Structuring Incentives within Organizations: The Case of Accountable Care Organizations

by

Brigham Frandsen,
Brigham Young University

and

James B. Rebitzer,
Boston University School of Management and
NBER, IZA, The Levy Economic Institute.

Corresponding author:
James B. Rebitzer
Boston University School of Management
595 Commonwealth Avenue
Boston, MA 02130
rebitzer@bu.edu
617 3837356

Abstract:

Accountable Care Organizations (ACOs) are a new model for integrated health care delivery created by the Patient Protection and Affordable Care Act. They allow a group of hospitals and providers to jointly contract with the Center for Medicare and Medicaid Services to provide care to a population of Medicare enrollees in an environment that rewards cost-efficiency through global budgeting and group-based pay-for-performance incentives. To improve the precision of performance measures, ACOs are required to have at least 5,000 enrollees. Achieving this scale requires pooling the patient panels of many physicians and this causes a free-riding problem. Working with a stylized model of ACO incentives, we establish that the negative effects of free-riding swamp the positive effects of increased precision. We calibrate the model using proprietary performance measures from a very large insurer. Our estimates suggest that even minimally sized ACOs will require unmanageably high stakes incentives. To achieve their goals ACOs will have to augment under-powered incentives with motivational strategies that complement pay-for-performance. Some of these alternative strategies can only be implemented in conventional integrated organizations, while others may prove workable in hybrid organizational forms that are more congruent with practice patterns in regions where care delivery is currently highly fragmented.

Introduction

Economists and others concerned about inefficiencies in the U.S. health care delivery system frequently worry about the fragmented structure of physician practices. Fragmented care delivered by physicians working as independent owners of small practices is, so the story goes, incapable of matching the high quality or low cost of care delivered by large integrated systems. The inefficient fragmented providers are not driven out of the market, however, because they operate in a largely “fee-for-service” payment environment that does not measure or reward cost-efficient delivery of health care services.

The view that fragmented care delivery has both lower quality and higher cost raises two fundamental questions for transaction cost economics. First, what is it about integration within an organization that enables superior performance? Secondly, do these advantages accrue only to traditional hierarchical organizations that own hospitals and clinics and hire physicians as employees? Alternatively, might the advantages of integration also accrue to hybrid forms that more closely resemble the organizational environment in which health care is currently delivered in much of the United States?

We explore these questions in the context of Accountable Care Organizations (ACOs). ACOs are a new model for integrated health care delivery created by the Obama administration’s Patient Protection and Affordable Care Act. ACOs are designed to promote the benefits of integrated care by allowing a network of hospitals and providers to jointly contract with the Center for Medicare and Medicaid Services (CMS) to provide care to a population of Medicare patients in an environment that rewards cost efficiency. The key feature of these contracts is the use of global budgeting to contain costs combined with incentives to maintaining care quality at or above acceptable levels.

In health care settings, there is a very compelling reason to aggregate incentives within organizations: quality measures are typically quite noisy and averaging measured performance across the members of an organization improves precision. A recent study finds, for example, that primary care physicians had annual median caseloads of 260 Medicare patients (Nyweide, Weeks et al. 2009). Of these, 25 were women eligible for mammography and 30 had diabetes**.** With such low numbers, individual primary care physician practices simply do not have a sufficient caseload to reliably detect, say, a 10 percent improvement in the rate of use of relevant preventative care measures such as routine breast exams and monitoring hemoglobin levels in blood. If real improvements in quality cannot be distinguished from changes due to random chance, rewarding performance measures comes uncomfortably close to rewarding luck. On this basis, Fisher, McClellan et al. (2009) argue that ACOs require a minimum of 5,000 beneficiaries for performance measures to have sufficient power to meaningfully identify performance improvements, a minimum size requirement that CMS has since adopted.

From an economic perspective this statistical approach to determining the minimum scale of an organization is incomplete. According to the canonical principal agent model, improving the precision of performance measures does indeed enhance the efficiency of incentive pay arrangements, but this gain comes at a cost. Increasing the size of patient populations necessarily requires bringing more physicians into the ACO. As the number of physicians grows, the effect of any physician’s action on the organization's overall performance is diminished and so incentives are diluted. On this basis one might expect that there exists some optimal ACO size that balances the marginal costs from free-riding against the marginal benefits of enhanced precision. Surprisingly we find that this is not the case. Increasing the size of ACOs simply makes the incentive problem more severe.

We establish this result using a stylized model of physician incentives under fee-for-service and ACO-style incentives. Our approach builds upon conventional principal-agent models, but is unusual in that it focuses on the sort of nonlinear incentives commonly used in health care generally and in the ACO program in particular – global budgets to encourage cost-efficient practice styles with bonuses linked to performance measures to prevent a decline in the quality of care. In this setting we solve for optimal values for the organization's two incentive instruments: bonus size and threshold performance levels.

Our analysis yields the following results. First, Medicare's traditional fee-for-service payment system discourages cost effective medical practices. Second, ACO-style global budgets produce cost-effective treatment practices but may also reduce care quality. This degradation in quality can be offset by linking bonuses to quality metrics – even when the measures of quality are noisy. Our third result is that as size increases ACO-style cost and quality incentives have to employ larger bonuses *and* higher threshold levels of performance: indicating that the motivation problem increases with organizational size.

With these results in hand, we then consider how a payer would set bonus pay and threshold performance levels in order to achieve cost and quality goals. We calibrate our incentive model using confidential claims data and quality measures from a very large sample of chronically ill patients. Our results indicate that achieving even modest quality targets under ACO-style incentives may require "high stakes" incentives with very large bonuses and very strict minimum performance thresholds.

These results suggest that implementing an ACO-like payment model may be not be manageable for organizations relying solely on financial incentives. We discuss three implementation challenges. First, it is often possible for provider organizations to take actions that improve measured performance without creating any clinical or economic value. The incentives to "game" the incentive system in this way are greatest under high stakes incentives, i.e. when bonuses and minimum performance thresholds are both very high. The second implementation problem is that "high-stakes" incentives linked to performance measures can discourage innovation if these innovations involve trials and ramp-up periods that cause even temporary declines in measured quality. The third implementation problem created by "high stakes" incentives is that they require organizations to bear great financial risk, and the amount of risk *increases* very dramatically as organizations increase in size. In the language of transaction cost economics, the three implementation issues we identify create substantial "costs of integration" for ACOs.

Gibbons (2010) in his overview of the field, argues that transaction cost economics has done a good job identifying the costs of transacting within a market, but has not yet provided a satisfactory account of variations in the cost of transacting within an organization.[1] Our analysis suggests that ACOs may be an interesting laboratory for examining these integration costs. We identify a number of factors that determine the cost of integration for ACOs. Of these, two seem particularly important: the proportion of Medicare enrollees in a physician's patient panel; and the ability of an organization to stimulate alternative motivators that can complement under-powered incentives to achieve high performance levels. Since ACOs can be constituted from either conventional or hybrid organizational forms, the latter factor raises the question of which arrangement better stimulates these complementary motivators. The former factor implies that successful ACOs will likely require a narrow physician network in which Medicare constitutes a large proportion of a physician's patient panel – a result of some importance for health care policy.

[1] Masten, Meehan et al. (1991) identify a similar gap in the literature.

The paper proceeds in four sections. Section one briefly introduces relevant institutional background on health care fragmentation and the structure of accountable care organizations. Section two develops our model of incentive pay and compares cost and quality outcomes in stylized fee-for service and ACO settings. Section three presents the results of our calibration exercise. Section four considers strategies to mitigate the integration costs resulting from high-stakes incentives within organizations.

I. Fragmented Care Delivery and the ACO as Policy Response[2]

Health services researchers have long argued that a central problem with health care delivery in the US is fragmentation (Cebul, Rebitzer et al. 2008). Individual patients are frequently treated by numerous care providers who have only weak organizational ties with one another and often little expertise in coordinating care. This results in poor information flows, heightened error rates and inadequate care coordination – problems that are especially troublesome for the management of patients with costly chronic diseases. The obvious fix, according to this view, is for physicians to join large integrated care delivery systems. Yet as late as 2001, 60 percent of physicians worked either in solo practice or in groups of 2 to 4 physicians and only 7 percent worked in groups with 50 or more physicians. In that same year, more than 65 percent of physicians were self-employed and only 35 percent were employees. Why, given their purported efficiency advantages, don't we see more physicians going to work for large integrated care organizations?

Surprisingly little research has been devoted to this important question, but conventional wisdom is that the answer lies in the ways health care services are financed and purchased.[3]

[2] Much of this section is adapted from Rebitzer and Votruba's (2011) review of the organizational economics of physician's practices.
[3] In contrast, there has been a large literature documenting the productive and allocative inefficiencies in our care delivery systems. For an incisive review see (Baicker and Chandra 2011).

More specifically, restrictions on Medicare's purchasing policies combined with coordination failures between non-governmental buyers and providers prevent the emergence of efficient integrated care delivery organizations.

Medicare, the largest single buyer of medical services, is locked by rules and legislation into a fee-for-service payment system and it is not allowed to selectively contract with more efficient physician groups. Compounding the problem is the fact that the Medicare's regulatory boards charged with evaluating new technologies are concerned primarily with whether new drugs or procedures offer positive benefits rather than whether they are cost-effective (Baicker and Chandra 2011). The failure to consider cost-effectiveness likely has system-wide repercussions because commercial health insurance plans are heavily influenced by Medicare coverage decisions (Baicker and Chandra 2011).

