Overlapping Generations, Moral Hazard, and
the Organization of Medical Partnerships*

Martin Gaynor
Carnegie Mellon University and NBER

David S. Salant
Law and Economics Consulting Group

October 17, 1994

* We thank Steve Hackett, Joe Harrington, Tom Lyon, Bentley Macleod, Mark Pauly, Louis Putterman, participants in seminars at Northwestern University, the Justice Department, and in the Fourth Annual Health Economics Workshop at Northwestern University, and the Health Economics Conference at the University of Chicago for helpful comments. This paper was written when the authors were at Johns Hopkins University and GTE Laboratories, respectively. The usual caveat applies.

I. INTRODUCTION

It is well known that first-best outcomes in work organizations can not be obtained as equilbria of a static, one-shot game (Holmström, 1982; Holmström and Tirole, 1989). Recent developments in game theory have shown that repetition of play can lead to very different outcomes than those that result from a static, one-shot game.[1] If the game is repeated long enough, and if the players do not discount the future too heavily, near-efficient outcomes can result. Intuitively, this seems sensible. If partners are in a group over some period of time, and if observable outcomes convey some information about their actions, then they clearly have to take play in future periods into account, providing they don't discount the future entirely. Looking at it another way, adding periods to the static problem increases the degrees of freedom, simply because there are more choices. Thus it has to be true that the outcomes from repeated games are no worse than those of the static game. However, infinite repetition is required in these games to improve on the static outcome.

In this paper we consider a finitely repeated game among overlapping generations of members of a partnership. We show that a near-efficient outcome can be achieved provided the future is not discounted too heavily, and that an arbitrarily efficient outcome can be achieved if the game is repeated often enough.

We fully characterize the payoffs of all workers in this organization. Older workers' payoffs are optimally greater than those of young workers. The reason for this is that efficient behavior can only be induced by rewarding workers when old for having worked hard when young. This is similar in spirit to the arguments of Lazear (1978) for non-partnership firms.[2]

Group practice has become the dominant setting in which most American physicians practice. Gonzalez and Emmons (1988) report that over 60% of physicians in private practice are in groups. Further, this number has been growing over time. Havlicek et al. (1993) report a 550% increase in the number of physicians in groups between 1965 and 1991.

In spite of this trend toward group practice in medicine, most economic models of the physician services market assume that physicians are engaged in solo practice. Given the importance of incentives in organizations, models of the physician firm as a solo practice may not provide an accurate picture of behavior.

The recent literature on the economics of medical group practice has concentrated on the internal organization of groups and the effects on incentives.[3] While these papers have shown that incentives affect physician behavior, and that these incentives are chosen rationally by groups, they have been couched in a static setting.

[1] E.g., Radner (1985), Fudenberg and Maskin (1986), Radner (1986), Radner, Myerson, and Maskin (1986).

[2] See also Harris and Holmström (1982).

[3] See Newhouse (1974), Sloan (1974), Reinhardt, Pauly, and Held (1979), Getzen (1984), Gaynor (1989), Hillman, Pauly, and Kerstein (1989), Gaynor and Pauly (1990), Lee (1990), Gaynor and Gertler (forthcoming).

In this paper we examine the internal organization of medical group practices in the context of a repeated game among overlapping generations of physician-members of a group. This builds on the work of Crémer (1986), Cooper and Daughety (1993), Shepsle and Nalebuff (1990), Salant (1991), Kandori (1992), and Smith (1992) on games with overlapping generations of players.

The paper is organized in three sections. The next section contains the theoretical model and results, and Section III contains a summary.

II. THE MODEL

The game to be considered is an infinitely repeated noncooperative game among players with finite but overlapping terms. Strictly speaking this implies that medical group practices must have infinite lives, but all that is really required is that they have long enough lives that they can be considered ongoing organizations. That is, these groups must have lifetimes which exceed the working lives of their members so that they can be said to exist for long periods of time.[4,5]

In what follows, we consider several issues related to the internal organization of medical practices in the context of this overlapping generations model. Section A contains the model without reputation effects. We consider the existence of a "nearly efficient equilibrium" which improves on the myopic equilibrium. We also consider the effects of seniority. A model with reputation effects is considered in Section B. We show that shirking is more likely when there is reputation. Section C contains a model with imperfectly observable types. We show that a separating equilibrium does not exist, and that screening may be optimal for the group, which may lead to an "up or out" system.

In order to focus the analysis, we make several simplifying assumptions. The first is that physicians' utilities are linear in income, i.e., they are risk neutral. We realize this is not a realistic assumption, but employ it to allow us to focus on generational issues.[6] The second is that profits are a function of all partners' efforts. While it is true that ambulatory care especially is typically produced independently by physicians, there certainly are joint fixed factors, such as space and equipment, and joint usage of non-physician variable factors, such as nursing time and clerical time. In addition, there is the possibility of externalities in demand, e.g., due to a firmwide reputation, although we will abstract from that initially.

