NBER WORKING PAPER SERIES

STRUCTURING INCENTIVES WITHIN ORGANIZATIONS: THE CASE OF ACCOUNTABLE CARE ORGANIZATIONS

Brigham Frandsen
James B. Rebitzer

Working Paper 20034
http://www.nber.org/papers/w20034

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2014

We would like to thank Aetna, a national health benefits company for its collaboration. We also thank the following people for comments and suggestions that greatly improved the paper: David Cutler, Robert Gibbons, Iver Juster, Daniel Kessler, Patrick Warren, and Julie Wulf as well as seminar participants at the Haas/Sloan Conference on the Law and Economics of Organizations; the NBER conference on The Organization and Productivity of Healthcare Delivery; The Conference on Healthcare Markets at Nortwestern's Kellogg School; the International Health Economics Association; and the Department of Health Care Policy at Harvard Medical School. Rebitzer also acknowledges support for this research from Microsoft. We are, of course, responsible for any remaining errors or oversights. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Structuring Incentives Within Organizations: The Case of Accountable Care Organizations
Brigham Frandsen and James B. Rebitzer
NBER Working Paper No. 20034
April 2014
JEL No. D23,D86,I12,L14,L24,M5

ABSTRACT

Accountable Care Organizations (ACOs) are new organizations created by the Affordable Care Act to encourage more efficient, integrated care delivery. To promote efficiency, ACOs sign contracts under which they keep a fraction of the savings from keeping costs below target provided they also maintain quality levels. To promote integration and facilitate measurement, ACOs are required to have at least 5,000 enrollees and so must coordinate across many providers. We calibrate a model of optimal ACO incentives using proprietary performance measures from a large insurer. Our key finding is that free-riding is a severe problem and causes optimal incentive payments to exceed cost savings unless ACOs simultaneously achieve extremely large efficiency gains. This implies that successful ACOs will likely rely on motivational strategies that amplify the effects of under-powered incentives. These motivational strategies raise important questions about the limits of ACOs as a policy for promoting more efficient, integrated care.

Brigham Frandsen
Department of Economics
Brigham Young University
Provo, UT 84602
frandsen@byu.edu

James B. Rebitzer
Professor of Management, Economics, Public Policy
Markets, Public Policy and Law Department
Boston University School of Management
595 Commonwealth Ave.
Boston, MA 02215
and NBER
rebitzer@bu.edu

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 economic questions. 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 shared savings 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. One study finds, for example, that primary care physicians had annual median caseloads of 260 Medicare patients (Nyweide 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 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 usually 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 physicians operating 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 relative to cost benchmarks 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; that is, the savings they produce will not cover the costs of the performance bonuses without extremely large economies of scale or productivity improvements within ACOs. It follows that ACOs committed to self-financed pay-for-performance will likely 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.

The difficulty of implementing these complementary motivational strategies determines the cost of transacting care delivery *within* organizations as opposed to markets. 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 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.

1. 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 et al., 2008). Individual patients are frequently treated by numerous care providers who have only weak organizational ties

[1] Masten et al. (1991) identify a similar gap in the literature.
[2] Much of this section is adapted from Rebitzer and Votruba (2011) review of the organizational economics of physician's practices.

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] 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 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

[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).

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 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

[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).

[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

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 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 et. al (2011)

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 et al. (2011), Shortell and Casalino (2010), Singer and Shortell (2011), and Song et al. (2011) .

[7] The exact nature of the payments to ACOs varies a good deal. All ACOs face a cost benchmark 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).

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 an organizational perspective, 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 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 hospital delivery system that employs 475 doctors and owns five

[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 are larger provider groups with experience in coordinated care started Medicare Pioneer ACOs.

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 additional upward pressure on the scale of ACOs.

2. 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 providing 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 cost benchmarks make it possible for providers to profit by not providing care to patients.

For this reason ACO contracts specify that providers can keep some portion of savings below the cost benchmark 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 cost benchmark 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.

Average 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, η_i^c:

$$C_N = \frac{1}{N}\sum_{i=1}^{N}(C + \eta_i^c - e_i^c)$$

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

$$Q_N = \frac{1}{N}\sum_{i=1}^{N}\left(e_i^q - \eta_i^q\right)$$

[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.

The cost and quality disturbances are not observable, have mean zero and variance σ_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 (or losses) relative to baseline, but receives a positive payout only if average quality exceeds a specified threshold $\bar{x}$. Doctors are risk neutral and maximize expected income minus effort costs. Doctor *i*'s payoff is therefore

$$U_i = E\big[b(C - C_N)\big(1 - 1(Q_N < \bar{x})1(C - C_N > 0)\big)\big] - \frac{\left(e_i^q + e_i^c\right)^2}{2} \tag{1}$$

The first term in the utility function in effect assumes that individual physicians are risk-neutral in bonus income. This assumption may be surprising given the importance of risk pooling in the structuring of ACOs - but organizations pool risk because of liquidity and credit constraints and the consequent risk of insolvency should an ACO incur much higher than expected costs. Our paper does not analyze these enterprise-level risk-management concerns. Rather we focus on how organization level incentives influence the actions of individual physicians. In this context, risk neutrality most realistically captures physicians' preferences over the relatively small individual payouts at stake in ACOs. [10] Introducing risk aversion into the model would also greatly complicate the analysis much without adding insight. The core question of our paper concerns incentive

[10] This is consistent with the conventional economic theory of insurance which finds that rational individuals ought to be approximately risk neutral over gambles that are small relative to individual net worth.

and moral hazard problems and these are handled entirely through our analysis of the variable component of compensation - just as they would be if the model allowed for risk-averse agents. Introducing risk aversion, however, would also require us to derive both the variable and fixed component of compensation because this latter component of compensation provides income insurance for risk-averse agents. Thus by abstracting from risk aversion, we are greatly simplifying the analysis without altering our conclusions in a substantive way.