If Medicare is hamstrung by regulations, the private sector is constrained by different considerations. Many employers who purchase insurance on behalf of their employees are not interested in or capable of evaluating the cost-effectiveness of the care their employees receive.[4] Sophisticated employers (typically large, self-insured companies) would like to reward high efficiency providers but their desires are thwarted by a thorny coordination problem. Suppose that the full efficiency gains of integrated care delivery can only be realized under bundled prospective payment systems (Crosson 2009). In communities with highly fragmented care delivery, it is hard to find providers with the capacity to succeed under such a payment system. As a result payers don't innovate away from the status quo fee-for-service payment system and

[4] In a case study of Geisinger's Provencare program, Clark and Rosenthal quote the results of conversations between Geisinger and the employers who buy their insurance. "We went with the health plan leadership and talked to a number of employers. We told them that we would guarantee delivery of the best care and that we wouldn't submit a bill otherwise. The employers didn't want any of that. Their eyes glazed over. They said, 'As far as we know, we're already buying best practices. The evidence we really care about is whether or not the patients need the procedure in the first place. In addition, we don't like all of the unpredictability in costs that you get with each patient. Give us one price per procedure and you worry about all the other stuff" (Clark and Rosenthal 2008) p. 8.

there is little competitive advantage for providers to move out of their currently fragmented delivery organizations.

Accountable Care Organizations are designed to overcome these impediments to payment reform.[5] First, and perhaps most important, ACOs offer a means by which Medicare can break away from traditional fee-for-service reimbursements and reward efficient providers. As a legal entity, ACOs are comprised of a network of hospitals and providers that contract with the Center for Medicare and Medicaid Services (CMS) to provide care to a large bloc of Medicare patients (5,000 or more). The contracts, which last for three years, create a single risk-bearing entity with incentives to control costs.[6] ACOs that come in under their specified cost benchmarks earn a fraction of the savings. In order to receive these payments the ACO must also clear stringent threshold quality levels on a number of indicators that reflect patient and caregiver experience, care coordination, patient safety, preventative health, and health of at-risk frail and elderly populations (Ginsburg 2011).

The goal of this incentive system is to reward efficient providers without sacrificing quality. By encouraging the formation of large provider organizations de-novo, CMS may also overcome the coordination failures that have prevented sophisticated private buyers from reforming their own payment practices. Indeed there is nothing stopping ACOs that contract with Medicare from also contracting with private payers. The prospect of emerging integrated

---

[5] Although ACOs are only a small part of a huge piece of legislation, they have attracted a great deal of attention from policy-makers, physicians and managers. As of October 2012, there were a total of 318 ACOs in 48 States. Medicare ACOs cover 2.4 million beneficiaries in 40 states plus Washington DC. (Meyer 2012). As an indication of the interest in ACOs, consider the following incomplete list of relatively recent articles in such leading journals as the *New England Journal of Medicine*, *The Journal of the American Medical Association* and *Health Affairs*: Burns and Pauly (2012), Crosson (2009), Crosson (2011), Ginsburg (2011), Meyer (2012), Shields, Patel et al. (2011), Shortell and Casalino (2010), Singer and Shortell (2011), Zirui Song, Safran et al. (2011).

[6] The exact nature of the payments to ACOs varies a good deal. All ACOs accept a global budget for taking care of defined groups of patients. If they meet performance standards, they share in any cost savings they achieve; in some cases, they also may share losses they incur. Medicare ACOs are being paid on a fee-for-service basis rather than on a per member per month basis- further clouding the picture (Meyer 2012).

delivery organizations may already be moving savvy insurance companies to rethink their payment policies. Song, Safran et al. (2011) analyze the effects of a recently introduced global double-sided payment incentive system implemented by Blue Cross Blue Shield of Massachusetts. The contract was similar in many respects to the shared savings program for Medicare but instead of Medicare patients it was implemented for HMO and point of service commercial populations.

From the perspective of organizational and transaction cost economics, ACOs have a number of novel features. ACOs cannot restrict their members to a specific network of physicians and there is nothing in the legislation requiring that ACOs be constituted as a traditional organization in which doctors are either employees or owners of a risk-bearing entity that also owns the relevant capital equipment. Indeed advocates who favor ACOs as a means of promoting integrated care systems see them emerging from five different practice arrangements: integrated delivery systems that combine insurance, hospitals, and physicians; multi-specialty group practices; physician hospital organizations; independent practice associations, and virtual physician organizations (Shortell, Casalino et al. 2010). As we shall see, the transformation of hybrid and virtual physician organizations into ACOs poses special problems and opportunities for incentive design. [7]

The ACO model may also transform the link between primary care and specialized care. ACOs may be able to improve their bottom line by introducing training and computer-assisted decision support that make it easier for generalists to substitute their own decisions for those of specialists. It may, for example, be efficient to train primary care physicians to treat rashes and acne rather than sending every case of rash or acne to a dermatologist. On the other hand, the

[7] Meyer (2012) reports that of the 114 provider groups in Medicare Shared Savings ACO Program, nearly half are physician driven organizations serving fewer than 10,000 beneficiaries. In addition 32 larger provider groups with experience in coordinated care started Medicare Pioneer ACOs.

vast explosion in medical knowledge implies that there are limits to the substitution of generalist for specialist care (Becker and Murphy 1992). In this case, Garicano and Santos's (2004) analysis suggests that efficiently managing referrals to specialists will likely entail bringing some specialists into the ACO. Keeping these specialists fully occupied may also exert upward pressure on the scale of ACOs.

II. Modeling Incentives in Fee-for-service and ACO Environments

In this section we present a simple model of physician incentives in the traditional Medicare fee-for-service environment and in an ACO-like environment. Our intention is not to build a model that closely resembles the very complicated real-world payment structures. Rather we present a highly stylized model that incorporates a few salient features of actual payment systems in order to establish three points. The first is that the incentives in fee-for-service payment systems do not deliver cost-effective health care. The second is that the prospective payment systems in ACOs can induce physicians to deliver cost-efficient care, but quality will suffer unless some effort is made to introduce quality incentives.

These points are not controversial but establishing them formally enables us to make a third point that is less obvious and more central to our story: increasing the size of ACOs makes incentive problems *more* severe. More specifically, as group size increases the ACO will choose to increase both bonus pay and threshold levels of performance. As we shall see, this result is due to the fact that under the appropriate scaling the precision of performance measures increases by a factor of the square root of group size while free-riding problems increase by a factor of group size.

The Fee-For Service Baseline:

Underlying our model is the observation that treatments with heterogeneous effects on patients are an important driver of health care costs (Chandra and Skinner 2012). Cardiac stents, for example, have a powerful and positive effect on patients who have experienced a very recent heart attack, but do not offer much better outcomes for patients with stable angina than conventional medical management. The key to cost-effective medicine is for physicians to identify which patient type they are treating and to withhold the expensive, but relatively inefficient care, from the patients who do not much benefit from it. Identifying patient type often requires time and effort from a physician and is complicated further by the fact that the boundary between patient types may be fuzzy. In addition, physicians who have already made investments in perfecting different treatment protocols may optimally choose different cutoffs to distinguish effective from ineffective treatments (Chandra and Staiger 2007).

We take this heterogeneity in treatment effects as the starting point for our stylized representation of the problem of delivering cost-effective care under fee-for-service. Patients in our model respond to treatments differently according to their (initially unknown) type $\theta \in \{0,1\}$ with $\Pr\{\theta = 0\} = p < 1$. When treating a patient, a provider first chooses $R \in \{0,1\}$, whether or not to undertake activity to reveal the patient's type. Revealing the patient's type (action $R = 1$) requires costly effort, r. Whether or not the provider chooses to exert this effort, she must select a mode of care $Z \in \{0,1\}$, and a treatment effort level, $S \in [0, \infty)$. In our notation, $Z = 1$ corresponds to care that is complicated and costly, but is effective for all patients. $Z = 0$ corresponds to care that is simple and low cost, but is effective for only patients with $\theta = 0$. The variable S represents effort that improves the quality of care received by patients, regardless of the mode chosen. Health outcomes depend on patient type, mode of care, and effort level:

$$h(\theta, Z, S) = \big(1 - \theta(1 - Z)\big)g(S),$$

Function, $\boldsymbol{g}(\boldsymbol{S})$ determines the "quality" of care delivered, and it assumed to be an increasing and concave function of physician effort.

The marginal cost of care, $c(Z)$, is incurred by the provider, and depends on the mode of care, with $c(1) = c(0) + q$, where $q \geq r/p$, so it is socially efficient to reveal θ. Provider reimbursement is on a fee-for-service basis, so the provider's (gross) income depends on the mode of care and is equal to the cost: $w(Z) = c(Z) + v$, where v is the physician's net compensation for her services.

Let $y \equiv w(Z) - c(Z)$ be a provider's net income and let e be total effort cost, which is equal to $S + rR$. It is analytically convenient and consistent with our setting, to assume that providers are risk neutral and so maximize $y - e$, subject to $e \geq e_0$ and $h \geq h_0 > 0$.[8] The terms e_0 and h_0 respectively capture intrinsic motivation providers have to devote effort to care quality and to realizing patient health outcomes. Assume $g(0) \leq h_0 \leq g(e_0)$ and $r \leq e_0$.

