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 so long as the ACO achieves certain quality thresholds. 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 model of ACO incentives, we establish that the negative effect of free-riding swamps the positive effect of increased precision. We then calibrate the model using proprietary performance measures from a very large insurer and find that free-riding within ACOs is so severe that the pay-for-performance plans will likely not be self-financing, i.e. the savings they produce will not be sufficient to pay for the requisite performance bonuses. It follows that to achieve their goals ACOs will have to find ways to augment under-powered pay-for-performance incentives with motivational strategies that complement pay-for-performance. We observe that some of these motivational strategies can only be implemented in conventional integrated organizations, while others may prove workable in hybrid organizational forms. This has implications for the efficacy of ACOs as a policy for promoting integrated care delivery.

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 care delivered 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? Or 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 maintain care quality at 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 preventive 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, pay-for-performance comes uncomfortably close to pay for 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 reliably identify meaningful performance improvements, a minimum size requirement that CMS has since adopted.

From an economic perspective this statistical approach to determining the optimal scale of an organization is incomplete. 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 in performance measures. 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 model of physician incentives under ACO-style incentives. Our approach builds upon conventional principal-agent models, but is unusual in that it focuses on the sort of nonlinear incentives built into the ACO program and commonly used elsewhere in health care – shared savings from global budgets to encourage cost-efficient practice styles with payouts conditional on achieving target levels of care quality.

We further investigate the nature of the ACO incentive problem by calibrating our incentive model using confidential claims data and quality measures from a very large sample of chronically ill patients. Here we find that the free-riding problems within ACOs of requisite size are so severe that pay-for-performance plans aimed at achieving meaningful cost reductions will typically not be self-funding, i.e. the savings they produce will not cover the costs of the performance bonuses. It follows that ACOs committed to self-financed pay-for-performance will operate with under-powered incentives. Successful ACOs will have to find ways to augment their under-powered incentives with motivational strategies that complement pay-for-performance.

In terms of transaction cost economics, the difficulty of implementing these complementary motivational strategies determines the cost of transacting care delivery *within* organizations. Interestingly some of the complementary strategies we identify will only be workable within conventional integrated organizations, while others might be implemented within hybrid organizational forms that are more congruent with fragmented practice patterns. 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, it but has not yet provided a satisfactory account of variations in the cost of transacting *within* an organization.[1] From this

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

perspective ACOs provide an interesting and policy-relevant laboratory for examining integration costs for varying organizational types.

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 section three presents the results of our calibration exercise. Section four considers the problem of augmenting under-powered pay-for-performance incentives within ACOs.

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 U.S. 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 (2011) review of the organizational economics of physician's practices.

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 cannot 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 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

[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).

[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).

payers don't innovate away from the status quo fee-for-service payment system and there is little competitive advantage for providers to move out of their currently fragmented delivery organizations. [5]

Accountable Care Organizations are designed to overcome these impediments to payment reform.[6] 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.[7] 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, preventive care, 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

---

[5] In comments on a previous draft of this paper, Daniel Kessler pointed out another contributing issue in the private sector. The fact that private insurance expenditures are tax exempt further reduces gains from eliminating inefficient spending.

[6] 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).

[7] 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 paid on a fee-for-service basis rather than on a per member per month basis- further clouding the picture (Meyer, 2012).

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 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 discuss below, the transformation of hybrid and virtual physician organizations into ACOs poses special problems and opportunities for incentive design. [8]

Larson, Van Citters, et al. (2012) offer an in-depth look at four recently formed ACOs that gives a tangible sense of the variety of organizations involved. One is an independent practice association that employs 700 physicians and has 2400 affiliated; another is an integrated

[8] 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.

hospital delivery system that employs 475 doctors and owns five hospitals. The third is a loose independent practice association with 40 employed physicians and 2500 affiliated ones that has affiliations with 18 hospitals but owns none of its own. The fourth is a community hospital system that employs 16 physicians and has 800 affiliated and owns two hospitals.

In addition to influencing primary care practice, 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 facilitates generalists substituting 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 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 (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 multi-task model of physician incentives. Physician effort and attention is divided between finding ways to generate income and provide quality care to patients. In a fee-for-service environment there is little tension between these two goals: physicians get reimbursed for all the medically necessary care their patients require. Things are different, however, in an ACO environment. Here global budget caps make it possible for providers to profit by not providing care to payments. For this reason ACO contracts specify that providers can keep some portion of savings below the global budget cap provided that the organization also clears specific quality thresholds. [9]

More formally, we model an ACO as a team of N doctors who accept a global budget for the care of a defined group of patients, and consider how the principal (i.e., CMS) should choose savings bonuses and quality thresholds in order to induce a desired level of cost savings and care quality. Following the typical contract theoretical framework, we first model the physicians' best responses to a given incentive scheme, and take those best responses as constraints in the principal's decision problem.

Total costs of care for the team depend on each doctor i's cost-control efforts, e_i^c, which are measured in money-metric units, as well as noise, ε_i:

$$\sum_{i=1}^{N} C + \varepsilon_i - e_i^c$$

where C is the average baseline cost of care. Quality of care likewise depends on noise and money-metric effort devoted to quality. Average quality for the team is

$$\frac{1}{N}\sum_{i=1}^{N} e_i^q - \varepsilon_i^q$$

The cost and quality disturbances are not observable, have mean zero and variance $\sigma_c^2$and σ_q^2 respectively, and are independent from each other and across doctors.