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. This specification also implies that physicians are intrinsically motivated to supply minimum levels of effort to cost and quality, but additional incentives are required to move them beyond these levels.[11]

Invoking a central limit theorem, the utility function can be rewritten as:

$$U_i \underset{p}{\rightarrow} b\left[\frac{1}{N}\sum_{j=1}^{N} e_j^c - \left(\frac{1}{N}\sum_{j=1}^{N} e_j^c + \frac{\sigma_c}{\sqrt{N}}\frac{\phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^c\right)}{\Phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^c\right)}\right)\left(1 - \Phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^q - \sqrt{N}\frac{\bar{x}}{\sigma_q}\right)\right)\Phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^c\right)\right] - \frac{\left(e_i^q + e_i^c\right)^2}{2},$$

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

[11] In this set-up, the intrinsic levels of effort devoted to quality and cost control are respectively normalized to zero.

$$\frac{\partial U}{\partial e_i^c} = \frac{b}{N}\left[1 - \Phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^c\right)\left(1 - \Phi\left(\frac{1}{\sigma_q\sqrt{N}}\sum_{j=1}^{N} e_j^q - \sqrt{N}\frac{\bar{x}}{\sigma_q}\right)\right)\right] - \left(e_i^q + e_i^c\right)$$

$$= 0$$

$$\frac{\partial U}{\partial e_i^q} = \frac{b}{\sigma_q\sqrt{N}}\left(\Phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^c\right)\frac{1}{N}\sum_{j=1}^{N} e_j^c\right.$$

$$\left. + \frac{\sigma_c}{\sqrt{N}}\phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^c\right)\right)\phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^q - \sqrt{N}\frac{\bar{x}}{\sigma_q}\right) - \left(e_i^q + e_i^c\right)$$

$$= 0$$

Consider a symmetric equilibrium where all doctors choose the same effort levels, e^q, e^c. Then provided the second order conditions are satisfied (which we verify is the case for the range of parameters we consider) the equilibrium conditions determining doctors' effort are:

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

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

For policy purposes we treat the Center for Medicaid and Medicare Services (CMS) as the principal but the same logic would apply to any private payer. 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. 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. With this set-up in place, we can then answer our main theoretical question: how does the required level of shared savings, b, change as the team size, N, increases? As we show in an appendix, it is straight-forward to show that

$$db/dN > 0.$$

Thus, to achieve any given level of cost reduction and quality level, principals must employ higher-powered cost 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 $1/\sqrt{N}$.

3. 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 two key parameters in the model, σ_c and σ_q, the standard deviations of the cost and quality measures, respectively, using data on patient health care costs and actual clinical quality measures. Plugging these values 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 the Standard Deviation of Clinical Cost and Quality Measures

Our observed cost and quality measures are derived from confidential insurance records on roughly a million chronically ill, commercial insurance members with health insurance from commercially insured employers. These data are well suited for this exercise in that the insurer combined billing records with data from pharmacies and labs to construct an ersatz electronic medical record for each patient that enables the construction of detailed cost and quality measures. An important limitation of this data is that they do not include information on Medicare patients. This limitation may not be that important for our purposes because: (1) this commercial population suffers from many of the chronic conditions afflicting a Medicare population; and (2) our cost and quality regressions include quite detailed clinical controls. In this regard it is also important to remember that commercial populations are themselves very relevant to the phenomenon we study because ACOs are expected to contract with both Medicare and commercial insurers.

Our cost measure is constructed by applying Medicare reimbursement rates to the diagnostic and procedure codes associated with each claim, and then summing within each patient. Our cost measure is intended not to reflect the actual amount paid for the care, which will depend on specific negotiations between providers and the insurer, but rather the resource costs of care. Our quality measure is also constructed from the individual claims. The claims records are 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 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,

[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.

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.

<<COMP: Place Table 1 about here>>

As reported in Table 1, average patient costs in our sample are $8,008, and the mean of *Any Potential Gap in Care* is 0.29. To calibrate our model we also require an estimate of the noise with which care cost and quality are measured. We base these on the residual variance from regressions of $Costs$ and *Any Potential Gap in Care* on variables for age and gender, physician fixed effects, 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 et al., 2004). The regression we estimate relates the cost or quality measure for patient i, treated by primary care physician j, to demographic and underlying clinical variables X_{ij} and a physician-specific effect α_j:

$$Y_{ij} = X_i'\beta + \alpha_j + \epsilon_{ij},$$

where Y_{ij} is either patient costs or the quality indicator. From this exercise we find that the root mean squared error is $10,976 for the cost regressions and 0.430 for potential gaps in care. Fischer et. al., (2009) report that the average physician group has 260

[13] Costs and 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.

Medicare patients. Adopting this as the relevant sample size for each physician, it follows that an empirical measure for the physician-level cost noise is $\$10{,}976/\sqrt{260} = \$681$, and for the physician level quality measure noise is $\sigma = .43/\sqrt{260} = .027$.

Converting Empirical Measures to Model Units

The final task is to convert the empirical measures of cost and quality noise to the money-metric units of the model. To do this, we first observe that money-metric effort implies effort and noise units are measured in terms of first-best cost-saving effort. The easiest way to see this is to consider optimal physician effort if team size were N=1, quality was at the status quo level $e^q = 0$, and physicians keep all cost savings induced by their efforts, $b = 1$. Under these conditions, equation (1) implies that first-best effort level occurs when $e^c = 1$. Since both the noise and effort terms are added together in the model, they must be in the same units – hence the money metric noise term is measured in terms of first best physician effort levels.

With this as background, converting observed estimates of noise in costs to money-metric noise simply requires an assumption about how much costs would fall if physicians operated under first-best cost reduction incentives. We assume in what follows that this number is 30 percent. [14] Remembering that average expenditures in our sample are \$8,008, it follows that a unit of money metric effort is equivalent to 0.3*8008 =\$2402.40. Using this conversion factor, our estimate of σ_c=\$681 becomes 0.283 in the money-metric units of the model.