Our first result is that under this fee-for-service system physicians do not provide cost-effective care. Since neither patient health outcomes nor the provider's net income is improved by choosing the cost-effective mode of care, a utility-maximizing provider will choose action $R = 0$ (that is, she will not exert effort to reveal è), treatment mode $Z = 1$ for all patients, and treatment effort $s^{FFS} = e_0$ yielding quality outcomes $g(e_0)$. If providers earn no rents, then $v = e_0$ and the expected cost of care under fee-for-service is $c(0) + q + e_0$. The cost-inefficiency of the fee-for-service system is captured by the additional cost q of expensive care for the fraction p of patients who do not benefit from it.

[8] Principal agent models typically analyze risk-averse agents working under a linear incentive contract. Risk neutrality is a reasonable starting point as Medicare enrollees are only a fraction of the patient panel of most primary care physicians. This fragmented financing means that incentive pay systems that put high stakes at risk for the ACO as an entity may put a much smaller fraction of an individual physician's compensation at risk.

Payments Under a Global Budget:

ACOs accept a global budget for taking care of a defined group of patients. This budget offers a mechanism for providers to internalize the costs of the mode of care they provide. Here, providers are no longer paid $w(Z) = c(Z)$ for treating patients, but instead receive a capitated payment equal to the expected cost of cost-effective care for patients

$$w = E[c(\theta)] + v = c(0) + (1 - p)q + v.$$

A provider's net income, $w - c(Z)$, now depends on the mode of treatment, and the provider will therefore find it worthwhile to choose $R = 1$ to reveal when the low-cost mode of treatment is appropriate. The provider will choose treatment mode $Z = \theta$, and exert treatment effort $s^{CPS} = e_0 - r$ if $g(e_0 - r) \geq h_0$ and $s^{CPS} = g^{-1}(h_0)$ otherwise. Patient health outcomes are equal to $g(s^{CPS}) \leq g(e_0)$. Assume for simplicity that $g(e_0 - r) \geq h_0$ so $s^{CPS} = e_0 - r$. If providers earn no rents, then $v = e_0$ and the expected cost of care is $c(0) + (1 - p)q + e_0$.

Thus we have arrived at our second result. The capitated payment system has induced the provider to give cost-effective care, for a savings of pq, but at the expense of less effort devoted to care quality. In order to induce cost-effective care without sacrificing quality, prospective payment systems must rely on quality incentives that can sustain higher levels of treatment effort. We now turn our attention to these quality incentives.

Group Incentives for Quality:

The model we develop in this section builds upon two essential features of ACO incentives— noisy care quality measures, and bonuses paid to groups of physicians contingent on aggregate targets—while abstracting from other complexities of actual ACO incentives. For example, our model posits that Medicare pays a fixed payment per enrollee, when the actual "per-member" payment is more complicated.[9] To underscore the stylized nature of our model of group quality incentives, in what follows we refer not to ACOs, Medicare and physicians, but instead to groups, principals and agents.

In order to maintain quality of care, suppose the designer of the incentive scheme (i.e., the "principal") structures payments to agents that depend on effort devoted to treatment. If effort were observable and contractible, this would be straightforward to implement. Suppose, however, only a noisy signal of effort is observed:

$$x_j = e_j + \epsilon_j$$

where $e_i$is the treatment effort exerted by provider i and ϵ_i is a random variable with mean zero and variance σ^2. We consider a compensation scheme where the signals are aggregated over a group of N providers:

$$\tilde{x}_N = \sum_{i=1}^{N} x_i \,/N$$

and compensation is conditioned on $\tilde{x}_N$.[10] The situation where the principal contracts with each agent individually is the special case where $N = 1$.

Adapting Ritter and Taylor (2011), and Rebitzer and Taylor (2010), we consider a simple scheme where the principal pays a fixed baseline wage w_0 to each agent, and in addition pays a bonus b to each agent if $\tilde{x}_N$ exceeds some cut-off $\bar{x}$. Linking payments to threshold levels of performance is common in many pay-for-performance contracts in health care settings – see for

[9] Meye**r (2012)** briefly describes a number of different compensation set-ups in Medicare's ACO program. In the Shared Savings program, ACOs receive bonuses if they achieve cost and quality targets. In the future, Shared Savings ACOs will have to accept "two sided risk" and pay CMS back if they exceed spending targets. The Shared Savings program also includes an Advance Payment ACO model in which smaller groups receive their potential savings up front to help them fund infrastructure costs. "Pioneer" ACOs are formed from large provider groups with more experience in coordinated care. These ACOs currently accept "two sided risk" and they must show that at least half of their revenues in the near future will come from similar contracts with other payers.

[10] To simplify the exposition, we are assuming that all physicians have the same number of patients and we normalize that number by setting it equal to 1. This assumption has no effect on the results in this section.

example Gaynor, Rebitzer et al. (2004). Assume that risk neutral providers will participate in these schemes only if their expected net income at least equals their effort cost, as in the fee-for-service regime. We want to find the optimal payment policy $\left(w_0^*(N), b^*(N), \bar{x}^*(N)\right)$ and, as our notation indicates, the optimal policy depends on, N, the number of agents in the group.

We develop our model in conventional fashion by first deriving each agent's best response to incentive pay. The best effort response by agent i to the incentive contract is $\hat{e}_i$. It maximizes:

$$bPr(\tilde{x}_N > x) - e_i = b \Pr\left(\frac{\sum_{i=1}^{N}(e_i+\epsilon_i)}{N} > \bar{x}\right) - e_i$$

This function can be rewritten

$$= b \Pr\left(\underbrace{\frac{1}{\sigma}\sum_{i=1}^{N}\epsilon_i / \sqrt{N}}_{\xrightarrow{d} N(0,1)} > \frac{1}{\sigma}\left(\sqrt{N}\bar{x} - e_i/\sqrt{N} - \sum_{-i} e_{-i} / \sqrt{N}\right)\right) - e_i$$

.

Using the central limit theorem we treat the term on the left side of the inequality as normally distributed and so can rewrite the agent's objective function as

$$\xrightarrow{p} b\left(1 - \Phi\left(\frac{1}{\sigma}\left(\sqrt{N}\bar{x} - e_i/\sqrt{N} - \sum_{-i} e_{-i} / \sqrt{N}\right)\right)\right) - e_i.$$

When the team contains more than one member, the first order condition for the best effort response of individual i, $\hat{e}_i(b, N, e_{-i})$, given the effort levels chosen by other team members e_{-i} is

$$b\frac{\phi\left(\frac{1}{\sigma}\left(\sqrt{N}\bar{x}-\hat{e}_i/\sqrt{N}-\sum_{-i}e_{-i}/\sqrt{N}\right)\right)}{\sigma\sqrt{N}}=1$$

We consider a symmetric equilibrium where all providers in the ACO choose the same effort level, $\hat{e}$, which is characterized by

(1)
$$b\frac{\phi\left(\frac{\sqrt{N}}{\sigma}(\bar{x}-\hat{e})\right)}{\sigma\sqrt{N}}=1.$$

The second order condition for the reaction function is

(2)
$$b\phi'\left(\frac{\sqrt{N}}{\sigma}(\bar{x}-\hat{e})\right)>0$$

which implies $\bar{x}<\hat{e}$ for positive b. This condition has an intuitive interpretation that will be useful for what follows. In our set-up, the marginal cost to agents of additional exertion is the same at every effort level, but the expected benefit of the marginal effort varies with the threshold performance level. If $\bar{x}$ far exceeds current effort level, the expected benefit of additional exertion is close to 0 – only a very rare draw would enable the agent to clear the threshold. The expected marginal benefit of effort increases as $\hat{e}$ approaches $\bar{x}$ and at $\bar{x}=\hat{e}$ the expected marginal benefit of additional effort is at its maximum and diminishes thereafter. Thus if the agent will choose to exert any effort, she will exert at least $\bar{x}$. An implication of this condition is that threshold levels of performance will always be set so that the probability of success exceeds 0.5.

Together with the participation constraint

(3)
$$w_0+b\left(1-\Phi\left(\frac{1}{\sigma}\left(\sqrt{N}(\bar{x}-\hat{e})\right)\right)\right)>\hat{e}+r,$$

and the incentive compatibility constraints

$$(4) \qquad b\frac{\phi\left(\frac{\sqrt{N}}{\sigma}(\bar{x}-\hat{e})\right)}{ó\sqrt{N}} = 1,$$

$$(5) \qquad \Phi\left(\frac{\sqrt{N}}{\sigma}\left(\bar{x}-\hat{e}\left(\frac{n-1}{n}\right)\right)\right) - \Phi\left(\frac{\sqrt{N}}{\sigma}(\bar{x}-\hat{e})\right) > \frac{\hat{e}}{b},$$

we can derive the ACO's optimal bonus payment.

If the principal wishes to induce effort $e > e_0 - r$, equation (1) requires that he set b as:

$$(6) \qquad b^*(e, N, \bar{x}) = \frac{\sigma\sqrt{N}}{\phi\left(\sqrt{N}\frac{(\bar{x}-e)}{\sigma}\right)}.$$

From this expression it is easy to show that $\partial b^*/\partial \bar{x} < 0$. If the threshold is far below any desired effort level, e, the agent will have a good chance to clear the performance cutoff by providing effort far below e. Inducing e will then require a very high bonus. Conversely, as the threshold increases towards e, the probability of success with effort less than e decreases and so smaller bonuses are sufficient to motivate agents.

The base component of pay would then be set to satisfy the participation constraint:

$$(7) \qquad w_0^*(e, N, \bar{x}) = c(0) + (1-p)q + e + r - b^*(e, N, \bar{x})\left(1 - \Phi\left(\frac{1}{\sigma}\left(\sqrt{N}(\bar{x}-e)\right)\right)\right).$$

So far the performance cutoff has been left unspecified other than $\bar{x} < e$. To pin down a value of $\bar{x}$ we assume that the organization wishes to operate with the lowest feasible level of bonuses.[11] This implies that the firm chooses the largest value of $\bar{x}$ consistent with the incentive compatibility constraint (5).