[4]This requires that the members in any period always perceive that there will be a new generation of members in the following period. If the group fails, but the failure is unexpected, the results of this paper should carry through for all but the period in which failure occurs.

[5]Although there is no direct information on group ages, the AMA reports that since the mid-1980s the rate of growth in the number of groups has slowed and that group size has been increasing. These facts are consistent with (although not proof of) groups existing for some period of time.

[6]Gaynor and Gertler (forthcoming) provide evidence that physicians are risk averse.

A. THE MODEL WITHOUT REPUTATION EFFECTS

At the outset we will consider a model in which only current effort affects profits. Let production be represented by a profit function, where group profits are a function of all physicians' current period efforts,

$$\Pi_t = \pi(\mathbf{e_t}), \qquad \mathbf{e_t} = (e_{1t}, e_{2t}, \ldots, e_{nt}) \tag{1}$$

where e_{it} denotes the effort of agent i in period t and Π is profits of the partnership. Let there be N partners with T period terms, so annual turnover is A = N/T.[7] Also, let B = 1/A be the number of periods between partners retiring. Let partners share profits equally, so that the share accruing to the i^{th} partner = 1/N. We write partner i's utility in any period t as

$$u_{it} = \frac{1}{N}\pi(\mathbf{e_t}) - \mathrm{v}(e_{it}), \tag{2}$$

where $\mathrm{v}(e_{it})$ is the i^{th} partner's private, non-monetary cost of effort, i.e., disutility. We assume all payoffs to be net of the value of outside opportunities. Partners maximize their lifetime discounted utility, which is

$$U_i = \sum_{t=s}^{s+T-1} \delta^{t-s} u_{it}, \tag{3}$$

for partners who begin their term in period s, where $\delta \in [0,1]$ represents the discount factor, which for our purposes here is assumed to be common to all members of the group and constant over time. We further assume that the function v is strictly convex, i.e., v' > 0, and v'' > 0, and that the function π is strictly concave, i.e., $\pi' > 0$, and $\pi'' < 0$. To make matters as simple as possible, let profits be a function of the sum of all members' efforts, $\pi(\mathbf{E_t})$, where $\mathbf{E_t} = \Sigma_i e_{it}$.

[7] We assume that there are an equal number of members in each cohort, so that the definition of turnover holds exactly. Otherwise, this can be interpreted as holding in expectation.

1. The Game

Now consider the game between the members in which everyone chooses their own effort. At each date, the members each decide their own effort levels. The rules they use for determining effort levels are dictated by strategies, and can depend on the entire history of the game up to that date. A subgame perfect equilibrium is a set of rules for each partner satisfying the condition that for any history at any date, no partner can change their rule to increase their lifetime utility. Multiple equilibria exist in this game. There is a myopic equilibrium, which is clearly suboptimal. There is also another equilibrium, which we will label "a nearly efficient equilibrium," in which only the oldest partner(s) shirk. This is also suboptimal, but better than the myopic equilibrium. By extension, the full optimum is not an equilibrium, since clearly the oldest member of the group will shirk.

2. Myopic Equilibrium

Consider a myopic equilibrium which has each partner i setting his effort, e_{it}, at levels solving

$$\max_{e_{it}} U_i = \sum_{t=s}^{s+T-1} \delta^{t-s} \left[\frac{1}{N} \Pi_t - v(e_{it}) \right] \qquad 4$$

Denote the myopic equilibrium levels of effort by $\hat{e}_{it}$ and $\hat{E}_t$. Let $\hat{e}_{it}$ satisfy

$$\frac{\partial U_{it}}{\partial e_{it}} = \frac{1}{N} \pi'(\hat{\boldsymbol{E}}_t) - v'(\hat{e}_{it}) = 0 \,, \quad \forall\, i,t,\ t = s,\ldots,s+T-1 \,, \qquad 5$$

where f'(·) denotes the first partial derivative of the function f with respect to e_{it}. This is the myopic equilibrium condition.

This outcome fails to maximize long run group welfare (or profits), as effort levels that maximize these payoffs must solve

$$\max_{e_{it}} \; W = \sum_{t=0}^{\infty}\left[\; \Pi_t - \sum_{i=1}^{N} v(e_{it})\right], \tag{6}$$

where W is group welfare. Call the first-best levels of effort e^*_{it} and E^*_t. Let e^* satisfy

$$\frac{\partial W}{\partial e_{it}} = \pi'(E^*) - v'(e^*) = 0 \;\; \forall\, i,t\,. \tag{7}$$

Equation (7) defines the first-best.

Now consider partner i's best response to the other partners' actions. Let $E_{-i} = \Sigma_{-i} e_{-i}$.[8] Define $\hat{e}_i(E_{-i})$ such that

$$\hat{e}_i(E_{-i}) = argmax\left\{\frac{1}{N}\pi(E_{-i} + e_i) - v(e_i)\right\}. \tag{8}$$

This is any partner's optimal best response. We now consider the nature of best responses to other partners' actions.