[14] Alternative assumptions of 10 percent or 50 percent change the simulation results only slightly (available upon request).

Converting the empirical measures for the noise in the care quality measure to the money-metric units in the model requires an additional assumption about the functional form relating the empirical quality measures to money-metric effort devoted to quality. We posit that the exponential distribution is a reasonable functional form because the transformation likely exhibits decreasing returns and a ceiling to the observed measure of quality. More formally, we can write the conditional mean of the quality measure as a function of effort devoted to quality, e^q, as $E[X_i|e^q] = 1 - \exp\left(-(1+e^q)/\beta\right) = h(e^q)$. The principal's money-metric effort is just the inverse of this transformation: $x_i = h^{-1}(X_i)$. This framework implies a locally linear mapping from outcome-denominated variance to input-denominated variance at the normalized status quo-input level. Applying a delta-method type approximation to convert the empirically observed mean and standard deviation of the quality measure, μ and σ, to the money metric quality noise parameter needed to calibrate the model, σ_q:

$$\sigma_q = \sigma_X h^{-1\prime}(\mu).$$

Using the exponential transformation described above, the calibration becomes

$$\sigma_q = \sigma \frac{1 + e^q}{(\mu - 1)\, ln(1-\mu)}.$$

Plugging in the values $\mu = 1 - .288 = .712$ and $\sigma = .027$ and normalizing so that e^q = 0 -implying that the target quality level is the status quo - the model's quality noise parameter is then σ_q =0.0741.

With our empirically-based estimates for the model's money-metric parameters in hand, we can proceed with the calibration. We find that for reasonable savings and quality targets, increasing the number of physicians quickly makes the incentive scheme untenable.

<<COMP: Place Fig. 1 about here>>

Figure 1 plots the principal's optimal cost-sharing parameter as the ACO size increases from 1 physician to 20 under the assumption that the principal is trying to achieve 20 percent of the first-best cost reductions. If, for example, first-best incentives yielded a 30% cost savings, the principal will be designing group incentives for the ACO to achieve a 6% reduction in costs. 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 highlights 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.

<<COMP: Place Fig. 2 about here>>

We present this result more fully in Figure 2. 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 17 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-funding under solo-practices. The obvious implication of Figure 2 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.

The analysis so far has assumed that team production is no different from individual production. In all likelihood, however, there may be efficiency gains from combining physicians into ACOs which may make self-financed incentives more feasible. Our calibrated model allows us to quantify how great the efficiency gains from team production would have to be in order for the required incentives to be self-financing. We do this by introducing a multiplier, $m(N)$, on cost-savings effort, so that physicians in a team of size N who each exert cost-savings effort e^c actually realize cost savings equal to $m(N) \times e^c$. This is a simple way of capturing the possibility that team production amplifies the benefits of each member's efforts. We do not unfortunately have estimates of $m(N)$, but our model allows us to ask how big the multiplier would have to be for each team size in order for incentives to be self-financing.

<<COMP: Place Fig. 3 about here>>

Our calibrations show that for the parameters used above, teams of reasonable size would need efficiency multipliers that are very large relative to estimates of the efficiency gains from team production in the health care literature. Figure 3 plots the efficiency multiplier, $m(N)$, that would be necessary for the ACO incentive scheme to be self-financing for team sizes from 1 to 20 physicians. As before, the savings target is set at 20 percent of the first-best level of savings, with quality being set at the status quo, and noise parameters as calibrated above. For teams of 4 or fewer physicians, the multiplier could actually be less than one, since as we saw in Figure 1, the required payouts for teams this small are already less than the generated savings. For larger teams, however, the required efficiency multiplier is greater than one and grows roughly proportionally to the team size. At $N = 10$ the required multiplier is greater than 200 percent, and at $N = 20$ the required multiplier is greater than 450 percent. Are team efficiencies of this magnitude realistic? Many studies have compared health care spending in multispecialty and integrated group practices to national averages and have found efficiencies up to about 30 percent (see Tollen, 2008 for a summary). More recently, Berwick and Hackbarth (2012) estimated that total waste in the health care system increased total spending by 21% in the low estimate to 47% in the high estimate. These studies do not speak directly to our model of incentives to reduce costs, but they do suggest that the likelihood of achieving effort efficiencies on the order of those in Figure 3 may be quite low.

<u>Sensitivity of Results to Assumptions:</u>

Empirical exercises of the sort we have conducted so far depend critically on key assumptions. 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, Gottlieb, Casalino and Fisher (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.

Similarly, we were concerned about the sensitivity of our results to sampling variation leading to different estimates of parameters μ, σ_c, and σ_q. To examine the role of sampling variation we bootstrapped our model and found almost no difference at all between calibrations using the minimum and maximum estimates. The unimportance of sampling variation makes sense given the very large samples we use to construct these estimates.

The model calibrations also required an assumption on the fraction of costs that could be saved under first-best incentives, which we set at 30 percent. Repeating the analyses for alternative assumptions of 10 percent and 50 percent produced only very slight changes, and no change at all in the qualitative conclusions.

Finally we were concerned about the sensitivity of our results to the fact that we weighted all potential gaps in care equally regardless of severity. To assess the importance of this assumption, we also experimented with replacing the quality measure described in the text with an alternative quality measure that gives greater weight to more severe potential gaps in care. The severity weighted measure produces bonus estimates that are very similar to those described in the text.[15]

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

The finding that self-financed pay-for-performance incentive schemes for large provider organizations are likely to be under-powered suggests that successful ACOs will have to make use of complementary motivators that augment the influence of pay-for-performance financial incentives.

In this section of the paper, we use our model to describe how two such complementary motivators would work. The first additional incentive instrument we consider is a performance bond posted by the provider. These bonds would be returned (with interest) to providers along with a payout based on realized savings should the ACO achieve its quality targets, but they would be forfeited in the event of failure.