With this set up we now proceed to establish our third result: that optimal bonus pay, b, and optimal threshold levels both are increasing in the number of group members, N. The optimal bonus increases with N:

[11] Firms may prefer the lowest feasible bonus levels because of agent's limited liability constraints or because of psychological factors that cause agents to respond best to incentives that offer the highest chance of success.

$$\frac{db^*}{dN} = \frac{\partial b^*}{\partial N} + \frac{\partial b^*}{\partial \bar{x}}\frac{d\bar{x}}{dN}$$

$$= \frac{\sigma\left[\Delta(\sigma,e,N)\frac{\sigma^2}{e(e-\bar{x}^*)} + \phi\left(\sqrt{N}\frac{\bar{x}^*-e}{\sigma} + \frac{e}{\sigma\sqrt{N}}\right)\right]}{2\sqrt{N}\phi\left(\sqrt{N}\frac{\bar{x}^*-e}{\sigma}\right)\left[\Delta(\sigma,e,N)\frac{\sigma^2}{e(e-\bar{x}^*)} + \phi\left(\sqrt{N}\frac{\bar{x}^*-e}{\sigma}\right)\right]}$$

The first partial derivative in this expression can be written:

(8) $$\frac{\partial b^*}{\partial N} = \frac{1}{2}\left(\frac{\sigma}{\sqrt{N}\phi\left(\sqrt{N}\frac{(\bar{x}-e)}{\sigma}\right)} + \frac{\phi'\left(\sqrt{N}\frac{(\bar{x}-e)}{\sigma}\right)}{\phi\left(\sqrt{N}\frac{(\bar{x}-e)}{\sigma}\right)^2}(e-\bar{x})\right) > 0.$$

We have already observed that $\partial b^*/\partial \bar{x} < 0$. Turning to $\bar{x}$, the bonus-minimizing cut-off is the solution $\bar{x}^*(e,N) < e$ to the following:

(9) $$\Phi\left(\sqrt{N}\frac{(\bar{x}-e)}{\sigma} + \frac{e}{\sigma\sqrt{N}}\right) - \Phi\left(\frac{\sqrt{N}}{\sigma}(\bar{x}-e)\right) = \frac{e}{\sigma\sqrt{N}}\phi\left(\sqrt{N}\frac{(\bar{x}-e)}{\sigma}\right).$$

Differentiating this equation, we find that the optimal cut-off increases with N:

(10) $$\frac{d\bar{x}}{dN} = \frac{1}{2N}\frac{\overbrace{\left(\phi\left(\frac{\sqrt{N}}{\sigma}(\bar{x}-e)\right) - \phi\left(\sqrt{N}\frac{(\bar{x}-e)}{\sigma} + \frac{e}{\sigma\sqrt{N}}\right)\right)}^{\Delta(\sigma,e,N)>0}\left(e-\bar{x}+\frac{e}{N}\right) + e\frac{(e-\bar{x})^2}{\sigma^2}\phi\left(\frac{\sqrt{N}}{\sigma}(\bar{x}-e)\right)}{\underbrace{\phi\left(\frac{\sqrt{N}}{\sigma}(\bar{x}-e)\right) - \phi\left(\sqrt{N}\frac{(\bar{x}-e)}{\sigma} + \frac{e}{\sigma\sqrt{N}}\right)}_{\Delta(\sigma,e,N)>0} + \frac{e(e-\bar{x})}{\sigma^2}\phi\left(\sqrt{N}\frac{(\bar{x}-e)}{\sigma}\right)} > 0.$$

Combining these results we also can show that b increases in group size, N:

(11) $$\frac{db^*}{dN} = \frac{\partial b^*}{\partial N} + \frac{\partial b^*}{\partial \bar{x}}\frac{d\bar{x}}{dN} = \frac{\sigma\left[\Delta(\sigma,e,N)\frac{\sigma^2}{e(e-\bar{x}^*)} + \phi\left(\sqrt{N}\frac{(\bar{x}^*-e)}{\sigma} + \frac{e}{\sigma\sqrt{N}}\right)\right]}{2\sqrt{N}\phi\left(\sqrt{N}\frac{(\bar{x}^*-e)}{\sigma}\right)\left[\Delta(\sigma,e,N)\frac{\sigma^2}{e(e-\bar{x}^*)} + \phi\left(\sqrt{N}\frac{(\bar{x}^*-e)}{\sigma}\right)\right]} > 0$$

where $\Delta(\sigma, e, N) \equiv \phi\left(\frac{\sqrt{N}}{\sigma}(\bar{x}^* - e)\right) - \phi\left(\sqrt{N}\frac{(\bar{x}^*-e)}{\sigma} + \frac{e}{\sigma\sqrt{N}}\right) > 0.$

III. Calibrating the Model

In this section we calibrate our incentive model and use the result to ask how high the stakes must be in a workable ACO-style group pay-for-performance system. Specifically we ask: how would the payer in an ACO-style payment system set bonus, b, and threshold performance, $\bar{x}$, in order to achieve the level of quality prevailing under fee-for-service, e_0?

The goal of the calibration in this setting is to determine the value of the key input to our model, σ, the standard deviation of the disturbance term in the money denominated performance measure. We do this in two steps. First, we obtain empirical estimates of the mean, u_x, and standard deviation, σ_x, of an actual clinical performance measure. The second step is to transform σ_x into σ.

Estimating Mean and Standard Deviation of Observed Quality

Our observed quality measure is derived from confidential insurance records on roughly a million chronically ill, commercial insurance members with health insurance from "fully insured" employers.[12] These data are well suited for this exercise in that the insurer combines billing records with data from pharmacies and labs to construct an ersatz electronic medical record for each patient. These records are then passed through a sophisticated artificial intelligence program to develop a quality measure which we label *Potential Gaps in Care*. The

[12] Fully insured employers are those who do not self-insure. They tend to be smaller employers who are less sophisticated in managing health insurance and associated costs.

adjective "potential" emphasizes that these are, in fact, noisy indicators of actual gaps in care. An illustrative issue identified by the system might be that the patient is a good candidate for an Ace inhibitor but there is no evidence that a prescription for the drug has been filled (the full list of targeted issues is provided in an appendix). This measured outcome could reflect a true gap in care arising from physician oversight. Alternatively, it might be a data error or it may reflect the patient's failure to fill the issued script, or an informed decision on the part of the physician not to offer Ace inhibitors because of some clinical issue not apparent to the software system.

The insurer invested substantial resources in developing these measures of potential gaps in care in order to track care quality and to communicate potential issues to physicians. It is important to note, however, that these measures were *not* tied to any incentive plan and there were no financial or other repercussions for physicians whose patients generated potential gaps in care. These quality measures are also useful for our purposes because they are based upon widely accepted quality indicators and because they are constructed from the same sort of billing records that are available to Medicare.

We restrict our sample to patients with a primary care doctor. Patients are defined as having a primary care doctor when a physician in a primary care specialty (internal medicine, family practice, pediatrics, general practice) is also the main provider of care as identified on the basis of claims information. Using this data we construct a dummy variable, *Any Potential Gap in Care,* which takes a value of one if any potential gap in care was observed over the period the patient is in the sample.[13] Descriptive statistics for our population are presented in Table 1.

As reported in Table 1, the mean of *Any Potential Gap in Care* is 0.29. Thus the mean value of the signal of quality success is μ_X =1-0.29 = 0.712. To calibrate our model we also

[13] Potential gaps in care were identified based on medical claims over a 30 month period. The median elapsed time between the first and last appearance of a patient in our sample is about 8 months.

require an estimate of the noise with which care quality is measured, σ_X. To obtain this we regress *Any Potential Gap in Care* on variables for age and gender as well as a vector of commonly used risk-adjustor variables known as *Hierarchical Clinicial Condition* (*HCC*) indictors. The *HCC* model is used by the Center for Medicare and Medicaid Services as a risk score to predict how costly a Medicare enrollee is likely to be relative to the national average beneficiary. It includes 70 hierarchical indicators that together describe an enrollee's clinical condition (for a full description see Pope., Kautter et al. (2004)). From this exercise we find that the standard deviation in the error term is 0.432. Fisher, McClellan et al. (2009) report that the average physician group has 260 Medicare patients. Adopting this as the relevant sample size for each physician, it follows that $\sigma_X = .432/\sqrt{260} = .027$, a result that plays a central role in following calibration exercise.[14]

<u>Converting to Money Metric Units</u>

Converting the empirical distribution of care quality to the money metric equivalent requires an assumption about the functional form relating the empirical performance measures to money-metric effort. If, as seems reasonable, the transformation of effort to measured quality exhibits decreasing returns and if there is a ceiling to the observed measure of quality, as is the case with our measure, then the exponential distribution offers a natural starting place. More formally, we can write $[X_i|e = e_0] = 1 - \exp\,(-e_0/\beta) = h(e_0)$. The principal's money-metric signal of effort is just the inverse of this transformation: $x_i = h^{-1}(X_i)$.

Having converted measured quality to money metric effort, we can then use a delta-method type approximation to determine the money metric standard deviation of the noise in the quality signal:

[14] The full regression is available in an unpublished regression.