Lemma 1: Any partner's best response is decreasing in other partners' actions. Further, it is less than fully proportionate to other partners' actions,

$$-1 < \frac{d\hat{e}(E)}{dE} < 0\,. \tag{9}$$

Proof: We can write utility as

$$\frac{1}{N}\pi\,[E_{-i} + \hat{e}_i(E_{-i})] - v[\hat{e}_i(E_{-i})]\,. \tag{10}$$

The first and second derivatives with respect to E_{-i} are

[8]We suppress the time subscript from here forward unless necessary for clarity.

$$\frac{1}{N}\pi'[E_{-i}+\hat{e}_i(E_{-i})]-v'[\hat{e}_i(E_{-i})]=0\,, \qquad \mathbf{11}$$

and

$$\frac{1}{N}\pi''[E_{-i}+\hat{e}_i(E_{-i})][1+\hat{e}_{i'}(E_{-i})]-v''[\hat{e}_i(E_{-i}]\,\hat{e}_{i'}(E_{-i})=0\,, \qquad \mathbf{12}$$

respectively.

Rearranging (12), we get

$$\hat{e}_i{}'(E_{-i})\cdot\left[\frac{1}{N}\pi''-v''\right]=-\frac{1}{N}\pi'' \qquad \mathbf{13}$$

Assuming concavity, $\pi''<0$ and $v''<0$, then

$$0>\hat{e}'(E)=\frac{-\frac{1}{N}\pi''(\cdot)}{\frac{1}{N}\pi''-v''}>-1$$

□

As long as $\pi''<0$ and $v''<0$, as seems reasonable, the myopic equilibrium will lead to partners working too little and investing too little. This is due to incentives, since each member receives $1/N<1$ of his marginal product.

<u>Proposition 1:</u> First-best partnership effort exceeds myopic partnership effort, $E^*=\Sigma_i e_i^*>\hat{E}=\Sigma_i\hat{e}_i$.

<u>Proof</u>: Note that e^* and E^* are determined by

$$\pi'(E^*)-v'(e^*)=0\,,$$
$$\pi'(\hat{E})-N\,v'(\hat{e})=0\,.$$

and that $\hat{e}$ and $\hat{E}$ are determined by
For $\hat{E}>E^*$, $\hat{e}'(E_{-i})$ must be ≥ 0. Thus Lemma 1 $\Rightarrow$ the result.[9] □

[9]Note that $\hat{e}_i'(E_{-i})>-1$ (Lemma 1) $\Rightarrow$ convergence to equilibrium.

3. Nearly Efficient Equilibrium

We now consider a nearly efficient equilibrium. When members of the group have overlapping terms, and when there is perfect information, i.e., everyone can observe everyone else's effort, then there is potential to improve on the myopic equilibrium, since anyone who shirks in the current period can be punished in the next period,[10] except for members who are in their last periods. The first-best equilibrium will not occur, because as the following lemma indicates, old partners cannot be induced to work hard or to sacrifice for the good of the group.

Lemma 2: The oldest member(s) in a group will always choose the level of effort that maximizes single period utility.

Proof: The oldest member(s) in a group are in their ultimate period, therefore they face no punishment for shirking in subsequent periods. Therefore they will choose the level of effort which maximizes their single period utility.[11] $\square$

Note that this implies that old partners shirk in some sense. More precisely, older partners have no incentive to sacrifice their current payoff for the benefit of the partnership's future income. However, younger associates can be induced to do so. Their inducement will be provided by their reward when they near retirement. The reward is a higher payoff due to exerting less effort. This essentially means shirking by the older partners.

It may be possible to mitigate shirking by the oldest members if mechanisms can be found to tie their payoffs beyond the last period to effort in the last period.[12] For example, retirement payoffs could be made conditional on effort while working. In partnerships, partners usually sell back their ownership share back to the group when departing. The valuation of this share could be made conditional on the last period's effort. These kinds of arrangements would be more likely the greater is the loss from the oldest members' shirking. This could be the case if older members are more productive than younger members, or if there are a disproportionate number of older members.

We now consider an equilibrium with the least amount of shirking, taking as given that the

---

[10] As long as shirking is always discovered before the next period's decision need be made.

[11] This may or may not be zero, depending on the shape and location of the disutility and profit functions.