More formally suppose that physicians are persuaded to post performance bonds of magnitude s with the principal or in escrow at the beginning of the period. At the end of the period, after team costs and quality outcomes are realized, the bond s is returned to each physician if the quality threshold was met, and not otherwise. The principal also pays out an amount equal to $b(C - C_N) - s$, unless the quality threshold was not met and $C - C_N > 0$, in which case the payout is zero. Since the marginal impact of effort on the physician's net payout is unaffected by the performance bond, the physicians' first-order

[15] The insurer gave each potential gap a severity code ranging from one to four. Applying these weights to Potential Gaps in Care, we ran the same regressions described in the text and used the mean square error from that regression to calibrate a new value of σ_q=0.095. This new parameter value produced results estimates very close to those in the text and group production multipliers that are, in fact, larger, so the approach described in the text was – if anything- conservative.

conditions and the principal's choice of b and $\bar{x}$ remain the same. Setting the amount of the performance bond at

$$s = b\left[e^c - \left(e^c + \frac{\sigma_c}{\sqrt{N}}\frac{\phi\left(\frac{e^c}{\sigma_c}\sqrt{N}\right)}{\Phi\left(\frac{e^c}{\sigma_c}\sqrt{N}\right)}\right)\left(1 - \Phi\left(\sqrt{N}\frac{e^q - \bar{x}}{\sigma_q}\right)\right)\Phi\left(\frac{e^c}{\sigma_c}\sqrt{N}\right)\right] - e^c,$$

where e^c and e^q are the cost and quality effort levels the principal wishes to induce, ensures that the average equilibrium payout is equal to e^c; that is, the payout is exactly financed by the savings the ACO generates.

The advantage of performance bonds is that they can greatly magnify the power of self-financing 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 organizational economics literature 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 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. The future commitments to handling the bonds and dismissal decisions fairly are presumably enforced by the organization's desire to maintain its reputation as a reliable counterparty for these sorts of agreements.

Another 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 via job design. The employer can simply mandate that the physician has to spend certain hours on process improvements and cannot see patients during that time. 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 achieve high levels of coordination and motivation with low powered incentives (Roberts, 2004).

To capture this idea formally in the model, suppose the principal can set the scope of a provider's job, denoted by $\delta \in (0,1]$. Setting δ to be very small corresponds to a narrow job scope and therefore a small opportunity cost of time. Setting $\delta = 1$ gives complete latitude to the provider and therefore corresponds to a high opportunity cost of time devoted to, say, process improvements. Introducing the job design parameter, the provider's utility function under the ACO incentive scheme becomes

$$U_i \underset{p}{\rightarrow} b\left[\frac{1}{N}\sum_{j=1}^{N} e_j^c - \left(\frac{1}{N}\sum_{j=1}^{N} e_j^c + \frac{\sigma_c}{\sqrt{N}} \frac{\phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^c\right)}{\Phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^c\right)}\right)\right.$$

$$\left. \times\left(1-\Phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^q - \sqrt{N}\,\frac{\bar{x}}{\sigma_q}\right)\right)\Phi\left(\frac{1}{\sigma_c\sqrt{N}}\sum_{j=1}^{N} e_j^c\right)\right]$$

$$-\,\delta\,\frac{\left(e_i^q + e_i^c\right)^2}{2}.$$

This utility function is the same as before except the quadratic effort cost term is multiplied by δ, capturing the notion that job design can reduce the opportunity cost of providing effort.

<<COMP: Place Fig. 4 about here>>

We can use this adapted model and the calibrated parameters to see how adjusting the opportunity cost of effort through job design ameliorates the ACO incentive problem. Figure 4 plots the required level of cost sharing for a physician team of size $N = 20$ for values of δ, the opportunity cost of effort parameter, ranging from zero to one. As before, the simulation assumes a cost savings target equal to 20 percent of the first-best savings while maintaining quality at the status quo. The figure shows that for values of δ near one, physicians would have to receive over 400 percent of the generated savings in order for incentives to be properly aligned. This result was also evident in Figure 1 which implicitly set $\delta = 1$. For smaller levels of the opportunity cost of effort, however, the required shared savings fraction is smaller, and falls below one when δ is less than 25 percent or so.

These calculations are, of course, purely illustrative. They demonstrate how reducing the opportunity cost of effort through job design can allow ACOs to succeed with lower-powered, self-financing incentives or, if that is infeasible, with smaller performance bonds. To the extent that performance bonds become more difficult to implement as they grow in size, this example also illustrates the complementary nature of the entire bundle of motivators available to employers in conventional employment relationships. Job design makes performance bonds more workable which in turn enhances the effectiveness of under-powered pay for performance systems.

Performance bonds and job design are, of course, not the only complementary motivational instruments that ACOs can employ to augment their under-powered group incentives. Another possibility, often described in the management literature as high-performance human resource systems, combines job design with training, screening, and socialization to motivate employees working under low-powered incentives (Holmstrom and Milgrom, 1991; Roberts, 2004). Related motivational mechanisms discussed in the personnel economics literature include peer pressure and mutual monitoring (Kandel and Lazear, 1992; Encinosa, Gaynor, et al., 2007). These complementary motivational strategies all have an important element of relationship-specific investments and so 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). In addition there is a growing body of theoretical argument and empirical research that highlights powerful psychological motivators and the sometimes problematic interaction of intrinsic motives and financial rewards (Rebitzer and Taylor, 2010; Bowles and Polania-Reyes, 2010).

It would not be difficult to expand our model to include each of the complementary motivational strategies we have sketched out in the preceding paragraph. Doing so, however, would greatly complicate the analysis while obscuring the point we make in our analysis of performance bonds and job design: that it is theoretically possible for successful ACOs to implement complementary motivational strategies that are sufficient to resolve the free-riding issues that plague principal-agent relationships in large groups.