(12) $$\sigma = \sigma_X \frac{\partial h^{-1}(\mu_X)}{\partial \mu_X}.$$

Using the exponential transformation described above, the calibration becomes

(13) $$\beta = -\frac{e_0}{ln\,(1-\mu_X)},$$

(14) $$\sigma = \sigma_X \frac{\beta}{1-\mu_X}.$$

Plugging in the values of μ_X and σ_X and normalizing e_0 = 1 so that quantities are multiples of a provider's annual net Medicare compensation under fee-for service, the money metric variation in the noise of the performance signal is σ =0.077.

With an estimated value for parameter σ, we are now in a position to explore how high must the incentive stakes be in an ACO-style group pay system to achieve target quality levels. With an average Medicare patient panel of 260 enrollees, the minimum size group of 5000 Medicare enrollees would include 20 doctors. To put the results in context, in each of the simulations we present, we compare incentive pay arrangements for groups ranging in size from 2 to 20.[15]

Figure 1 plots the bonus required to achieve fee-for-service quality levels for groups of different sizes. Consistent with our comparative static results, we observe that bonus levels rise from 1.05 to 1.26 as we move from groups of 2 to groups of 20 physicians. The number 1.05 at N=2 can be interpreted as follows: a quality bonus sufficient to achieve fee-for-service levels of quality would have to be equal to 105 percent the net physician compensation for Medicare patient enrollees under fee-for-service. For groups of 20 physicians, the requisite bonus would have to increase to nearly 126 percent of physician net compensation under fee-for-service.

[15] Details on these calculations are available in an unpublished appendix.

Figure 2 plots the payers second incentive instrument, the threshold level of performance or $\bar{x}$. Since the target level of performance has been normalized to 1, it appears that at N=2, the threshold level of performance is a bit under 90 percent of target performance levels. At N=20, threshold performance approaches 98 percent of target levels. Taken together with the results from Figure 1, this means that as groups grow in size, threshold levels grow so that only tiny deviations in measured performance can have a very large effect on payouts.

Figure 3 plots the probability that groups have of clearing the performance target and earning bonuses. This probability is determined by the effort provided by physicians, by the precision of the performance measure, and by the magnitude of the performance threshold. At N=2 groups have a 97 percent chance of winning the bonus. As groups increase in size, performance measures become more precise but this effect is swamped by the increase in free-riding, so that by N = 20, the chances of winning the bonus falls to 81 percent.

The patterns in Figures 1 and 3 indicate that the variability of payments under these ACO-style incentives increase vary dramatically as we move from small groups to large. The standard deviation of the variable proportion of pay is $\sqrt{b^2 \Pr(Success) * (1 - \Pr(Success))}$ At N=2, the standard deviation of pay is 0.18, while at N=20 the equivalent number is 0.35. To put these numbers in perspective, it is worth remembering that the participation constraint in our model requires that total expected earnings be the same for all group sizes. Thus the standard deviation of earnings relative to expected pay is twice as large for groups with N =20 than for groups with N = 2.

The finding that free-riding problems increase very quickly with group size is consistent with findings in other medical settings. Gaynor, Rebitzer et al. (2004), for example, find that a bonus of 20 percent of physician income had a powerful and positive effect on achieving cost targets within a network HMO, but the effect was largest in groups of three physicians and declined markedly as group size approached six physicians.

It is worth noting that our results are not the result of choosing a very noisy performance measure or unrealistically small panel sizes. Taking the second issue first, the median number of Medicare beneficiaries in a *practice* in Nyweide, Weeks et al. (2009) is 260, suggesting that the median caseload for a physician would be much smaller. Also, the caseload for any given quality measure is a small fraction of the total caseload (see their Table 2). To compare their results with ours, we would need to know the caseload per physician not per practice, which they do not show. To get a rough sense of magnitudes, however, we plugged the median caseload from single physician practices and their mammography performance measures (with a mean of 0.695) into our model. For N=2, the required bonus is 175% of fee-for-service net income with a probability of successfully clearing the performance threshold = 0.7. For N=20, the bonus is 525% of fee-for-service income and the probability of successfully earning the bonus = 0.56. Obviously these are much higher-stakes incentives than our calibration produces.

The finding that requisite bonuses increases strongly with size while the probability of success falls suggests a novel interpretation of the generally small and highly variable results of pay-for-performance in the health care setting ((Rosenthal 2008); (Rosenthal 2009); (Rosenthal and Frank 2006); (James 2012)). A very recent example is Carrie H. Colla, David E. Wennberg et al. (2012) study of the groups that participated in the Physician Group Practice Demonstration Project. They found that allowing these groups to keep 80 percent of *any savings* (after the first two percent) elicited only small and uneven reduction in costs. From the perspective of our model, this would be expected because the incentive intensity under the Group Pay Demonstration Project was simply too low to overcome the free-riding problem in large physician groups. It seems reasonable then to wonder if the groups that did manage to mobilize physicians around cost-conscious medical practice must have had alternative motivators that complemented the low power of the monetary incentives. We take this issue up in the next section of the paper.

IV. High Stakes Incentives, Integration Costs and Mitigating Strategies.

Integration Costs

In this section we discuss the difficulties that very high stakes incentives pose for organizations as well as strategies for mitigating these problems. In the language of transaction cost economics, these difficulties are best understood as integration costs, i.e. the costs of structuring incentives within an organization.

The first of these integration costs emerges from the high payoff to gaming the performance measurement system just a little bit. According to our calibration, groups with 20 members have a minimum performance threshold that is set at 98 percent of target effort levels and a bonus of nearly 130 percent of physician net earnings under fee-for service Medicare. Free riding in groups this size so depresses effort that the probability of clearing the threshold is only 81 percent. In this situation, the payoff to manipulating the performance measure may prove to be irresistible. Manipulation could take the form of distorting the performance measure directly – something analogous to the widespread backdating of stock options in firms that used these options as part of their incentive systems (Heron and Lie 2009).[16] A more likely and harder to detect form of manipulation in the health sector, however, might be for groups to find subtle ways to turn away or discourage patients likely to have low quality scores. Dranove, Kessler et al. (2003) for example find evidence of such steering in a cardiac report card system. Medicare,

[16] Similarly Dafny (2005) finds that hospitals up-coded with great speed and subtlety when changes in Medicare prices made it profitable to do so. This was true even though the penalties for up-coding were potentially severe.

or any other payer trying to operate a high-stakes incentive system of the sort we model, would have to devote considerable resources to preventing these privately profitable but socially wasteful manipulation activities.

A second integration cost emerging from high-stakes incentives is the cost of multi-tasking. If incentives cause providers to devote lots of attention and effort to achieving quality targets, they may divert attention from other activities that create clinical or economic value. In the case of ACOs, high stakes incentives are most likely to displace experiments or innovations that lower measured quality in the present in the hopes of producing higher quality at lower cost in the future. Given that the determinants of high quality, cost-effective health care are not well understood, discouraging of experimentation is likely to be a particularly important integration cost.

The third type of integration cost is the result of heightened financial risks emerging from high-stakes incentives. As we have seen, this risk grows dramatically as ACOs approach minimum mandated size. For ACOs organized by large organizations with deep pockets and easy access to credit markets, it may be feasible to self-insure against these risks. Less wealthy ACOs may be able to purchase reinsurance to protect themselves from very adverse outcomes just as relatively small, self-insured employers do. Reinsurance contracts typically require the reinsurer to bear risk on individuals above some threshold amount – thereby protecting the policy holder from a patient suddenly acquiring a very expensive condition. Insurance contracts against adverse quality outcomes, however, would likely be much harder to write because quality outcomes are harder to agree upon, harder to audit externally, and more subject to manipulation than cost outcomes.

Mitigation Strategies

Having identified the important types of integration costs, we turn to strategies for mitigating these costs.

As we have discussed, an important cause of very high-stakes incentives is the low signal to noise ratio in quality metrics. Because the precision of performance metrics increases by a factor of $\sqrt{N}$, ACOs would function better under small, closed networks of primary care physicians where the share of each physician's patient panel exposed to the incentive scheme is large. In this sense, the fragmented system of health care financing in the United States makes it more difficult to move away from patterns of fragmented care delivery.

Restricting ACOs to small, closed networks, however, runs contrary to current legislation and to the goal of promoting integrated care into regions where care is quite fragmented. An alternative strategy would be for ACOs to persuade all their payers to use the same quality and cost metrics, but the precision of performance measures won't be improved by this unless payers are also persuaded to pay on the basis of the ACO's aggregate performance as well.

Improving the precision of performance measures does not eliminate the other important driver of the costs of integration: the free-riding problem that causes incentives to weaken dramatically as the number of physicians in the group grows. Attacking this problem requires that the organization find ways to reduce the cost to the physician of providing effort. In the case of ACOs, the most important determinant of the cost of effort is likely to be the opportunity cost of the physician's time. A doctor, for example, who spends more of the day meeting quality targets for Medicare patients, loses the opportunity to see more fee-for-service commercial patients. For physicians who are employees, the obvious way to reduce this opportunity cost of effort, is to restrict the scope of the physician's practice to Medicare patients. This is an illustration of a more general point made by Holmstrom and Milgrom (1991). In employment relationships, incentive pay and job design are powerful and complementary motivational instruments. By narrowing the scope of work, employers can greatly reduce the opportunity cost of effort and so operate with low powered incentives (Roberts 2004).