ACO members are compensated based on the entire team's level of costs and quality. The team splits evenly a fraction, b, of savings relative to baseline provided that average quality

[9] The actual incentive contracts are more varied than this. **Meyer (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.

exceeds a specified threshold $\bar{x}$. Doctors are risk neutral and maximize expected income minus effort costs. Doctor *i*'s payoff is therefore

$$(1)\ U_i = E\left[b\left(C - \frac{1}{N}\left(\sum_{j=1}^{N} C + \varepsilon_j - e_j^c\right)\right)1\left(\frac{1}{N}\sum_{j=1}^{N} e_j^q - \varepsilon_j^q \geq \bar{x}\right)\right] - \frac{\left(e_i^q + e_i^c\right)^2}{2}$$

The last term in the utility function reflects the multi-task nature of effort devoted to cost reduction and to quality. An increase in e_q increases the marginal cost of providing effort for cost-reduction activities and an increase in e_c similarly increases the marginal cost of quality improving efforts. It is straight-forward to rewrite the utility function as:

$$(2)\ U_i = \frac{b}{N}\left(e_i^c + \sum_{j\neq i} e_j^c\right)\Pr\left(\frac{1}{N}\left(e_i^q - \varepsilon_i^q + \sum_{j\neq i} e_j^q - \varepsilon_j^q\right) \geq \bar{x}\right) - \frac{\left(e_i^q + e_i^c\right)^2}{2}$$

$$(3)\ \underset{p}{\rightarrow} \frac{b}{N}\left(e_i^c + \sum_{-i}^{N} e_{-i}^c\right)\Phi\left(\frac{1}{\sigma_q}\left(e_i^q/\sqrt{N} + \sum_{-i}^{N} e_{-i}^q/\sqrt{N} - \sqrt{N}\bar{x}\right)\right) - \frac{\left(e_i^q + e_i^c\right)^2}{2},$$

where for physician teams of sufficient size the normal cdf, Φ , will be a good approximation. The first order conditions for doctor $i's$ best response to shared savings fraction b and quality threshold $\bar{x}$ are:

$$(4)\ \frac{\partial U}{\partial e_i^c} = \frac{b}{N}\Phi\left(\left(e_i^q/\sqrt{N} + \sum_{-i}^{N} e_{-i}^q/\sqrt{N} - \sqrt{N}\bar{x}\right)/\sigma_q\right) - \left(e_i^q + e_i^c\right) = 0$$

$$(5)\ \frac{\partial U}{\partial e_i^q} = \frac{b}{N\sqrt{N}\sigma_q}\left(e_i^c + \sum_{-i}^{N} e_{-i}^c\right)\phi\left(\left(e_i^q/\sqrt{N} + \sum_{-i}^{N} e_{-i}^q/\sqrt{N} - \sqrt{N}\bar{x}\right)/\sigma_q\right) - \left(e_i^q + e_i^c\right) = 0$$

The second derivatives – upon which the second order conditions will depend – are:

$$(6)\ U_{cc} = -1$$

$$(7)\ U_{qq} = \frac{b}{N^2\sigma_q^2}\left(e_i^c + \sum_{-i}^{N} e_{-i}^c\right)\phi'\left(\left(e_i^q/\sqrt{N} + \sum_{-i}^{N} e_{-i}^q/\sqrt{N} - \sqrt{N}\bar{x}\right)/\sigma_q\right) - 1$$

$$(8)\ U_{cq} = \frac{b}{N\sqrt{N}\sigma_q}\phi\left(\left(e_i^q/\sqrt{N} + \sum_{-i}^{N} e_{-i}^q/\sqrt{N} - \sqrt{N}\bar{x}\right)/\sigma_q\right) - 1$$

Consider a symmetric equilibrium where all doctors choose the same effort levels. Then the first-order conditions are:

(9) $\frac{b}{N}\Phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right) - \left(e_i^q + e_i^c\right) = 0$

(10) $\frac{b}{\sqrt{N}\sigma_q} e^c \phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right) - \left(e_i^q + e_i^c\right) = 0$

The second order conditions (evaluated at the first-order conditions) are

(11) $U_{cc}U_{qq} - U_{cq}^2 > 0 \Leftrightarrow \frac{e^q + e^c}{be^c}\left(2 - \frac{1}{N}\frac{e^q + e^c}{e^c}\right) > \frac{e^c}{\sigma_q^2}\phi'\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right)$

For policy purposes we treat the Center for Medicaid and Medicare Services (CMS) as the principal. CMS determines the levels of cost sharing, b, and quality thresholds, $\bar{x}$, that must be set in order to induce desired levels of cost savings, e^c, and care quality, e^q. A sufficient condition for the second-order conditions to be satisfied is that: (a) the desired level of cost effort, e^c, is at least as great as the target level of quality effort, e^q, and (b) that the quality threshold is less than the target quality effort. The intuition for (a) comes from the fact that the quality incentive derives from shared savings - if the potential shared cost savings are tiny so must be the incentive to provide quality effort. Condition (b) is similarly intuitive and emerges from the simple mathematics of threshold incentives. [10]

From the principal's point of view, therefore, the first-order conditions determine the required choice of savings bonus, b, and quality threshold, $\bar{x}$, for any desired effort levels:

(12) $F^1 = \frac{b}{N}\Phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right) - (e^q + e^c) = 0$

(13) $F^2 = \frac{b}{\sqrt{N}\sigma_q} e^c \phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right) - (e^q + e^c) = 0$

[10] To see this consider that the expected return to marginal effort varies with the threshold performance level. If $\bar{x}$ far exceeds current effort levels, 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 $\bar{x}$.

With this set-up in place, we can then solve for the comparative static properties of the model. More specifically we consider how the principal will set her two incentive instruments (the fraction of savings shared parameter, b, and the minimum quality threshold, $\bar{x}$) as team size N increases. As we show in an appendix, it is straight-forward but tedious to prove that

(14) $db/dN > 0$ and $d\bar{x}/dN > 0$.

Thus, to achieve any given level of cost reduction and quality level, principals will employ higher powered cost and quality incentives in larger ACOs. The reason for this is that the free-riding problem swamps any gains from improved precision in performance measures. This result is, in turn, due to fundamental properties of group incentives and performance measures. Specifically it reflects the fact that the free-riding incentive dilution worsens with $1/N$ while precision improves with $1/\sqrt{N}$.

III. Calibrating the Model

In this section we calibrate our incentive model in order to consider the conditions under which the ACO pay-for-performance scheme will be self-financing. More precisely, we ask under what combinations of cost targets, quality targets and group size will the savings generated by the pay-for-performance incentives be enough to pay for the requisite performance bonuses.

Our calibration proceeds in two steps. First, we derive empirical estimates of the mean and standard deviation of actual clinical quality measures and use these to construct values for our model's key unobserved parameter, σ_q, the noisiness of the quality measure. Plugging this value into our model, we then calculate the maximum ACO size consistent with a self-financing pay-for-performance incentive that achieves a given cost/quality target.