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 to provide care to a population of enrollees in an environment that rewards cost-efficiency through shared savings 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 generally not be self-financing unless extremely large economies of scale or productivity improvements accompany ACOs. 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 are best 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, the key parameters of the model are estimated from a population of chronically ill commercial insurance patients and the 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. In considering this limitation, it should be noted, however, that one of the goals of the ACO program is to promote the use of population based pay for performance in commercial populations as well. 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: cost benchmarks and bonuses linked to noisy cost and quality measures averaged over groups of physicians. 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 the 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 precise 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, optimal 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 et al. (2012) which describes a demonstration project for large physician groups. Comparing our calibration with the incentive payouts used in the actual experiment suggests that the incentives in these experiments were 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 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. To the extent that relatively small commercial payers will require larger groupings of physicians to achieve target panel sizes, they will likely 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 transaction cost economics, our main conclusion is that free-riding in large groups creates serious incentive problems but that it is also possible for ACOs to implement complementary motivational strategies that mitigate these problems. Conventional organizations routinely use these complementary strategies to solve team production and incentive problems and variations in the effectiveness of these complementary strategies are likely to be an important determinant of the costs of transacting within an organization. What is not conventional or routine about ACOs is the fragmented state of the US health care system. In much of the US, it is hard to imagine the large-scale migration of independent practice physicians into truly integrated organizations. In this context, a large fraction of ACOs will have to begin their existence by working with providers operating within hybrid organizations that sit somewhere on a continuum between true integration and a collective of small independent practices.

Understanding which augmenting motivational strategies are most effective at different points on this continuum is therefore critical for understanding the potential of ACOs as a policy for improving the efficiency of the US health care system. If hybrid organizations can find ways to implement complementary motivational strategies as effectively as conventional organizations, ACOs might be a very effective tool for improving efficiency. If not, then gains from ACOs may be limited to regions where care is “mostly integrated” already.

Appendix A: 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? From the principal's point of view, b and $\bar{x}$ are jointly determined by the agents' first order conditions evaluated at the symmetric equilibrium:

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

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

Implicit differentiation of this system gives us:

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

where subscripts denote partial differentiation.

To sign this, we compute and sign each element:

$$F_b^1=\frac{1}{N}\left[1-\Phi\left(\sqrt{N}\frac{e^c}{\sigma_c}\right)\left(1-\Phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q-\bar{x})\right)\right)\right]$$

$$>0.$$

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

$$<0.$$

.

$$F_N^1 = -\frac{1}{N}\left(1 - \frac{1}{2}\frac{(e^q - \bar{x})}{e^c}e^q + e^c\right.$$

$$+ \frac{b}{N}\frac{1}{\sigma_q}\left(\frac{1}{2}\frac{e^q - \bar{x}}{e^c}\sigma_c\phi\left(\sqrt{N}\frac{e^c}{\sigma_c}\right)\phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right)\right.$$

$$\left.\left. + \frac{1}{2}\sqrt{N}\frac{e^c}{\sigma_c}\sigma_q\phi\left(\sqrt{N}\frac{e^c}{\sigma_c}\right)\left(1 - \Phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right)\right)\right)\right)$$

$$< 0,$$

provided $e^q > \bar{x}$ and $e^c > \frac{e^q - \bar{x}}{2}$. The first condition holds whenever the first order conditions hold, and emerges from the simple mathematics of threshold incentives, and is true whenever the first order conditions are satisfied. [16] The intuition for the second condition is that the monetary reward to quality effort is proportional to the cost-savings effort, so the target cost savings effort must be sufficiently high to induce quality effort. This condition constitutes a restriction on the set of cost- and quality-efforts the principal will be able to induce, but is not binding in the natural setting we consider in which the principal wishes to cut costs while maintaining quality.

Continuing,

---

[16] To see the intuition, 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 $\bar{x}$.

$$F_b^2 = \frac{1}{\sigma_{q\sqrt{N}}}\left(e^c + \frac{\sigma_c}{\sqrt{N}}\frac{\phi\left(\sqrt{N}\frac{e^c}{\sigma_c}\right)}{\Phi\left(\sqrt{N}\frac{e^c}{\sigma_c}\right)}\right)\Phi\left(\sqrt{N}\frac{e^c}{\sigma_c}\right)\phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right)$$

$$> 0.$$

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

$$> 0$$

provided $e^q > \bar{x}$. Finally,

$$F_N^2 = -\left(\frac{b}{N}\frac{e^c}{\sigma_q}\left(\frac{1}{2\sqrt{N}}\Phi\left(\sqrt{N}\frac{e^c}{\sigma_c}\right) + \frac{1}{N}\frac{\sigma_c}{e^c}\phi\left(\sqrt{N}\frac{e^c}{\sigma_c}\right)\right)\right.$$

$$\left. + \frac{1}{2}\frac{b}{\sqrt{N}}\frac{(e^q - \bar{x})^2}{\sigma_q^2}\left(\frac{e^c}{\sigma_q}\Phi\left(\sqrt{N}\frac{e^c}{\sigma_c}\right) + \frac{1}{\sqrt{N}}\frac{\sigma_c}{\sigma_q}\phi\left(\sqrt{N}\frac{e^c}{\sigma_c}\right)\right)\right)\phi\left(\frac{\sqrt{N}}{\sigma_q}(e^q - \bar{x})\right)$$

$$< 0.$$

Given the sign of the components, we have that

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

Appendix B: 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

References

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), 1137-60.

Berwick, D. M. and A. D. Hackbarth. 2012. "Eliminating Waste in Us Health Care." *JAMA*, 307(14), 1513-16.

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), 2407-16.

Carrie H. Colla, PhD; MD David E. Wennberg, MPH; PhD Ellen Meara; PhD Jonathan S. Skinner; MS Daniel Gottlieg; PhD Valerie A. Lewis; PhD Christopher M. Snyder and MD Elliott S. Fisher, MPH. 2012. "Spending Differences Associated with the Medicare Physician Group Practice Demonstration." *Journal of the American Medical association,* 308(10), 1015-23.