Looking beyond conventional organizations, it is helpful to think of the ACO's incentive problem as analogous to the provision of effort when effort is a public good. Physicians have intrinsic motives for providing effort – a fact that is captured in the model parameter, e_0. Large, hybrid ACOs may be able to achieve quality targets with lower-stakes incentives if they could find an instrument for increasing e_0. Kandel and Lazear (1992) offer a model of peer pressure and mutual monitoring among partners in a partnership that is essentially a reduced form description of this possibility. Encinosa, Gaynor et al. (2007) offer some theory and evidence that these effects matter for incentive design in medical groups.[17] The large experimental literature on the provision of public goods suggests, however, that incentives do not always have this positive effect on intrinsic motives. In their extensive review of the experimental literature, Bowles and Polania-Reyes (2010) conclude that the effects of incentives on public good provision depend critically on the *meaning* agents give to the incentive. They speculate that well-designed incentives should be implemented in a way that helps the agent understand that "the desired modification in her actions will serve to implement an outcome that is socially beneficial so that the target is more likely to endorse the purpose of the incentive, rather than being offended by it as either unjust or a threat to her autonomy…".[18]

[17] . The key to the incentive power of norms is that violating norms triggers actions or changed perceptions in *other* agents and this can greatly magnify the costs of violating the norm. In some models, agents compare their actions with a prescribed set of behaviors (Akerlof 1976), (Akerlof 1980) or with the actions of others in their reference groups (Akerlof and Kranton 2000), (Akerlof and Kranton 2002).
[18] This point is echoed in the growing literature on behavioral agency theory. See Rebitzer and Taylor (2010) and Ellingsen and Johannesson (2007) for extensive reviews and discussions.

Thus structuring incentives within ACOs likely involves paying careful attention to assigning meaning to the payments, but it is unclear if this meaning is more easily constructed within conventional employment relationships or within hybrid organizations. Given the medical profession's long history of battling to preserve its status as an autonomous and learned profession, low powered incentives in ACOs built upon a hybrid organizational form might be workable.[19] On the other hand, conventional organizations may have greater opportunities to train, screen and socialize for physicians who might respond well to low-powered incentives.

To the extent that successful ACO's have organizational capabilities that rely on training, screening, socialization and the "meaning" of incentives they likely also involve relational contracts. Relational contracts are based upon informal trusting arrangements whose credibility is enforced by the continuing value of the relationship between parties, i.e. by the "shadow of the future" (Gibbons and Henderson 2011). The great advantage of relational contracts for our purposes is that they can complement more formal relationships such as those involved in pay for performance. Incentives that would be under-powered in the sense of our principal-agent model may be quite a bit more effective if performance this period determined the continuation of a valuable on-going relationship.[20] Relational contracts can also be used to reduce some of the distortions of high-powered formal incentives.[21] The great disadvantage of relational contracts is that the persistent performance benefits they offer are hard to realize within

[19] Starr (1984); Cebul, Rebitzer et al. (2008)) and Rebitzer and Votruba (2011) discuss implications of this autonomy for the evolution of care delivery in the United States.

[20] The value of an on-going relationship can be enhanced simply by increasing the fixed component of compensation as in efficiency wage models. Alternatively, being in a relationship where the problems of credibility and clarity are acceptably resolved increases the value of the on-going relationship relative to the value of starting anew in another organization.

[21] Production workers at Lincoln Electric, for example, received a weekly salary determined entirely by piece rates. These high piece rates produced a furious work pace and high worker earnings, but they also made it difficult to experiment with new production methods or quality improvements because any change that interfered with production cost workers money. This undesirable side-effect of high-powered piece rate incentives was partly offset by a very large annual bonus that was distributed to production workers based on managers' subjective assessment of their willingness to cooperate and help improve the overall functioning of the enterprise (Gibbons and Henderson 2011).

organizations and hard to diffuse across organizations.[22] Even if it becomes apparent that successful ACOs depend upon relational contracts, it may be difficult and slow for imitators of successful ACOs to themselves become successful.

Conclusions

ACOs are a new model for integrated health care delivery created by the Obama Administration's Patient Protection and Affordable Care Act. ACOs are designed to promote the benefits of integrated care by enabling groupings of hospitals and providers to jointly contract with the CMS to provide care to a population of Medicare enrollees in an environment that rewards cost-efficiency through global budgets and pay-for-performance incentives. By aggregating the experience of many enrollees, ACOs improve the signal to noise ratio in performance measures. For this reason, ACOs are required to have at least 5,000 enrollees. Combining enrollees requires, however, combining physicians and as the numbers of physicians grow so does the free-riding problem. Working with a stylized model of ACO-like incentives, we establish that the negative effects of free-riding swamp the positive effects of increased precision. We then calibrate the model using proprietary performance measures from a very large insurer. Our estimates suggest that even minimally sized ACOs will likely require unmanageably high stakes incentives. These very high-stakes pay-for-performance systems increase incentives to game the measurement system, increase the difficulties of experimenting with new care delivery methods, and increase the financial risk ACOs must bear. To achieve their goals under these circumstances, ACOs will likely have to operate with under-powered

[22] Gibbons and Henderson (2011) argue that this difficulty stems from the fact that relational contracts must be both credible and clear. Credibility requires that parties to the agreement can trust in the future actions of their counterparties. Clarity requires that parties have a shared understanding of what it means to act in support of the agreement even in circumstances that are novel and unanticipated.

incentives and combine these with alternative motivational strategies that complement "pay-for-performance." Some of these complementary motivational strategies can only be implemented in conventional integrated organizations, while others may prove workable in hybrid organizational forms that more are more congruent with practice patterns in regions where care delivery is currently highly fragmented.

Our analysis has a number of limitations. First, our measure of performance is taken from a population of chronically ill commercial insurance patients and the performance measures we use are not the same that CMS might use in tracking care quality for its population of Medicare patients who are primarily over age 65. Secondly, our formal model is not a detailed depiction of each facet of the ACO payment system; rather, it is a stylized representation of the essential features of the system: global budgeting and bonuses linked to noisy performance measures. Finally, our calibration of the model is based on a number of simplifying, but restrictive functional form assumptions – most notably that physicians are risk neutral in responding to Medicare's group incentive.

While none of these limitations is likely to overturn our qualitative conclusions, they do suggest that our analysis is not likely to provide a quantitative prediction of behavior in actual ACOs. Rather, the contribution of our model and its calibration is that it helps analysts think systematically about the key determinants of incentive intensity and their likely effects. If, for example, CMS used much more precise quality metrics than the commercial health insurer, ACOs could operate successfully with lower powered incentives. If physicians were highly risk averse with respect to their ACO payments, incentive intensity would similarly decline, but the cost of compensating providers would also increase.

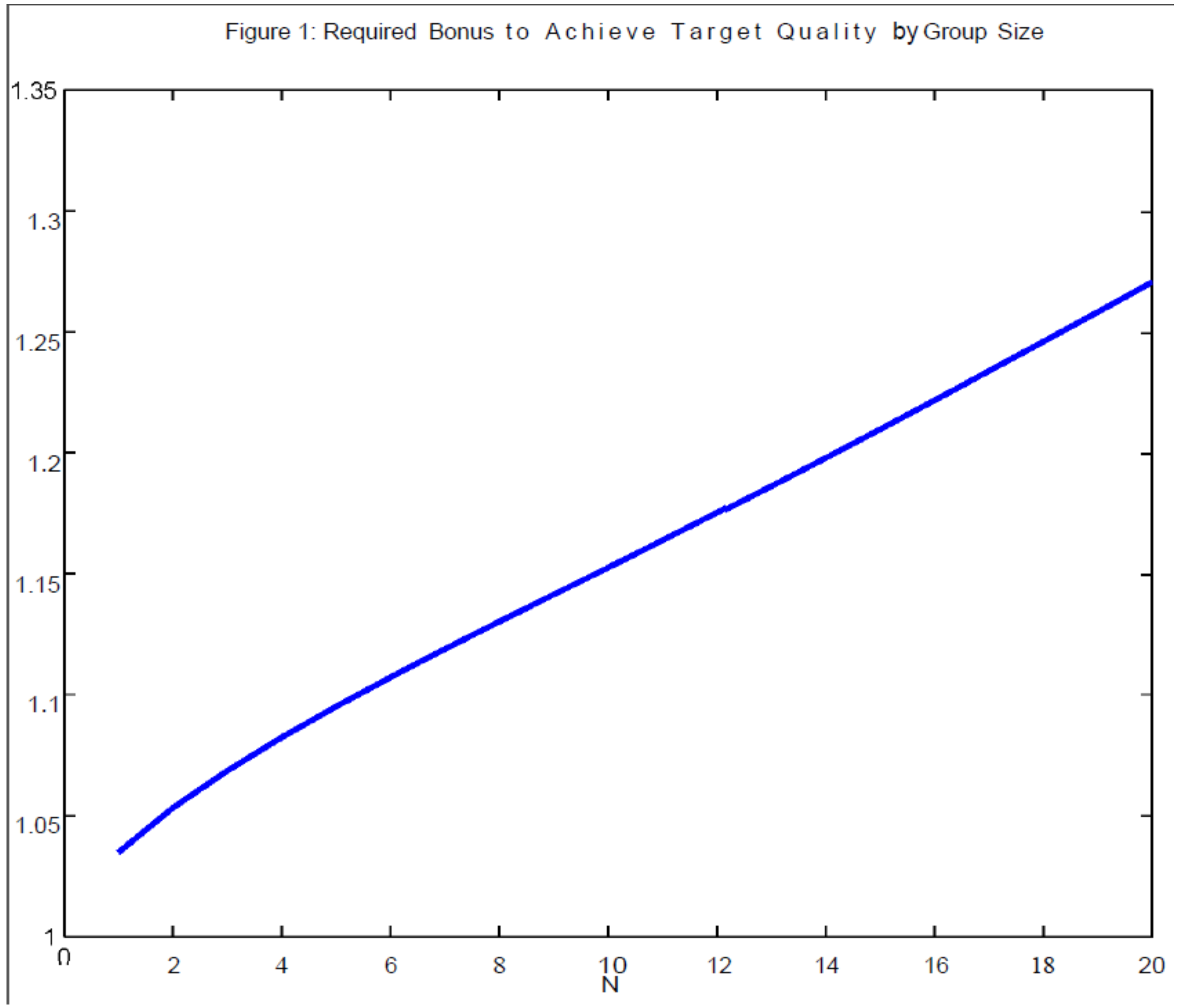
From the perspective of health care policy, our analysis has two important implications. The first is a novel interpretation of prior pay-for-performance experiments that find small but highly variable results. The most recent these (Carrie H. Colla, David E. Wennberg et al.), concerns large physician groups. Comparing our calibration with the incentive payouts used in the actual experiment suggests that the incentives in these experiments are far too low to overcome free-riding problems. The great variability in outcomes across sites might, therefore, have to do with the ability of different organizations to employ alternative "motivators" that complement the effect of low stakes incentives. The second implication is that ACOs will have difficulty writing workable incentive contacts in the sort of loose, open networks envisioned in the legislation. Achieving quality and cost targets though pay-for-performance incentives is most feasible when physicians are dependent on Medicare for a large fraction of their patient panel. This implies that the relatively small commercial payers will have an even harder time using pay-for performance contracts to induce cost-conscious, quality-preserving practices among their physician networks.