[12] Harrington (1992) shows that a lame duck politician will act in the interests of his political party, as long as the politician cares about policies which are enacted after his term in office. Salant (1992) shows that the performance of regulated firms can be enhanced when regulators are permitted to work post-retirement for the firm they regulated, and vice versa.

oldest members of the group will shirk. Consider a situation in which the K oldest members shirk, $K < N$, and the remaining $N-K=M$ members cooperate. Consider the effort levels e^M and e^K that solve

$$\max_{e^M, e^K} \quad \frac{\pi(E^M + E^K) - M\, v(e^M) - K\, v(e^K)}{N} \qquad \textbf{17}$$

$$s.t. \quad e^K = argmax \; \left\{ \frac{1}{N} \pi \; [\; E^M + (K - 1)\, e^K + e \;] - v(e) \right\} \qquad \textbf{18}$$

where M denotes working associates, and K denotes non-working associates. Call this level of effort the M cooperative level, e^M. [13] This maximizes the average payoff across all N partners subject to K of them choosing myopic levels. Let S be the duration or number of periods in which associates work at the beginning of their terms and $R \equiv T-S$ be the period in which they begin shirking.[14]

We construct trigger strategies in which cheating results in reversion to Ê, the Nash equilibrium of the stage game, through the duration of the cheater's tenure. The cooperating agents are no worse off in the continuation game at the M cooperative levels of effort than at the myopic levels of effort. We note that whenever

$$\begin{aligned} &\frac{1}{N}\pi\left[E^M + E^K\right] - v\left(e^M\right) + \left\{\frac{1}{N}\pi\left(NE^M + E^K\right) - v\left(e^K\right)\right\}\frac{\delta\left(1-\delta^R\right)}{1-\delta} \\ &\geq \frac{1}{N}\pi\left[\left(M-1\right)e^M + E^K + \tilde{e}\right] - v\left(\tilde{e}\right) + \frac{\delta\left(1-\delta^R\right)}{1-\delta}\left[\frac{1}{N}\pi\left(\hat{E}\right) - v\left(\hat{e}\right)\right] \end{aligned} \qquad \textbf{19}$$

no associate would wish to deviate from a cooperative path which provided payoffs of $(1/N)\pi(E^M+E^K) - v(e^M)$ in the first S periods of each player's term and payoffs of $(1/N)\pi(E^M+E^K) - v(e^K)$ in the last R periods.

[13]Notice that $e^M > e^N$, by the concavity of the profit function and the convexity of the disutility function.

[14]Note that $K=(R/T)N$ and $M=(S/T)N$.

We suppose that

$$\frac{1}{N}\pi(E^M + E^K) - v(e^K) > \frac{1}{N}\pi(\hat{E}) - v(\hat{e}) \qquad \textbf{20}$$

Condition 20 says there are gains to cooperation. The left hand side of the expression is utility evaluated at M-cooperative effort levels. The right hand side is utility evaluated at myopic effort levels. Utility at the cooperative level must exceed that at the myopic level for this problem to be of any relevance. Note too, that for fixed R, if T is large and δ is near one, then the per period payoff that solves (17) subject to (18) can approximate $(1/N)\pi(E^*) - v(e^*)$ arbitrarily closely.[15]

Proposition 2 (Folk Theorem): Any outcome that dominate ê can be approximated at a subgame perfect equilibrium. Assume (20) is satisfied, so $\exists$ gains to cooperation. Then given $\varepsilon > 0$, $\exists$ $\overline{R}$ and $\overline{T}$ and $\bar{\delta}$ such that for $\delta > \bar{\delta}$, $T > \overline{T}$, $R = \overline{R}$ [16] there is a subgame perfect equilibrium with average payoffs within ε of

$$\frac{\delta}{1-\delta}(1-\delta^T)\cdot\left[\frac{\pi(E^*)}{N} - v(e^*)\right]. \qquad \textbf{21}$$

Proof: Let R and S safisfy (17) subject to (18).[17] By (20) such R and S exist. Let the strategies determining effort levels be such that:

1) (a) Associates of age less than S work at rate e^M.

(b) Associates of age more than S work at rate e^K,

provided this has been the pattern for the duration of all active players' terms.

2) Should any deviation occur, let $e_i = \hat{e}_i$ for all i as long as the last deviant is still around. Then revert to 1).

By (17) no associate will ever wish to deviate from cooperative play. Further, once a deviation has occurred:

---

[15]The expression is really $[(1/N)\pi(E^*) - v(e^*)][(1-\delta)/\delta(1-\delta^T)]$. Note that $(1-\delta)/\delta(1-\delta^T) \rightarrow 1$ as $\delta \rightarrow 1$.

[16]Note that R cannot be too large or it will not be possible to approximate optimal outcomes.

[17]Given T and N, R and S define K and M, respectively. See footnote 14.

(a) no associate can do better in the short run by deviating from ê,

and

(b) deviation can only reduce long run payoff to that associate by "restarting" the punishment period.

Then for any T > R, the above strategies form a subgame perfect equilibrium. In particular, for T larger than $\overline{T}$, this equilibrium approximates the optimal outcome, for δ sufficiently large.[18] □

Remarks:

1) Suppose ê is the minmax threat. Then we have optimal punishments.