Cebul, Randall; James B. Rebitzer; Lowell J. Taylor and Mark Votruba. 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.

____. 2011. "The Accountable Care Organization: Whatever Its Growing Pains, the Concept Is Too Vitally Important to Fail." *Health Affairs*, 30(7), 1250-55.

Fisher, Elliott S.; Mark B. McClellan; John Bertko; Steven M. Lieberman; Julie J. Lee; Julie L. Lewis and Jonathan S. Skinner. 2009. "Fostering Accountable Health Care: Moving Forward in Medicare." *Health Affairs*, 28(2), W219-W31.

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), 263–88.

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.

Kandel, Eugene and Edward P. Lazear. 1992. "Peer Pressure and Partnerships." *Journal of Political Economy*, 100(4), 801-17.

Masten, Scott E.; James W. Meehan and Edward A. Snyder. 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), 2363-67.

Nyweide, David J.; William B. Weeks; Daniel J. Gottlieb; Lawrence P. Casalino and Elliott S. Fisher. 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), 2444-50.

Pope, G. C.; J. Kautter; R. P. Ellis; A. S. Ash; J. Z. Ayanian; L. I. Lezzoni; M. J. Ingber; J. M. Levy and J. Robst. 2004. "Risk Adjustment of Medicare Capitation Payments Using the Cms-Hcc Model." *Health Care Financ Rev*, 25(4), 119-41.

Rebitzer, James B. and Lowell J. Taylor. 2010. "Extrinsic Rewards and Intrinsic Motives: Standard and Behavioral Approaches to Agency and Labor Markets," O. Ashenfelter and D. Card, *Handbook of Labor Economics.* North-Holland, 701-72.

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.

Ryan, Andrew M and Tim Doran. 2012. "The Effect of Improving Processes of Care on Patient Outcomes: Evidence from the United Kingdom's Quality and Outcomes Framework." *Medical Care*, 50(3), 191-99.

Shields, Mark C.; Pankaj H. Patel; Martin Manning and Lee Sacks. 2011. "A Model for Integrating Independent Physicians into Accountable Care Organizations." *Health Affairs*, 30(1), 161-72.

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), 1747-48.

Shortell, Stephen M.; Lawrence P. Casalino and Elliott S. Fisher. 2010. "How the Center for Medicare and Medicaid Innovation Should Test Accountable Care Organizations." *Health Affairs*, 29(7), 1293-98.

Singer, Sara and Stephen M. Shortell. 2011. "Implementing Accountable Care Organizations." *JAMA: The Journal of the American Medical Association*, 306(7), 758-59.

Song, Z.; D. G. Safran; B. E. Landon; Y. He; R. P. Ellis; R. E. Mechanic; M. P. Day and M. E. Chernew. 2011. "Health Care Spending and Quality in Year 1 of the Alternative Quality Contract." *N Engl J Med*, 365(10), 909-18.

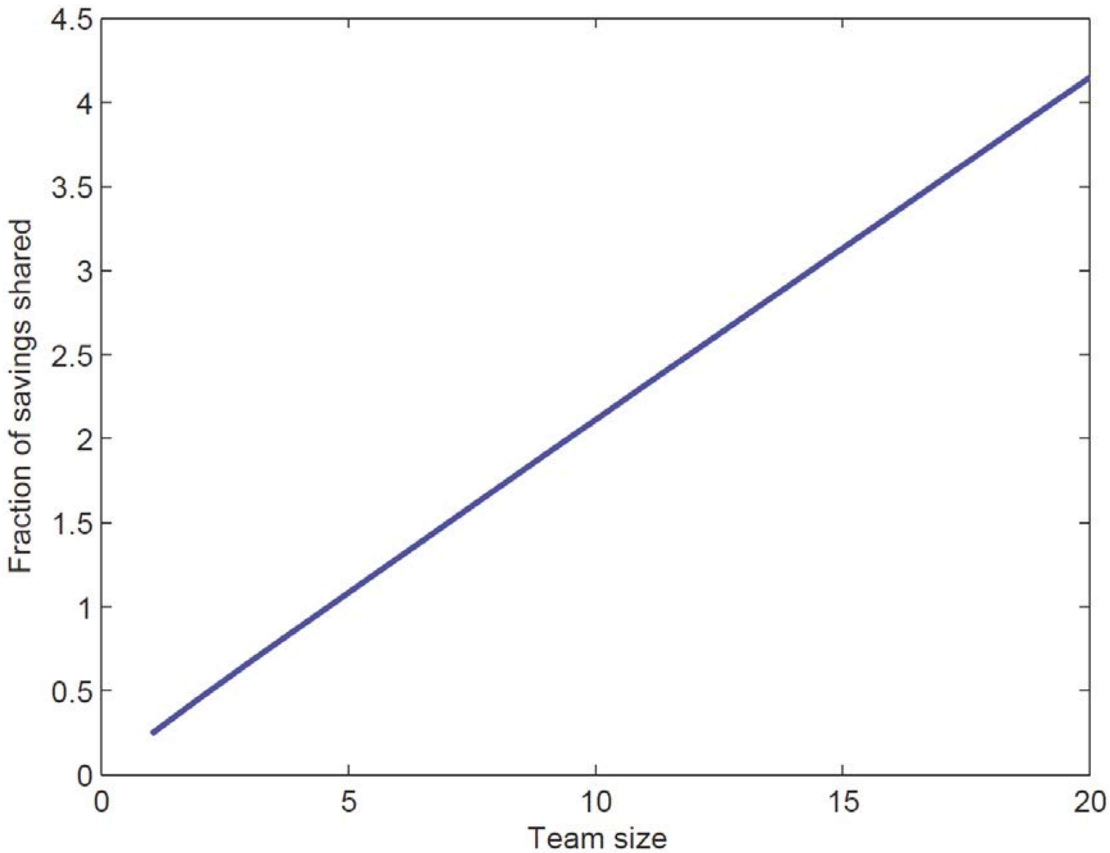
Tollen, Laura**.** 2008. "Physician Organization in Relation to Quality and Efficiency of Care: A Synthesis of Recent Literature," The Commonwealth Fund,