From the perspective of transactions cost economics, the study of ACOs offers a rich laboratory for exploring the transaction costs of bringing activities within an organization. It particularly offers the opportunity to compare how effectively hybrid and conventional organizations can mitigate the costs of integration. In a large proportion of US regions, health care delivery is still quite fragmented and it is hard to imagine the large scale migration of independent practice physicians into employment relationships. If hybrid organizations can mitigate integration costs as effectively as conventional organizations, ACOs might be a very effective tool for catalyzing the spread of integrated health care organizations. If not, then any gains from ACOs may be limited to regions where care is “mostly integrated” already.

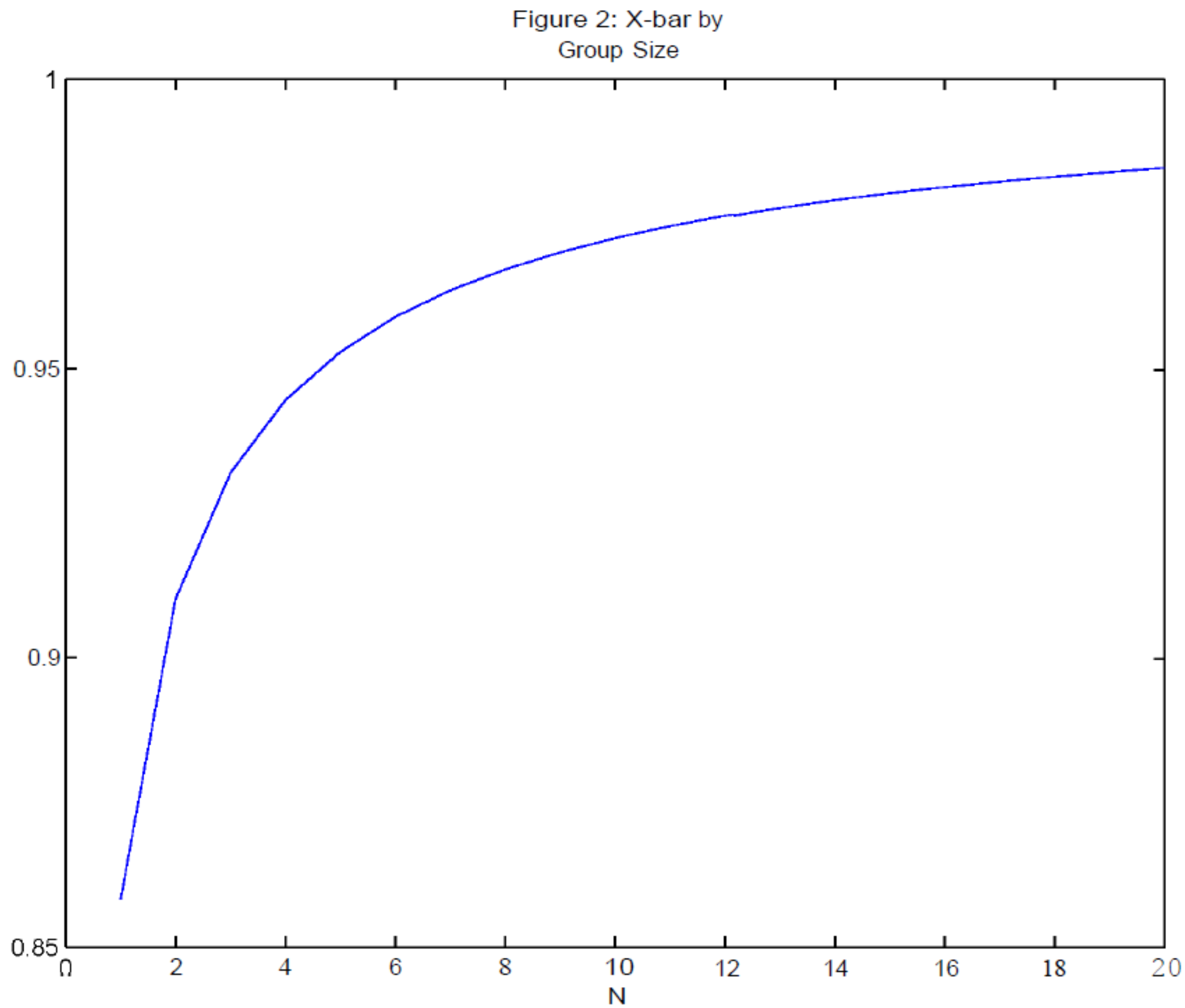
Table 1

Variable	Mean	Std. Dev.
Any Potential Care Gap	0.29	0.45
Age	45.95	15.15
Fraction Female	0.57	0.49
Incidence of Common Chronic Diseases		
Fraction with Diabetes	0.18	0.39
Fraction with Hypertension	0.45	0.5
Fraction with Ischemic Heart Disease	0.13	0.33
Fraction with Congestive Heart Disease	0.03	0.17
Fraction with Chronic Obstructive Pulmonary Disease	0.06	0.24
Fraction with Two or More Common Chronic Diseases	0.21	0.41
Number Patients	564,049	
Number of Primary Care Physicians	59,087	

The patient sample contains commercial insurance members whose employers are fully insured, and who have evidence of chronic illness. The provider sample consists of the primary care providers identified as the main providers for these patients on the basis of claims information

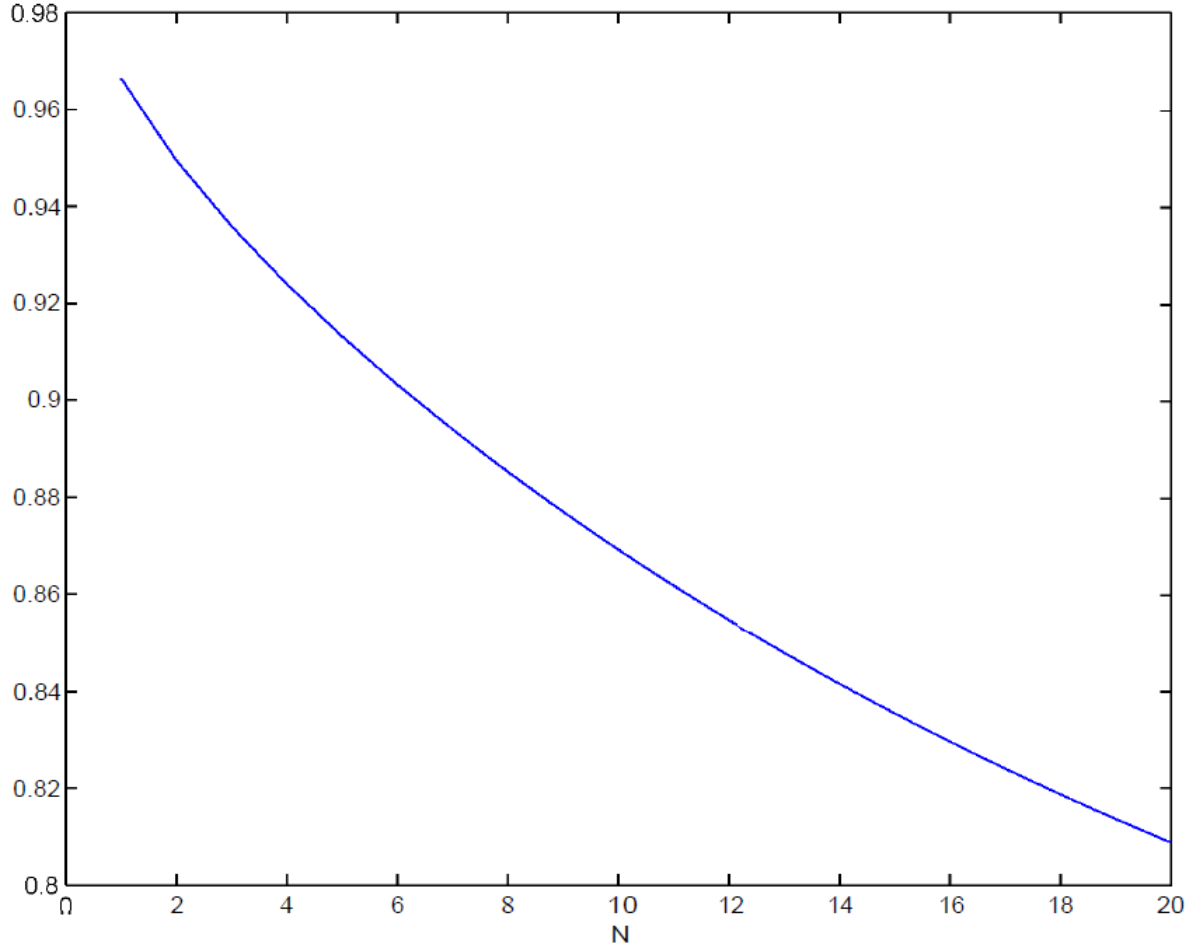


This figure plots the bonus required to achieve fee-for-service levels of quality while practicing cost-conscious medicine. The calibration takes as given the number of doctors in the group, *N*, the size of their panel of Medicare enrollees, 260, and our estimate of the standard deviation of the noise component of the performance measure; $\sigma = 0.077$. The bonus is expressed relative to physician earnings under fee-for-service compensation. Details in text.