2) a) This result is stronger when "shirkers" can be expelled, i.e., the condition

$$\frac{1}{N}\pi(E^S+E^R)-v(e^R)>\frac{1}{N}\pi(\hat{E})-v(\hat{e}) \qquad \mathbf{22}$$

would be replaced with

$$\frac{1}{N}\pi(E^R+E^S)-v\left(e^R\right)>0\,. \qquad \mathbf{23}$$

There exists a cooperative equilibrium for R and T large enough.

b) With increasing returns to scale, a (trigger) strategy of punishment by reversion to Nash equilibrium might yield higher payoffs to remaining members than does a strategy of punishment by expulsion of a shirker.

This is an extension and strengthening of Crémer's (1986) result, cast in the context of our model. We show that anything better than the minmax is an equilibrium, thus our Folk Theorem (Proposition 2) is stronger than Crémer's result. Some overlap is required in partners' terms.[19] If there is no overlap, the game reduces to a repeated game among a single cohort. This is a finite period game, with length equal to the players' terms, and suffers from the familiar problem of "the

[18]Note the following: 1) as T increases for fixed R and N, K decreases toward 0 and M approaches N, which affects (17) and (18); and 2) the fact that $\delta(1-\delta^T)/(1-\delta) \to 1$ as $\delta \to 1$ implies average payoffs converge to $(1/N)\pi(E^*)-v(e^*)$.

[19]With arbitrary games the amount of overlap needed will depend on the structure of the payoffs. See Salant (1991), Kandori (1992), and Smith (1992).

chainstore paradox" if there is a unique equilibrium.[20]

One thing this suggests is, *ceteris paribus*, that groups with little or no overlap in their members should be less efficient. In addition, if turnover in the group is too high, there may not be sufficient overlap, and there will not be sufficient opportunity for future punishment. Similarly, if a group has members with short terms, the possibilities for future punishment are restricted, and less efficiency should result. Higher discount rates reduce the efficacy of future punishment, so groups consisting of members with higher discount rates should be less efficient, all other things remaining equal.

In addition, it may not always be the case in the real world that punishment is subgame perfect. For small groups the punishment may be equivalent to disbanding the group, and thus could leave members worse off. Also, we obtained results assuming decreasing returns to scale in production. This is not completely realistic, since one of the reasons for forming a group is to exploit economies of scale. Nonetheless, most evidence suggests that economies of scale are exhausted at relatively small group sizes of three to five physicians (Reinhardt, Pauly, and Held, 1979), thus this issue would only arise for small groups or large groups where punishments might drive them below minimum efficient scale. It is likely, however, that monitoring is effective in small groups, thus mitigating the need for the strategies we have described in order to achieve efficiency. Further, we restricted cohorts to be of equal size. It it not clear if there is a degree of inequality between the cohorts at which the second-best cannot be supported, although we suspect this is the case.

4. Seniority

Seniority seems to be important in the real world. In medical practices new members are not allowed to join as partners immediately. They typically work as salaried employees for 2 to 4 years before becoming eligible for ownership (Kralewski et al., 1985; Lee, 1990). Nonetheless, it seems to be the case that medical practices are relatively horizontal, in that shares are not much differentiated by seniority. In this section we address the issue of the extent to which payoffs should optimally be differentiated by seniority.

Suppose that members of the group have two period terms, and in any period exactly half the members are young and half are old. Thus, there are N/2 partners of each type, and N partners in total. Now suppose that $\mathbf{e}^{\mathbf{s}}$ is the sustainable effort level of junior partners, and Π^{s} is the sustainable per period profit when there is equal sharing among all partners. Since the senior partners do not work, there doesn't seem to be any reason to pay them anything. Senior partners could be paid 0 and junior partners could each be paid 2/N. However, in this case there is no future punishment available to discipline junior partners who cheat. We now characterize effort levels for

[20]If there are multiple Nash equilibria in the stage game, the subgame perfect equilibrium of the repeated game can have multiple outcomes. Thus, improvement on the myopic equilibrium is possible. See, e.g., Benoit and Krishna (1985) and Harrington (1987). In our model, other equilibria emerge even when the stage game lacks multiple equilibria, due to the overlap in players' terms.

junior and senior partners which are sustainable in equilibrium.

Proposition 3: Suppose partners have two period terms, and at any date half the partners are old and half are young. Then the sustainable equilibrium effort levels for young and old partners, e^y and e^o, satisfy

$$e^o \in argmax \left\{ \frac{1}{N} \Pi \left[\frac{N}{2} e^y + e + \left(\frac{N}{2} - 1 \right) e^o \right] - v(e) \right\} \qquad \textbf{24}$$

and

$$\begin{aligned} &\frac{1}{N}\Pi\left[\frac{N}{2}e^y + \frac{N}{2}e^o\right] - v\left(e^y\right) + \delta\left\{\frac{1}{N}\Pi\left[\frac{N}{2}e^y + \frac{N}{2}e^o\right] - v\left(e^o\right)\right\} \geq \\ &\max_e\left\{\frac{1}{N}\Pi\left[\frac{N}{2}e^o + e + \left(\frac{N}{2} - 1\right)e^y\right] - v(e)\right\} + \delta\left[\pi(\hat{e}) - v(\hat{e})\right]. \end{aligned} \qquad \textbf{25}$$