Table 1

Variable	Mean	Std. Dev.
Any Potential Gap in Care	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	

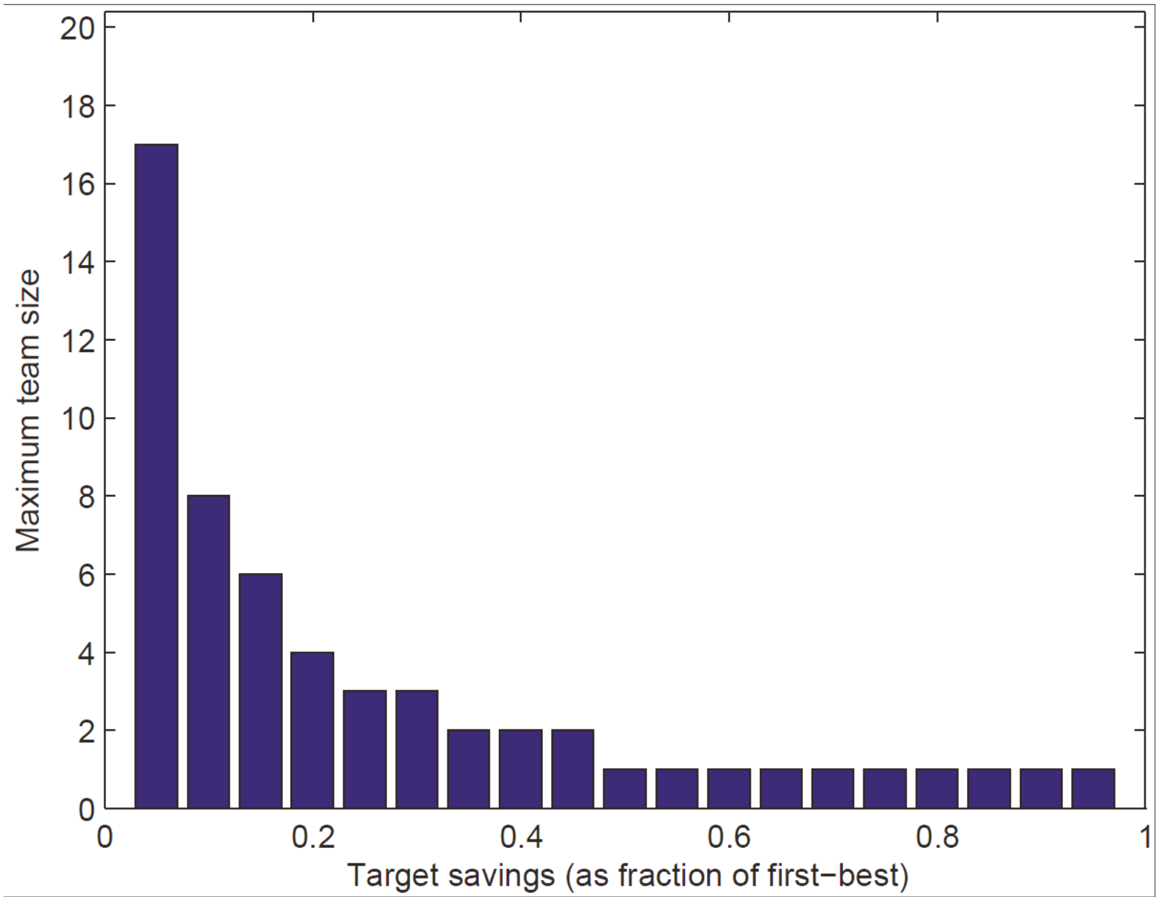
The patient sample contains commercial insurance members whose employers are commercially insured, and who have evidence of chronic illness. The provider sample consists of the primary care providers identified as the main providers for these patients on the basis of claims information

Figure 1: Optimal Sharing by Team Size



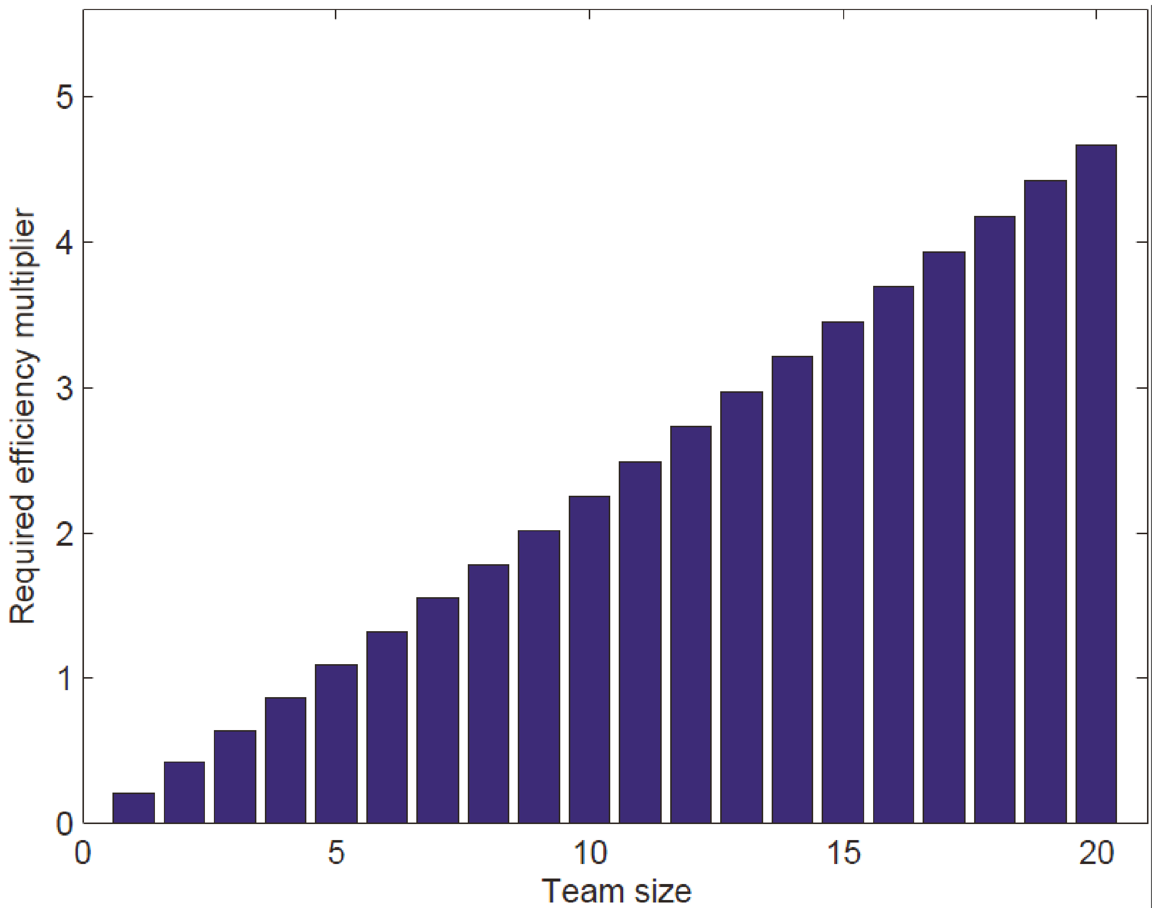
This Figure plots the cost sharing parameter, b, required to achieve a target level of cost savings by group size, N. The target level of savings are 20% of the savings achievable under first-best incentives. The calibration takes as given the number of doctors in the group, *N*, the size of their panel 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 2: Maximum Self-funding Team Size by Target Cost Savings