Estimating Mean and Standard Deviation of Clinical Quality Measures

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.[11] 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 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 (a partial list of targeted issues is provided in an appendix).[12] 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

[11] 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.
[12] The system used by the insurer identified 1246 unique gaps in care. The most common gaps involved well-known preventive care guidelines while some of the rarer ones involved more immediately threatening issues.

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 determined from 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 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 Clinical 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 0.43. 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 $\sigma_X = .43/\sqrt{260} = .027$, a result that plays a central role in following calibration exercise.[14]

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

Converting Quality to Money Metric Units

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 the conditional mean of the quality measure as a function of effort devoted to quality, e^q, as $E[X_i|e^q] = 1 - \exp\left(-(1+e^q)_0/\beta\right) = h(e^q)$. 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:

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

Using the exponential transformation described above, the calibration becomes

(16) $\beta = -\frac{1+e^q}{ln\,(1-\mu_X)}$,

(17) $\sigma_q = \sigma_X \frac{\beta}{1-\mu_X}$.

Plugging in the values of μ_X and σ_X and normalizing so that $e^q = 0$ – implying that the target quality level is the status quo- the model parameter is then σ_q =0.0741.

Specifying Cost Targets

With an estimate of σ_q in hand, we turn to specifying cost targets. The socially optimal level of cost containment (given a care quality threshold of $\bar{x}$) could be induced by making a risk neutral physician the residual claimant in a solo practice, i.e. by setting $b = 1$. The physician's payoff in this case is

$$(18) \qquad U(e^c, e^q) = e^c \times 1(e^q \geq \bar{x}) - \frac{(e_i^q + e_i^c)^2}{2}$$

Normalizing $\bar{x} = 0$, the first-order condition for cost control implies the optimal level of cost containment is 1. The level of cost containment effort produced under the ACO style group incentives, $e^c \leq 1$, can thus be interpreted as the fraction of first-best cost-containment effort elicited by the incentives.[15] For calibration purposes, the principal's cost containment targets can be stated as the fraction of first-best cost-control efforts that CMS aims to achieve with pay-for-performance – conditional, of course, on achieving quality thresholds.

The results of our calibration show that for realistic parameter values and savings and quality targets, the incentive benefits of more precise quality measurement holding the size of physician teams fixed are substantial, but achieving these precision improvements by increasing the number of physicians quickly makes the incentive scheme untenable. Figures 1-4 present some results from the model calibration under the assumption that ACOs are trying to achieve modest cost savings equal to 20% of those possible under first-best incentives. Figures 1 and 2 plot optimal sharing and quality thresholds in a solo practice as a function of the precision with which quality is measured. In Figure 1 we observe that as the precision of quality measures increases, the optimal degree of cost sharing falls. This is an intuitive result: the less noisy is the quality measure the more likely it is that cost control efforts will result in a payoff to providers. In contrast Figure 2 demonstrates that as the performance measure becomes more precise, threshold levels of quality are increased. This result is also to be expected because reducing the

[15] This interpretation rests on our assumption of money-metric effort which is in turn equivalent to assuming that the maximum cost savings achievable under first-best incentives is the same as the income the physician would earn as a residual claimant under these incentives.

variance of the performance measure reduces the marginal benefit of providing effort above threshold levels of quality.[16]

Figures 3 plots the principal's optimal cost-sharing parameter and quality threshold as the ACO size increases from 1 physician to 20 under the assumption that the principal is trying to achieve the same modest cost reductions assumed in Figures 1 and 2. Consistent with the comparative statics results from the previous section, we see that the fraction of savings that are shared increases with group size. Notice, however, that as the size of the ACO exceeds 5, the sharing parameter required to achieve the target exceeds one. This surprising result leads us to the primary result of this calibration: even with modest cost and quality targets, it will often be the case that ACO style pay-for-performance incentives may not be self-financing.

We present this result more fully in Figure 4. This figure plots the maximum team size consistent with achieving the presumed cost target and having a self-financing pay-for-performance system (that is, the fraction of shared savings is less than 100 percent). Examining the figure we find that a self-funding pay-for-performance incentive aimed at achieving 5 percent of the cost reductions obtainable by first-best incentives (i.e. those possible with risk neutral physicians working as residual claimants in solo practice) cannot involve a group larger than 15 physicians. Above this size level, the free-riding problem becomes so severe, that the requisite cost containment bonus exceeds the savings generated under such a system. If the principal's aim was to achieve 10 percent of the gains possible under first-best incentives, the maximum size of a self-funded pay-for-performance system would be eight physicians. As the desired level of cost savings rise, the maximum ACO size shrinks dramatically. Cost savings goals aimed at achieving 50% or more of the savings possible under first-best incentives could only be self-

[16] In the extreme case, where $\sigma_q = 0$, there is no marginal benefit to providers of offering e_q above $\bar{x}$ and hence no reason for the principal not to set $e_q = \bar{x}$.

funding under solo-practices. The obvious implication of Figure 4 is that with all but the most trivial cost reduction targets, ACOs with self-funding pay-for-performance systems must operate with under-powered financial incentives.

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) – although to compare their results with ours, we would need to know the caseload per physician not per practice, which they do not show.

The finding that self-financed pay-for-performance incentive schemes for large provider organizations are likely to be under-powered 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 up this issue in the next section of the paper.

IV. Mitigating Strategies for Organizations with Under-powered Incentives.

In this section we analyze the various strategies available to ACOs to augment their under-powered pay-for-performance incentives.

Perhaps the simplest and most straight-forward strategy for augmenting under-powered incentives is to increase the Medicare patient panel size. Increasing the number of Medicare ACO patients in a physician's panel helps in two ways. First, it allows the ACO to attract the requisite number of patients with smaller numbers of physicians, thus reducing the free-riding problem. Secondly, as we have seen in our calibration exercise, improving the signal to noise ratio in our quality measure allows ACOs to achieve cost reduction targets with less revenue sharing. Thus ACO incentives would more likely be self-financing in 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 for ACOs to successfully 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 in regions where care is currently 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.