Sketch of Proof: Equation (24) characterizes a senior partner's effort level. It shows that all senior partners will set their effort levels at the private noncooperative level, i.e., the myopic level.[21] Equation (25) characterizes a junior partner's lifetime utility. The left hand side is his utility if he does not cheat when young. The right hand side is his utility if he cheats when young. Under the trigger strategy described in Proposition 2 he gets the myopic payoff when old if he cheated when young. The two conditions merely indicate the conditions needed to ensure that neither junior nor senior partners have any incentive to defect. □

Notice that the implied payoff structure is that senior partners' payoffs are higher. Although all partners receive the same shares of profits, senior partners do not exert as much effort, therefore their payoffs are higher. In the case we have considered, it is irrelevant if productivity rises with seniority, since senior partners never work. If we consider more than two periods, however, these concerns may become relevant, and the optimal structure of payoffs could be quite different.

[21]This is just a restatement of Proposition 1 in this context.

B. THE MODEL WITH REPUTATION EFFECTS

We now vary the structure of the model from that of the previous section to allow profits to depend not just on current period effort, but on reputation as well. Some authors (e.g., Getzen, 1984) have asserted that reputation effects are quite important in medical group practices. We model reputation as a dynamic effect in profits, i.e., current period profits depend not only on this period's effort, but on last period's effort. This period's effort captures production, plus any contemporaneous demand effects (e.g., through production of quality). Last period's effort affects the group's reputation, which affects demand.

To keep things simple, suppose that profits, Π, are high, i.e., equal to Π^h, provided that group effort (defined as the sum), $E_t = \Sigma e_{it}$, is at least equal to some level E_1, and that the group's reputation, R_t, is good. Further, let the reputation in period t, R_t, be good if last period's total effort, E_{t-1}, was at least as great as some level, E_2, which exceeds E_1, $E_2 \geq E_1$. When only this period's effort is good ($E_t \geq E_1$), or only this period's reputation is good ($E_{t-1} \geq E_1$), suppose profit is strictly less than the high level of profits, $\Pi = \Pi^m < \Pi^h$. Thus, both effort devoted to production (current period effort) and good reputation (last period's effort) are required for profits to be high. Let the profit level when reputation does not matter and when a partner shirks be less than the profit level when reputation matters and a partner shirks, $\Pi^s < \Pi^m$. Also suppose that $v(0) = 0$. As before, let there be N partners who each have T period terms.

Compared to the case in which reputation does not matter, shirking is more likely. The costs of shirking may be lower. When reputation matters, its benefits can still be captured by shirking partners. When reputation doesn't matter, the dropoff in profits associated with shirking will likely be greater. In addition, when reputation matters, the costs of shirking are partially in the future, since current effort affects future reputation. If there is discounting, shirking may be more likely. We formalize this below.

Proposition 5: Consider two partnerships having the same maximum profit potential, Π^h. Suppose that in one partnership this level of profit can be realized when $E \geq E_2$. For this partnership the levels of effort which maximize output and which maximize reputation are identical. In the other partnership $E \geq E_1 \leq E_2$ maximizes output, and a good reputation requires that $E \geq E_2$ in the previous period. In this partnership it is not necessary to work as hard to get high output as it is to achieve a good reputation. Then equilibrium profits can be higher without the reputation effects.

Sketch of Proof: Look at maximum partnership profit without reputation. The equilibrium condition for the oldest working partner(s) (i.e., the second oldest partner(s)) not to shirk requires that the present value of their utility in the last two periods is greater than their share of the shirking level of profits in the current period (their payoff in the next period will be zero if they cheat and are expelled),

$$\frac{1}{N}\Pi^h - v(e) + \delta \frac{1}{N}\Pi^h \geq \frac{1}{N}\Pi^s . \qquad \textbf{26}$$

The partner's utility in the last period when he does not work is $(1/N)\Pi^h$, since disutility equals zero when effort equals zero. When reputation matters, the equivalent condition is more restrictive. The present value of the second oldest partner's utility must still exceed his payoff when he shirks,

$$\frac{1}{N}\Pi^h - v(e) + \delta \frac{1}{N}\Pi^h \geq \frac{1}{N}\Pi^m . \qquad \textbf{27}$$

In this case, however, the payoff is $(1/N)\Pi^m$, which strictly exceeds the shirking payoff without reputation,$(1/N)\Pi^s$. Thus, shirking is more likely with reputation effects. □

This result gives a flavor of what some of the effects of reputation might be. It implies that cooperation among all but the oldest partners will be more difficult to induce. Thus it may be true that there will be less efficiency when group reputation is important. HMOs, for example, typically have a "brand name" which supercedes that of any of their physicians. It may be that, *ceteris paribus*, HMOs are less efficient than fee-for-service groups, although empirically it may be a very difficult matter to hold all other things equal. We have not considered seniority in this model. It may be that seniority works differently when there is reputation, although since reputation is a lagged effect, the actions of the junior members assume even more prominence.