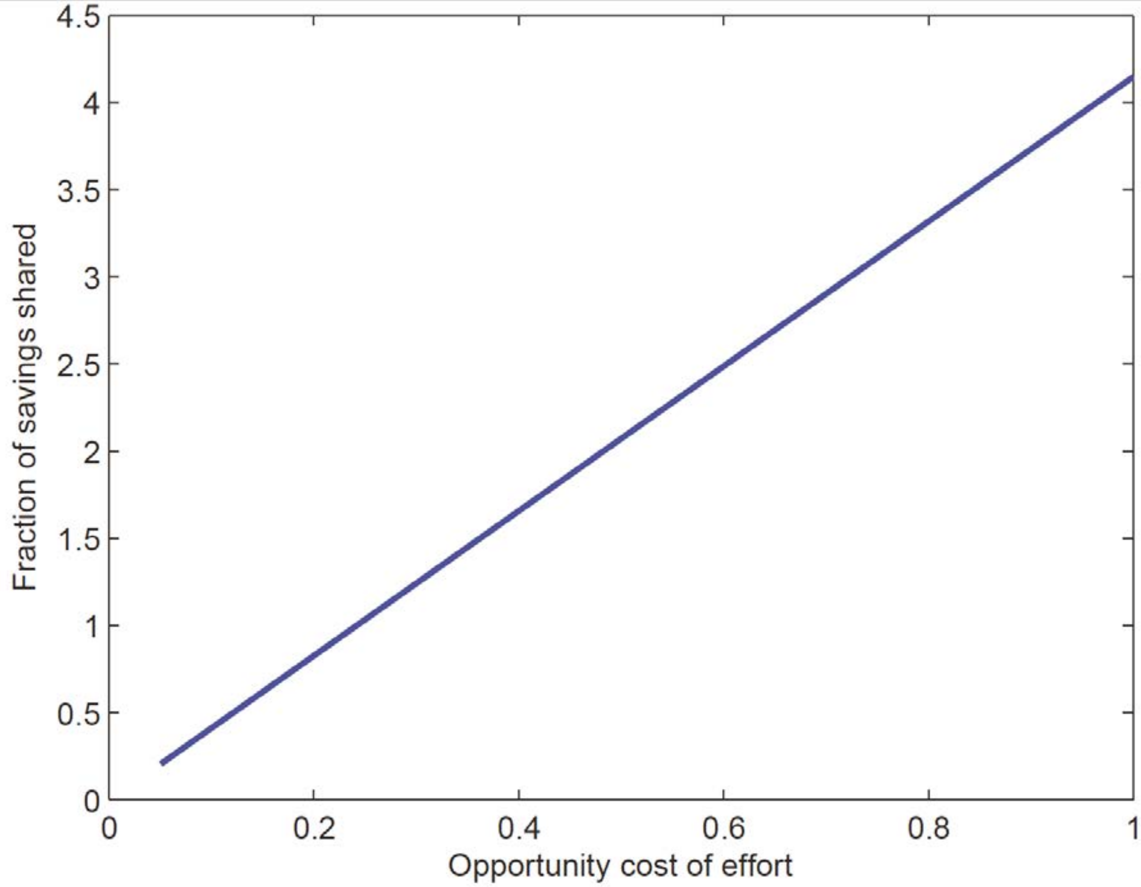
This Figure plots the maximum team size (on the vertical axis) that is consistent with self-financing incentives by target cost savings (the horizontal axis). The target level of cost savings are expressed as a percent of the savings achievable under first-best incentives. Savings are only paid out if 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.

Figure 3: Required Team Production Multiplier by Team Size



This Figure plots the efficiency "multiplier" that would have to be achieved in order for team incentives to be self-financing, as a function of the team size, N. The target level of cost savings is set at 20 percent of the first-best level of savings, and the quality target is set at zero (status quo). Positive savings are only paid out if the quality threshold is cleared, but negative savings (losses) are shared regardless. 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.

Figure 4: Optimal Sharing by Opportunity Cost of Effort



This Figure plots the sharing parameter, b, required to achieve a target level of cost savings for a team size of $N = 20$, by the opportunity cost of effort parameter, δ. 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.