Another promising possibility for augmenting under-powered incentives is 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 the opportunity cost of the physician's time. A doctor, for example, who spends more of the day in meetings devoted to making care processes more cost-efficient loses the opportunity to see more fee-for-service commercial patients. For physicians who are employees, an 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).

Physicians have intrinsic motives for providing effort and ACO incentive schemes may be able to be far more effective than our model predicts if they can appeal to these motives. One way to stimulate intrinsic motives is through peer pressure and mutual monitoring among professionals in the ACO (Kandel and Lazear, 1992). Encinosa, Gaynor, et al. (2007) offer some theory and evidence that these peer and monitoring effects matter for incentive design in medical groups.[17] A growing literature in economics and psychology suggests, however, that financial incentives may not always coexist easily with intrinsic motivators (Rebitzer and Taylor, 2010). In their extensive review of the experimental literature on public good provision, Bowles and Polania-Reyes (2010) conclude that the effects of incentives on intrinsic motivators depends 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

[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).

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]

From this perspective, structuring incentives within ACOs likely involves paying careful attention to the meaning providers assign to the payments, and it is unclear if positive meanings are 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.

Another possibility for augmenting the under-powered incentives inherent in self-financed pay-for-performance systems is to ask physicians and other providers to post performance bonds. These bonds would be returned (with interest) to providers should the ACO achieve its’ cost and quality targets, but they would be forfeited in the event of failure. The great advantage of performance bonds is that they can greatly magnify the power of pay-for-performance systems. Their great disadvantage is that it might be very difficult to persuade agents to post them and to trust that they will be returned under the right circumstances. In the context of employment relationships this difficulty is often addressed through the device of deferred compensation in the context of long-term employment relationships. Employees post bonds by accepting pay less than their marginal product early in their relationship and this is returned later on in the relationship through severance pay, pensions and other forms of deferred

[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.
[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.

payments. A closely related employment strategy is the efficiency wage strategy under which employees receive a salary greater than their next best alternative. The discounted present value of this pay premium, when combined with a threat to sever relationships should performance targets be missed, would also have the effect of augmenting under-powered pay-for-performance incentives. To the extent that performance bonds and efficiency wage strategies are best implemented in the context of employment relationships, it would seem that they constitute another advantage that conventional organizations have over hybrid organizations in forming successful ACOs.

The mitigating strategies we have discussed so far - job design, training, screening, socialization that gives incentive pay a positive "meaning", deferred compensation and efficiency wages- all have an important element of relationship-specific investments. It is natural to think of these investments as the foundation of a relational contract 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). As we have discussed, 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. In addition, relational contracts can also be used to reduce some of the distortions created by high-powered formal incentives.[20] An important limitation of relational contracts, however, is that the persistent performance benefits they offer are hard to realize within organizations and hard to

[20] 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).

diffuse across organizations.[21] Even if it becomes apparent that successful ACOs depend upon relational contracts to make their pay-for-performance systems work, 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 groups of hospitals and providers to jointly contract with 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. Achieving this scale, however, requires combining physicians and as the numbers of physicians grow so does the free-riding problem. Working with a model of ACO incentives, we establish that the negative effects of free-riding swamp the positive effects of increased precision.

We also calibrate our model using proprietary performance measures from a very large insurer. Our estimates suggest that even minimally sized ACOs with modest cost reduction targets will likely not be self-financing. As a result, successful ACOs will have to find ways to operate with under-powered pay-for-performance incentives augmented by alternative motivational strategies. Some of these complementary strategies can only be implemented in conventional integrated organizations, while others may prove workable in hybrid organizational

[21] 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.

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 quality 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 would operate 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 to participate in ACOs 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 of these is Carrie H. Colla, David E. Wennberg et al. (2012) which 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 of our work for health policy 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 relatively small commercial payers will have an even harder time than Medicare in using pay-for-performance contracts to induce cost-conscious, quality-preserving practices among their physician networks.

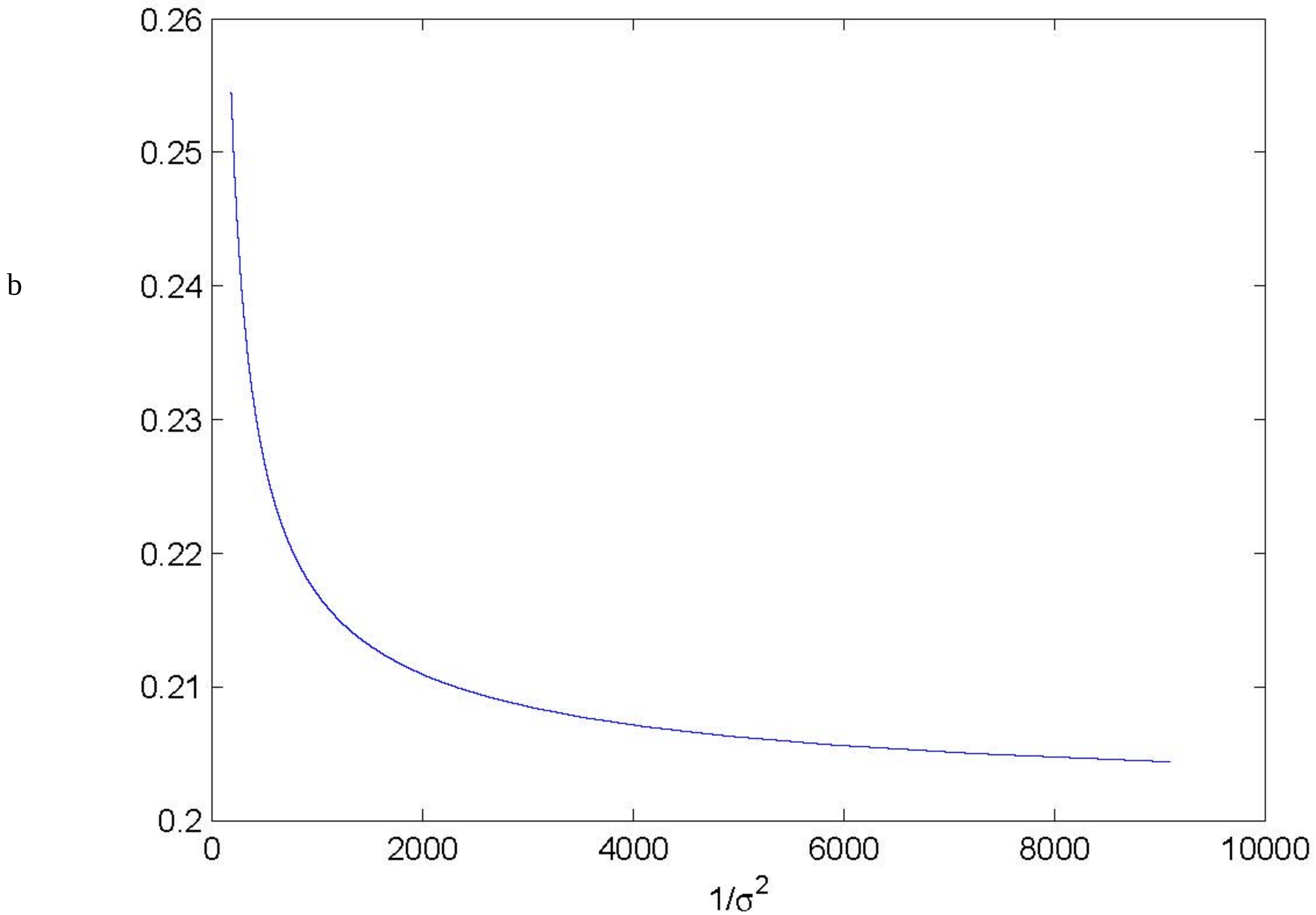
From the perspective of transaction cost economics, we suggest that the study of ACOs offers a rich laboratory for exploring the transaction costs arising from bringing activities *within* an organization. It particularly offers the opportunity to compare how effectively hybrid and conventional organizations can augment under-powered pay-for-performance systems. In much of the US, 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	

