction cost engineering requires a careful matching of governance structures to transaction characteristics. The attorney must look to the characteristics of the underlying transaction that make contracting costly and design governance structures accordingly. Occasionally this will demand efficient contract terms; but often the design of more complex governance structures will be required. The attorney who fails to recognize that contracts must be drafted with an eye to the accompanying governance structure not only fails as a transaction cost engineer, but can contribute to the eventual demise or restructuring of the underlying relationship.

I have illustrated some of the conditions under which a more complex private orderings will be required. In the context of organizing an IPA I have argued that uncertainty and complexity in the underlying contracting, opportunistic conduct, and the extent to which investments are nonredeployable, makes specialized governance structures necessary. These structures will be devised by the parties as part of their ongoing relationships.

My argument strongly suggest that an attorney counsel his clients that a contract document alone will not guarantee the success of an enterprise. Simply because a group of doctors agrees to a set of con-
tract terms does not mean that the HMO will be a success. The lawyer
must make clear to the parties that all contractual relationships re-
quire some method of enforcing, monitoring, and adapting the agree-
ments and he must be prepared to make concrete suggestions about
the specific methods of governing the transaction. This in turn re-
quires the lawyer to understand a good deal about his client's enter-
prise. Where the parties involved are, like physicians, not particularly
familiar with the role of business relationships the attorney's role be-
comes all the more important.54

54. Macaulay implicitly makes this same point. Macaulay, Non-Contractual Re-