This figure plots the threshold quality level that triggers payment of the bonus. The calibration takes as given the number of doctors in the group, *N*, the size of their panel of Medicare enrollees, 260 and our estimate of the standard deviation of the noise component of the performance measure, $\sigma = 0.077$. The target level of effort is that required to achieve the same quality prevailing under fee-for-service. It is normalized to 1, so the threshold is expressed as the fraction of the target level. Details in text.

Figure 3: Probability of Reaching Target by Group Size



This figure plots the probability a group has of exceeding performance thresholds and receiving the bonus. The calibration takes as given the number of doctors in the group, *N*, the size of their panel of Medicare enrollees (260), and our estimate of the standard deviation of the noise component of the performance measure; $\sigma = 0.077$. The target level of effort is that required to achieve the same quality prevailing under fee-for-service. Details in text

Bibliography

Akerlof, G. A. (1976). "The Economics of Caste and of the Rat Race and Other Woeful Tales." Quarterly Journal of Economics **90**(4): 599-617.

Akerlof, G. A. (1980). "A Theory of Social Custom, of Which Unemployment May be One Consequence." The Quarterly Journal of Economics **94**(4): 749-775.

Akerlof, G. A. and R. E. Kranton (2000). "Economics and Identity." Quarterly Journal of Economics **115**(3): 715-753.

Akerlof, G. A. and R. E. Kranton (2002). "Identity and Schooling: Some Lessons for the Economics of Education." Journal of Economic Literature **40**(4): 1167-1201.

Baicker, K. and A. Chandra (2011). Aspirin, Angioplasty, and Proton Beam Therapy: The Economics of Smarter Health Care Spending. Jackson Hole Economic Policy Symposium.

Becker, G. S. and K. M. Murphy (1992). "The Division of Labor, Coordination Costs, and Knowledge." Quarterly Journal of Economics **107**(4): 1137-1160.

Bowles, S. and S. Polania-Reyes (2010). "Economic Incentives and Pro-social Preferences as Complments or Substitutes."

Burns, L. R. and M. V. Pauly (2012). "Accountable Care Organizations May Have Difficulty Avoiding The Failures Of Integrated Delivery Networks Of The 1990s." Health Affairs **31**(11): 2407-2416.

Carrie H. Colla, P., M. David E. Wennberg, MPH, et al. (2012). "Spending Differences Associated with the MEdicare Physician Group Practice Demonstration." Journal of the American Medical association **308**(10): 1015-1023.

Cebul, R., J. B. Rebitzer, et al. (2008). "Organizational Fragmentation and Care Quality in the US Health Care System." Journal of Economic Perspectives **22**(4): 93-113.

Chandra, A. and J. Skinner (2012). "Technology Growth and Expenditure Growth in Health Care." Journal of Economic Literature **50**(3): 645-680.

Chandra, A. and D. O. Staiger (2007). "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks." Journal of Political Economy **115**(1): 103-140.

Clark, J. and M. Rosenthal (2008). ProvenCare: Geisinger Health System and Pay for Performance. Harvard School of Public Health. Boston, Harvard University.

Crosson, F. J. (2009). "21st Century Health Care- The Case for Integrated Delivery Systems." New England Journal of Medicine **36**(14): 2.

Crosson, F. J. (2011). "The Accountable Care Organization: Whatever its Growing Pains, The Concept is Too Vitally Important to Fail." Health Affairs **30**(7): 1250-1255.

Dafny, L. S. (2005). "How Do Hospitals Respond to Price Changes?" American Economic Review **95**(5): 1525-1547.

Dranove, D., D. P. Kessler, et al. (2003). "Is more Information Better? The Effects of 'Report Cards" on Health Care Providers." Journal of Political Economy **111**(3): 555-588.

Ellingsen, T. and M. Johannesson (2007). "Paying Respect." Journal of Economic Perspectives **21**(4): 135-140.

Fisher, E. S., M. B. McClellan, et al. (2009). "Fostering Accountable Health Care: Moving Forward In Medicare." Health Affairs **28**(02782715): W219-W219-W231.

Garicano, L. and T. Santos (2004). "Referrals." American Economic Review **94**(3): 499-525.

Gaynor, M., J. B. Rebitzer, et al. (2004). "Physician Incentives in Health Maintenance Organizations." Journal of Political Economy **112**(4): 915-931.

Gibbons, R. (2010). "Transaction-Cost Economics: Past, Present, and Future?" The Scandinavian Journal of Economics **112**(2): 263–288.

Gibbons, R. and R. Henderson (2011). "Relational Contracts and Organizational Capabilities." Organization Science (forthcoming).

Ginsburg, P. B. (2011). "Spending to Save – ACOs and the Medicare Shared Savings Program." New England Journal of Medicine **364**(22): 2.

Heron, R. and E. Lie (2009). "What Fraction of Stock Option Grants to Top Executives Have Been Backdated or Manipulated?" Management Science **55**(4): 513.

Holmstrom, B. and P. Milgrom (1991). "Multi-task Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design." Journal of Law Economics and Organization **7**(special issue): 24-52.

James, J. (2012). "Health Policy Brief: Pay for Performance." Health Affairs.

Kandel, E. and E. P. Lazear (1992). "Peer Pressure and Partnerships." Journal of Political Economy **100**(4): 801-817.

Masten, S. E., J. W. Meehan, et al. (1991). "The Costs of Organization." Journal of Law, Economics and Organizations **7**(1): 1-25.

Meyer, H. (2012). "Many Accountable Care Organizations Are Now Up And Running, If Not Off To The Races." Health Affairs **31**(11): 2363-2367.

Nyweide, D. J., W. B. Weeks, et al. (2009). "Relationship of Primary Care Physicians' Patient Caseload With Measurement of Quality and Cost Performance." JAMA: The Journal of the American Medical Association **302**(22): 2444-2450.

Pope., G. C., J. Kautter, et al. (2004). "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model." **25**(4).

Rebitzer, J. B. and L. J. Taylor (2010). Extrinsic Rewards and Intrinsic Motives: Standard and Behavioral Approaches to Agency and Labor Markets. Handbook of Labor Economics. O. Ashenfelter and D. Card, North-Holland. **4A:** 701-772.

Rebitzer, J. B. and M. Votruba (2011). "Organizational Economics and Physician Practices." National Bureau of Economic Research(17535).

Ritter, J. A. and L. J. Taylor (2011). "Racial Disparity in Unemployment." Review of Economics and Statistics **93**(1): 30-42.

Roberts, J. (2004). The Modern Firm: Organizational Design for Performance and Growth. Oxford, Oxford University Press.

Rosenthal, M. (2008). "Beyond Pay for Performance - Emeging Models of Provider Payment Reform." New England Journal of Medicine **359**(12): 4.

Rosenthal, M. B. (2009). "What Works in Market-Oriented Health Policy?" New England Journal of Medicine **360**(21).

Rosenthal, M. B. and R. G. Frank (2006). "What Is the Empirical Basis for Paying for Quality in Health Care?" Medical Care Research and Review **63**(2): 135-157.

Shields, M. C., P. H. Patel, et al. (2011). "A Model For Integrating Independent Physicians Into Accountable Care Organizations." Health Affairs **30**(1): 161-161-172.

Shortell, S. M. and L. P. Casalino (2010). "Implementing Qualifications Criteria and Technical Assistance for Accountable Care Organizations." JAMA: The Journal of the American Medical Association **303**(17): 1747-1748.

Shortell, S. M., L. P. Casalino, et al. (2010). "How The Center For Medicare And Medicaid Innovation Should Test Accountable Care Organizations." Health Affairs **29**(7): 1293-1293-1298.

Singer, S. and S. M. Shortell (2011). "Implementing Accountable Care Organizations." JAMA: The Journal of the American Medical Association.

Song, Z., D. G. Safran, et al. (2011). "Health Care Spending and Quality in Year 1 of the Alternative Quality Contract." New England Journal of Medicine **1056**(10).

Starr, P. (1984). The Social Transformation of American Medicine, Basic Books.

Zirui Song, D. G. Safran, et al. (2011). "Health Care Spending and Quality in Year 1 of the Alternative Quality Contract." New England Journal of Medicine.