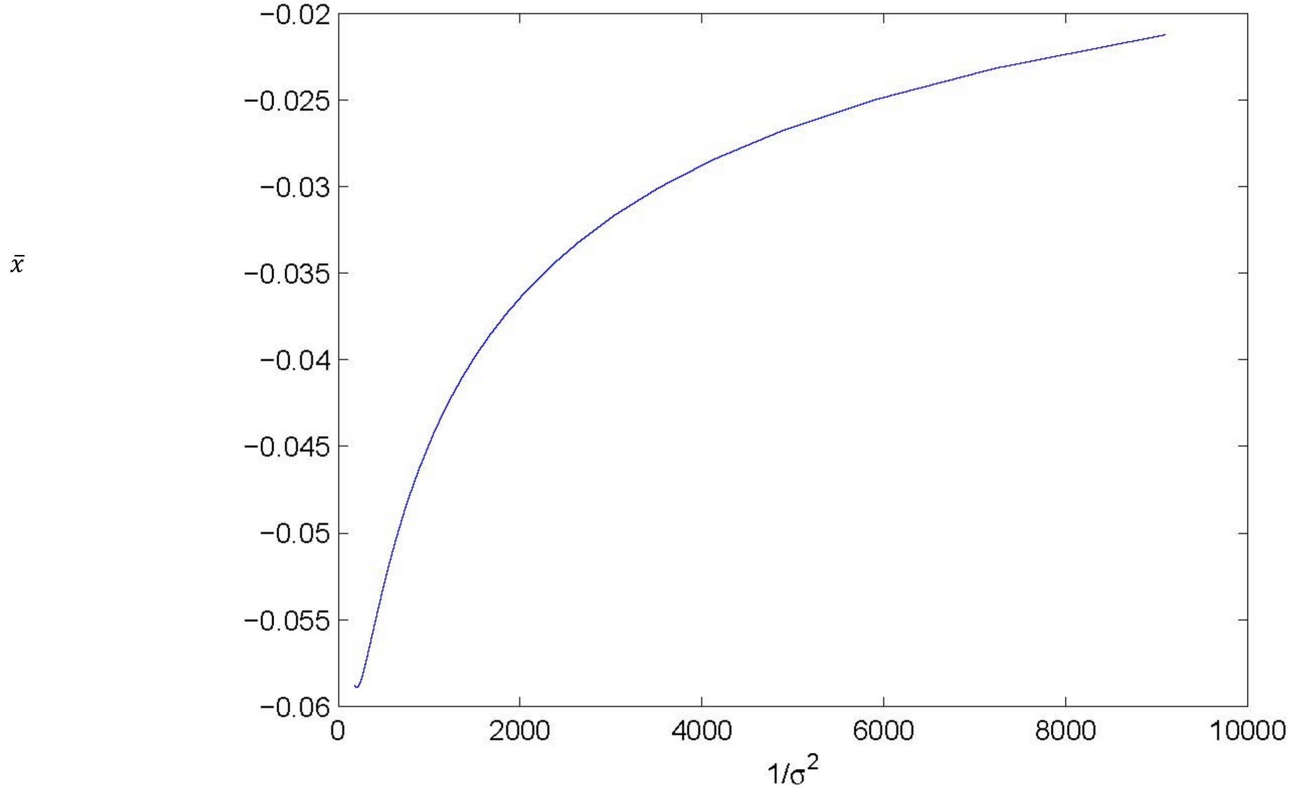
The patient sample contains commercial insurance members whose employers are fully insured, and who have evidence of chronic illness. The provider sample consists of the primary care providers identified as the main providers for these patients on the basis of claims information