C. IMPERFECT OBSERVABILITY AND SCREENING

In this section we consider issues of imperfect observability. The previous sections assumed perfect information ex post, which is clearly not a realistic assumption. In particular, let us expand the analysis to allow for different types of members, so that screening becomes an issue.

Assume that there are differing types of junior members, of varying levels of productivity. Further, senior members can only observe output of other partners with noise. Although it may be desirable for productive physicians only to associate with other productive physicians, we speculate that a separating equilibrium in compensation may not exist.

If self-selection will not result in a separating equilibrium, groups may wish to screen out the unproductive junior partners ex post. We concentrate on screening here. This could require setting up payoffs such that unproductive junior partners will have incentives to defect all the time and productive junior partners will not.

This could explain "up or out" seniority systems in which some good junior partners are expelled in order to weed out bad ones. It could also explain the existence of different types of groups with different "cultures." When multiple equilibria exist, some partnerships will be more productive than others, and each will have a different "culture" characterized by a different reward scheme.

Now suppose that all agents have T period horizons. There are three types of agents: "unproductive," "moderately productive," and "productive." There are M agents of each type. Thus in each period, 3M associates/junior partners join each partnership. Let $x_{it} = \theta e_{it} = \varepsilon_{it}$ denote the observed input of an associate i in period t.

Suppose the utility that associate i receives in period t is

$$u_{it} = w_{it} - .5v_i e_{it}^2$$

where w_{it} is the wage the i^{th} associate receives in period t, v_i is a scalar (the coefficient on disutility), and $v_i \in \{v^l, v^m, v^h\}$. $v^h > v^m > v^l$, thus h, m, and l denote high, medium, and low disutility of effort, respectively. Thus h-types are "unproductive," m-types are "moderately productive," and l-types are "productive."

Suppose output of the partnership in period t is $F(X_t)$, where $X_t = \Sigma x_{it}$. Also, suppose x_{it} is observed after the wages are paid. Further, suppose that associates can be denied partnership after one or more periods, but no more than N periods. Note that senior partners will not work at any equilibrium. Further, suppose that the production function is such that $F(X) = 0$ whenever X is no larger than some value X_c , or that reputations can only be maintained by large groups,[22] so that no individual can receive any benefits, beyond reservation wages, outside a partnership.

<u>Conjecture:</u> If $F'' < 0$, i.e., there are decreasing returns to scale, a partnership maximizing per capita payoff will keep only type l's (i.e., the most productive types, since they have the lowest disutility of effort), and maximize $F(Me)/M - .5v^l e^2$.

For $F(\bullet)$ exhibiting some increasing returns, a group may wish to keep m-types and possibly h-types. If the wages for type l's and other types differ, then the senior partners might impose higher standards on l-types. For if (e^{*l}, e^{*m}) maximize $(1/2M)\{F(\theta(Me^l + Me^m) - M[.5v^l(e^l)^2] - M[.5v^m(e^m)^2]\}$, then at the optimum, $e^{*l}/e^{*m} = v^m/v^l > 1$. However, the partnership won't be able to distinguish between types unless there are several observations of outputs.

Further, only old partners shirk, thus there is an incentive to cull out unproductive workers. This can be done using a compensation scheme which provides standards that unproductive associates will have no incentive to try and meet.

This analysis indicates that groups will need to screen new members. The screening will

[22] Getzen (1984) argues in favor of this.

take the form of standards that junior associates need to maintain to make partner. These standards can be higher than those maintained by senior members, especially where the promotion serves an ability based screening role. Finally, greater economies of scale, which make larger partnerships profitable, may lead to relaxation of these standards in order to achieve the optimal number of members.

III. SUMMARY AND CONCLUSIONS

This paper represents a first attempt to analyze medical group practices as ongoing organizations. Since the majority of practicing physicians now practice in some kind of group setting, and since many groups have been in existence for some time, this seems appropriate. We have tried to show that the outcomes which result from a model of this sort are quite different from those which arise in a static setting. Near-efficient outcomes can result when there are overlapping generations of physician-members in medical groups. Near-efficiency is more likely the lower are physicians' discount rates, the less turnover there is in the group, and the more overlap there is in physicians' terms. These may be empirically testable propositions. Further, we have explored some aspects of medical groups as ongoing organizations, especially seniority and screening. There seem to be efficiency justifications for these institutions. It seems that the greater are economies of scale, the less intense should be screening, which may also be an empirically testable hypothesis.

REFERENCES

Becker, Gary S. (1993) Human Capital, Third Edition, Chicago: University of Chicago Press.