Figure 1: Shared Savings by Precision of Performance Measure



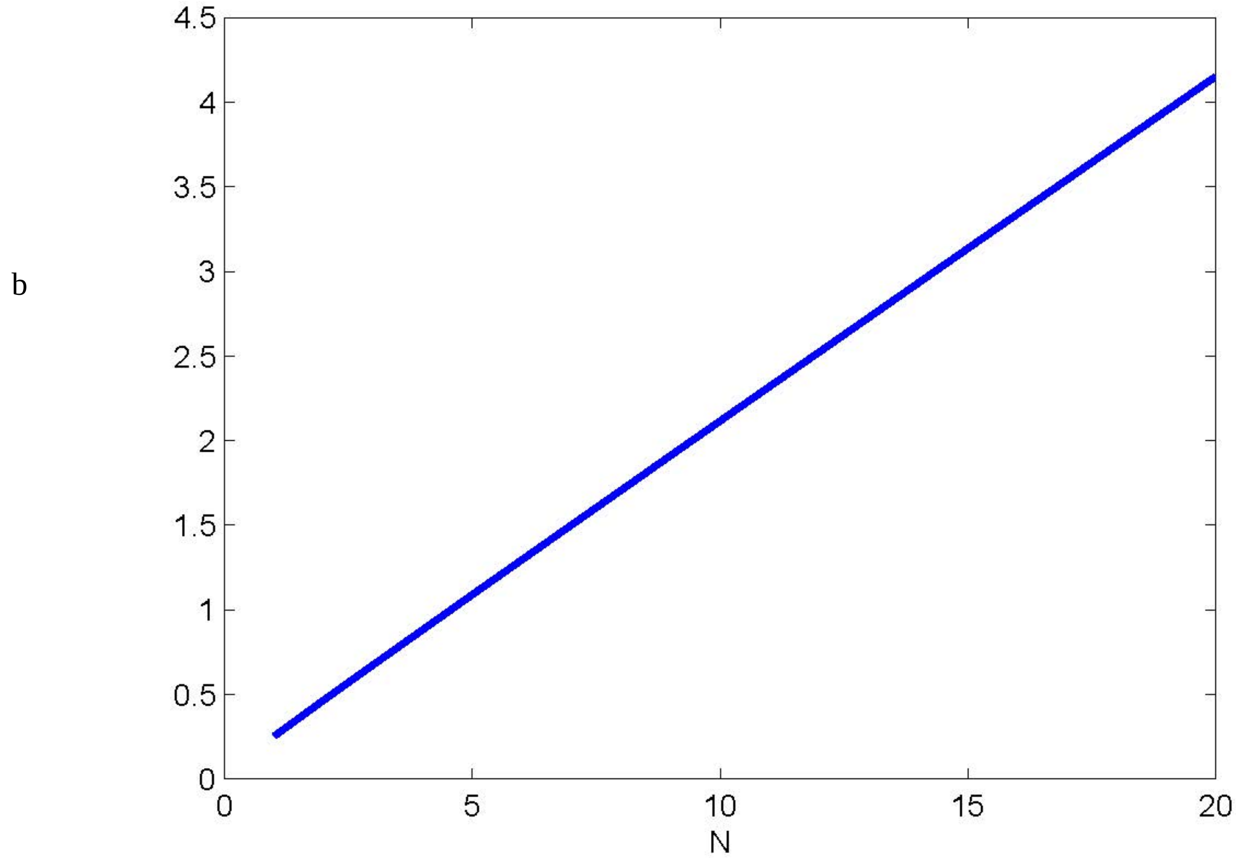
This Figure plots the relationship between the precision of quality measures, $1/\sigma^2$, and b, the fraction of savings that are shared. This graph assumes a risk-neutral physician in a solo practice and that the goal of the principal is to achieve cost savings that are 20% of the savings under first-best incentives. 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; $\sigma_q = 0.0741$. Details in text.

Figure 2: Quality Targets by Precision of Performance Measure



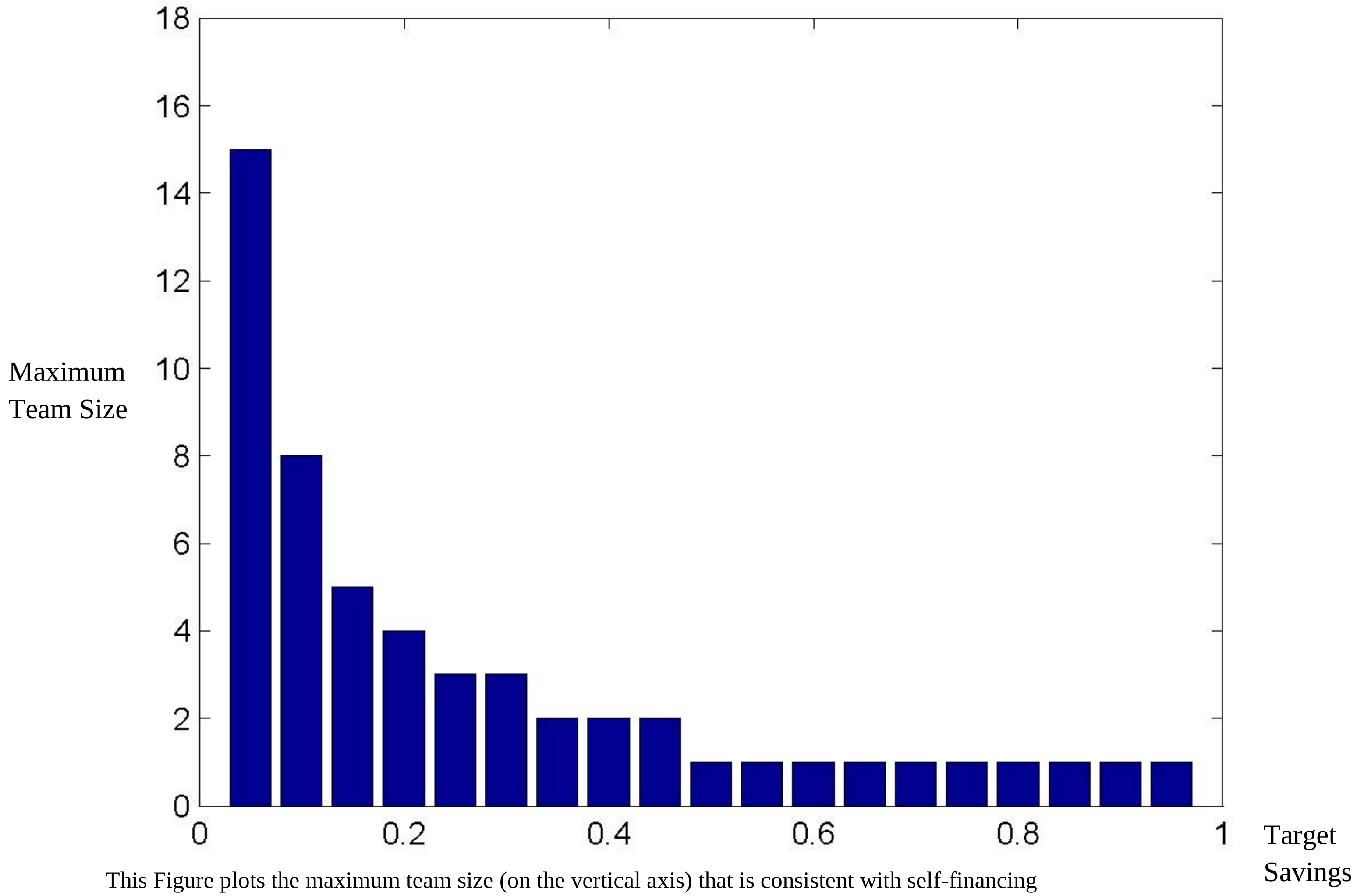
This Figure plots the relationship between the precision of quality measures, $1/\sigma^2$, and $\bar{x}$, the threshold level of quality that must be cleared for savings to be shared. This graph assumes a risk-neutral physician in a solo practice and that the goal of the principal is to achieve cost savings that are 20% of the savings under first-best incentives. 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; $\sigma_q = 0.0741$ under fee-for-service compensation. Details in text.