Benoit, Jean-Pierre and Vijay Krishna (1985) "Finitely Repeated Games," Econometrica, July, 17(4):905-922.

Cooper, Russell and Andrew Daughety (1993) "Cooperation in Infinite Horizon Games with Finitely Lived Players," Games and Economic Behavior.

Crémer, Jacques (1986) "Cooperation in Ongoing Organizations," Quarterly Journal of Economics, 101:33-49.

Fudenberg, Drew and Eric Maskin (1986) "The Folk Theorem in Repeated Games with Discounting and Incomplete Information," Econometrica, 54:533-554.

Gaynor, Martin (1989) "Competition Within the Firm: Theory Plus Some Evidence From Medical Group Practice," Rand Journal of Economics, Spring, 20(1):59-76.

Gaynor, Martin and Mark V. Pauly (1990) "Compensation and Productive Efficiency in Partnerships: Evidence from Medical Group Practice," Journal of Political Economy, 98(3):544-573.

Gaynor, Martin and Paul J. Gertler (forthcoming) "Moral Hazard and Risk Spreading in Medical Partnerships," Rand Journal of Economics.

Getzen, Thomas E. (1984) "A 'Brand Name Firm' Theory of Medical Group Practice," Journal of Industrial Economics, 33:199-215.

Gonzalez, Martin and David Emmons (1988) Socioeconomic Characteristics of Medical Practice in the U.S., Chicago: American Medical Association.

Harrington, Joseph E. (1987) "Collusion in Multiproduct Oligopoly Games under a Finite Horizon," International Economic Review, February, 28(1):1-14.

Harrington, Joseph E. (1992) "The Role of Party Reputation in the Formation of Policy," Journal of Public Economics, 49:107-121.

Harris, Christopher and Bengt Holmström (1982) "A Theory of Wage Dynamics," Review of Economic Studies.

Havlicek, Penny L., Eiler, Mary Ann, and Ondria T. Neblett (1992) Medical Groups in the U.S., 1993, Chicago: American Medical Association.

Held, Philip J. and Uwe E. Reinhardt (eds.) (1979) Analysis of Economic Performance in Medical Group Practices, Project Report 79-05, Princeton: Mathematica Policy Research, Inc.

Hillman, Alan L., Pauly, Mark V., and Joseph J. Kerstein (1989) "How do Financial Incentives Affect Physicians' Clinical Decisions and the Financial Performance of Health Maintenance Organizations?" New England Journal of Medicine, July 13, 321(2):86-92.

Jovanovic, Boyan (1979) "Job Matching and the Theory of Turnover," Journal of Political Economy, 87: 972-990.

Kandori, Michihiro (1992) "Repeated Games Played by Overlapping Generations of Players," Review of Economic Studies, 59(1): 81-92.

Kralewski, John E., Pitt, Laura, and Deborah Shatin (1985) "Structural Characteristics of Medical Group Practices," Administrative Science Quarterly, 30:34-45.

Lazear, Edward (1979) "Why is There Mandatory Retirement?" Journal of Political Economy, 87: 1261-1284..

Lee, Robert H. (1990) "The Economics of Group Practice: A Reassessment," in Advances in Health Economics and Health Services Research Richard Scheffler and Louis Rossiter (eds.) Greenwich, CT: JAI Press.

Newhouse, Joseph P. (1974) "The Economics of Group Practice," Journal of Human Resources, Winter, 37-56.

Radner, Roy (1985) "Repeated Principal Agent Games with Discounting," Econometrica, 53:1173-1198.

Radner, Roy (1986) "Repeated Partnership Games with Imperfect Monitoring and No Discounting," Review of Economic Studies, 53:43-58.

Radner, Roy, Myerson, Roger, and Eric Maskin (1986) "An Example of a Repeated Partnership Game with Discounting and Uniformly Inefficient Equilibria," Review of Economic Studies 53:59-70.

Reinhardt, Uwe E., Pauly, Mark V., and Philip J. Held "Financial Incentives and the Economic Performance of Group Practitioners,"in Held and Reinhardt (eds.), *op cit.*

Salant, David S. (1991) "A Repeated Game with Finitely Lived Overlapping Generations of Players," Games and Economic Behavior, 3:244-259.

Salant, David S. (1992) "Behind the Revolving Door: A New View of Public Utility Regulation," unpublished manuscript, GTE Laboratories.

Salop, Joanne and Steven Salop (1976) "Self-Selection and Turnover in the Labor Market," Quarterly Journal of Economics, 90: 619-627.

Shepsle, Kenneth, and Barry Nalebuff (1990) "The Commitment to Seniority in Self-Governing Groups," Journal of Law, Economics, and Organization, 6, Special Issue, 45-72.

Smith, Lones (1992) "Folk Theorems in Overlapping Generations Games," Games and Economic Behavior, July 57(2): 426-449.