Figure 3: Optimal Sharing by Team Size



This Figure plots the sharing parameter, b, required to achieve a target level of cost savings by group size, N. The target level of savings are 20% of the savings achievable under first-best incentives. 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; $\sigma_q = 0.0741$. Details in text.

Figure 4: Maximum Self-funding Team Size by Target Cost Savings



This Figure plots the maximum team size (on the vertical axis) that is consistent with self-financing incentives by target cost savings (the horizontal axis). The target level of cost savings are expressed as a percent of the savings achievable under first-best incentives. Savings are only paid out if a quality threshold is cleared. The calibration takes as given the size of each physician's panel of Medicare enrollees (260), the quality threshold, and our estimate of the standard deviation of the noise component of the performance measure; $\sigma_q = 0.0741$. Details in text.

Appendix I: Comparative Static Properties of the Model

How do the level of shared savings, b, and the quality threshold, $\bar{x}$, vary with team size, N, for given desired levels of quality and cost-control effort? Implicit differentiation gives us:

$$\begin{bmatrix} \frac{db}{dN} \\ \frac{d\bar{x}}{dN} \end{bmatrix} = -\begin{bmatrix} F_b^1 & F_{\bar{x}}^1 \\ F_b^2 & F_{\bar{x}}^2 \end{bmatrix}^{-1} \begin{bmatrix} F_N^1 \\ F_N^2 \end{bmatrix} = -\frac{1}{F_b^1 F_{\bar{x}}^2 - F_b^2 F_{\bar{x}}^1} \begin{bmatrix} F_{\bar{x}}^2 F_N^1 - F_{\bar{x}}^1 F_N^2 \\ -F_b^2 F_N^1 + F_b^1 F_N^2 \end{bmatrix}$$

where subscripts denote partial differentiation.

The determinant is positive:

$$F_b^1 F_{\bar{x}}^2 - F_b^2 F_{\bar{x}}^1 = \frac{e^c b}{N\sigma_q^2}\left(\phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right)^2 - \Phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right)\phi'\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right)\right) > 0$$

because the slope of the standard normal density at positive arguments is negative.

The numerator of db/dN simplifies to

$$F_{\bar{x}}^2 F_N^1 - F_{\bar{x}}^1 F_N^2 = \frac{e^c b^2}{\sigma_q^2 N^2}\left(\Phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right)\phi'\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right) - \frac{1}{2}\phi'\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right)^2\right) < 0$$

We therefore have our first result:

$$\frac{db}{dN} = -\frac{F_{\bar{x}}^2 F_N^1 - F_{\bar{x}}^1 F_N^2}{F_b^1 F_{\bar{x}}^2 - F_b^2 F_{\bar{x}}^1} > 0.$$

The fraction of shared savings required to induce a given level of cost and quality effort is increasing in team size.

Turning now to $\bar{x}$:

$$d\bar{x}/dN = -\frac{-F_b^2 F_N^1 + F_b^1 F_N^2}{F_b^1 F_{\bar{x}}^2 - F_b^2 F_{\bar{x}}^1}$$

The denominator of this expression is the determinant identified above and is positive. The numerator is:

$$-F_b^2 F_N^1 + F_b^1 F_N^2 = -\frac{A}{2}\frac{(e^q + e^c)}{N\sqrt{N}}\frac{1}{\sigma_q}\phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right)\left(\frac{e^q}{e^q + e^c} + N\frac{(e^q - \bar{x})}{\sigma_q^2}e^c\right) < 0.$$

From this it follows that $d\bar{x}/dN > 0$. As group size increases, so does the threshold level of quality that must be achieved in order to receive the bonus payment.

In summary, for given desired levels of effort, as the team size increase, both the fraction of savings to be shared and the quality threshold increase.

Appendix II: Quality Measures

The quality measure dataset contains records for potential gaps in care associated with 1,246 specific issues. The following list gives the twenty most frequently occurring issues for which potential gaps in care were detected, collectively accounting for two-thirds of the total potential gaps in care in the dataset:

- Diabetes - Consider Eye Exam
- Heart Protection Study - Consider Adding a Statin
- Breast Cancer Screening - Females 50 Years and Older
- Diabetes - Consider HbA1C Monitoring
- Cervical Cancer Screening - Females Age 21 and older
- Diabetes - Consider Screening for Microalbuminuria
- Breast Cancer Screening - Females 50 Years and Older
- Hyperlipidemia - Primary Prevention - Consider Lifestyle Changes and/or Lipid Lowering Therapy
- Colorectal Cancer Screening - Adults 50 Years and Older
- Diabetes Mellitus - Consider Pneumococcal Vaccine
- Breast Cancer Screening - Females Age 40-49 Years
- Diabetes - Consider Lipid Panel Monitoring
- High Risk Diabetic (HOPE Trial) - Consider Adding an ACE Inhibitor
- Levothyroxine - Consider TSH Monitoring
- Metabolic Syndrome - Consider Treatment
- Concomitant use of SSRIs and NSAIDs increases the risk of GI bleeding
- Diabetes and LDL Greater than 100 - Consider Adding a Lipid Lowering Agent
- Hyperlipidemia (Primary Prevention) - Candidate for a Lipid Lowering Agent
- Age 6-59 mos - Consider Influenza Vaccine
- Statin Use - Consider LFT Monitoring

Quality Measure Regression Results

Dependent Variable: Any Potential Gap in Care

N = 567050

Regressor	Coefficient	Standard Error
hcc1	-0.0047	0.0214
hcc10	0.0499	0.0020
hcc100	0.2130	0.0070
hcc101	0.1147	0.0029
hcc104	0.0750	0.0066
hcc105	0.0482	0.0095
hcc107	0.0282	0.0182
hcc108	0.1556	0.0056
hcc111	0.0852	0.0123
hcc112	0.1531	0.0072
hcc119	0.0113	0.0077
hcc130	0.0638	0.0061
hcc131	0.0991	0.0094
hcc132	0.0303	0.0207
hcc148	0.0662	0.0155
hcc149	0.0545	0.0042
hcc15	0.0181	0.0154
hcc154	0.0432	0.0031
hcc155	0.0270	0.0023
hcc157	0.1137	0.0083
hcc158	0.1626	0.0078
hcc16	0.0262	0.0277
hcc161	0.0831	0.0049
hcc164	0.0037	0.0055
hcc17	-0.0392	0.0262
hcc174	0.0070	0.0119
hcc176	-0.0012	0.0065
hcc177	0.0293	0.0163
hcc18	-0.0125	0.0304
hcc19	-0.0555	0.0128
hcc2	0.0516	0.0083
hcc21	-0.0688	0.0198
hcc25	0.0334	0.0181
hcc26	0.0186	0.0172
hcc27	-0.0110	0.0208
hcc31	0.0340	0.0781
hcc32	0.0249	0.0163

(Quality Measure Regression Results Continued)

hcc33	0.0167	0.0163
hcc37	0.0335	0.0232
hcc38	0.0247	0.0123
hcc44	0.0105	0.0043
hcc45	-0.0566	0.0085
hcc5	0.0430	0.0121
hcc51	0.1067	0.0048
hcc52	0.0577	0.0092
hcc54	0.0862	0.0056
hcc55	0.0842	0.0038
hcc7	0.0239	0.0024
hcc77	0.0352	0.0342
hcc78	-0.0108	0.0302
hcc79	-0.0014	0.0221
hcc8	0.0288	0.0068
hcc80	0.0500	0.0105
hcc81	0.1010	0.0075
hcc82	0.0933	0.0142
hcc83	0.0891	0.0059
hcc9	0.0355	0.0159
hcc92	0.0168	0.0024
hcc95	0.0499	0.0025
hcc96	0.0218	0.0038
hccDM_CHF	0.0330	0.1595
hccDM_CVD	-0.0700	0.0170
hccCHF_COPD	-0.1552	0.0783
hccCOPD_CV~D	-0.1270	0.0300
hccRF_CHF	-0.0505	0.1036
hccRF_CHF_DM	0.0000	(omitted)
hccDM_CHF1	0.0000	(omitted)
hccRF_CHF1	0.0000	(omitted)
hccagecat_35	0.0718	0.0033
hccagecat_45	0.1801	0.0030
hccagecat_55	0.2058	0.0035
hccagecat_60	0.2221	0.0037
hccagecat_65	0.2385	0.0046
hccagecat_70	0.2973	0.0074
hccagecat_75	0.3157	0.0114
hccagecat_80	0.2818	0.0167
hccagecat_85	0.2739	0.0291
hccagecat_90	0.3491	0.0575
hccagecat_95	0.1696	0.1428

(Quality Measure Regression Results Continued)

hccfemale_1	0.0027	0.0029
hccageX~35_1	0.0061	0.0040
hccageX~45_1	-0.0102	0.0037
hccageX~55_1	0.0225	0.0044
hccageX~60_1	0.0228	0.0046
hccageX~65_1	0.0593	0.0061
hccageX~70_1	0.0393	0.0104
hccageX~75_1	0.0381	0.0162
hccageX~80_1	0.0621	0.0231
hccageX~85_1	0.0782	0.0400
hccageX~90_1	-0.0879	0.0807
hccageX~95_1	-0.0554	0.1851
Physician fixed effects		
F(88, 507893)	435.980	
R-squared:	0.1949	
Root MSE	0.4293	

Notes: Least squares coefficients and standard errors from a regression of an indicator for any potential gap in care on the Hierarchical Clinical Condition regressors, which include diagnostic categories interacted with age and gender, controlling for primary care physician fixed effects.

Bibliography

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

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

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

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

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

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

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

Burns, Lawton R. and Mark 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, PhD, MD 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, Randall, James B. Rebitzer, et al. (2008). "Organizational Fragmentation and Care Quality in the US Health Care System." Journal of Economic Perspectives **22**(4): 93-113.

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

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

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

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

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

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

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

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

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

Holmstrom, Bengt and Paul 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, Julia (2012). "Health Policy Brief: Pay for Performance." Health Affairs.

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

Larson, Bridget K., Aricca D. Van Citters, et al. (2012). "Insights From Transformations Under Way At Four Brookings-Dartmouth Accountable Care Organization Pilot Sites." Health Affairs **31**(11): 2395-2406.

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

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

Nyweide, David J., William 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., Gregory C., John Kautter, et al. (2004). "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model." **25**(4).

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

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

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

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

Rosenthal, Meredith B. and Richard 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, Mark C., Pankaj H. Patel, et al. (2011). "A Model For Integrating Independent Physicians Into Accountable Care Organizations." Health Affairs **30**(1): 161-161-172.

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

Song, Zirui, Dana Gelb 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, Paul (1984). The Social Transformation of American Medicine, Basic Books.

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