Information and Quality When Motivation is Intrinsic: Evidence from Surgeon Report Cards[†]

By Jonathan T. Kolstad*

If profit maximization is the objective of a firm, new information about quality should affect firm behavior only through its effects on market demand. I consider an alternate model in which suppliers are motivated by a desire to perform well in addition to profit. The introduction of quality "report cards" for cardiac surgery in Pennsylvania provides an empirical setting to isolate the relative role of extrinsic and intrinsic incentives in determining surgeon response. Information on performance that was new to surgeons and unrelated to patient demand led to an intrinsic response four times larger than surgeon response to profit incentives. (JEL D83, I11, L15)

A defining feature of health care markets is imperfect information (Arrow 1963). Standard models show that in settings where consumers are poorly informed about product quality there are welfare losses due to the less-than-optimal supply of costly quality (Dranove and Satterthwaite 1992) or the absence of markets for products and services that consumers otherwise value (Akerlof 1970). Losses generally stem from profit maximizing suppliers who are able to profit from information to which they are privy, but cannot be obtained (or verified) by the demand side (e.g., Arrow 1963; Gaynor 2006; Salop and Stiglitz 1977). Consequently, policies to correct market failures due to asymmetric quality information focus on demand. In health care markets, where quality is widely believed to be suboptimal, this is the rationale for efforts to gather performance information, such as mortality, and offer it directly to consumers. The existing evidence on such quality reporting in health care, however,

generally finds improvements in measured quality, but little evidence for corresponding changes in consumer demand (Epstein 2006; Steinbrook 2006).

To better explain the observed behavior of surgeons facing public quality reporting and to explore the nature of incentives among firms and individuals receiving public ratings, I consider the role of information in determining market outcomes when suppliers also have non-pecuniary incentives. I use the term "intrinsic motivation" to refer to incentives unrelated to profit and model it as a function not only of quality itself but of the ability to observably perform well relative to a reference group. In this context, information enters profit motives and alters intrinsic incentives when collecting and disseminating information provides the individual with a better sense of his or her own quality compared to peers.

The empirical setting for this study is the introduction of quality "report cards" for surgeons performing Coronary Artery Bypass Graft (CABG) surgery in Pennsylvania. Utilizing a detailed panel of data on surgeons and patients, I explore the effects of quality report card release on subsequent surgeon performance. Exploiting the information contained in the report card's risk adjustment scheme, I model surgeons' prior and posterior beliefs about market quality levels (both their own and that of their peers). I find an impact of this information on changes in surgeon quality but not on demand—evidence for the presence of non-pecuniary incentives resulting from quality reporting.

To incorporate profit incentives, I estimate a structural model of consumer demand for surgeons. Consumer utility is modeled as a function of the detailed set of individual patient and surgeon observables. In addition, I account for unobserved (to the econometrician) influences on choice, including the role of physician agency, using a random coefficients demand model (Berry, Levinsohn, and Pakes 1995; Train 2003). Simulations, relying on the estimated demand parameters, produce a measure of the additional market rewards for quality due to reporting. Variation in the ex ante distribution of patient demand for quality and the competitive structure of the markets leads to large differences in extrinsic incentives between surgeons. Individuals facing stronger profit incentives following the release of quality report cards show greater improvements in performance. This effect, however, is relatively small. Extrinsic incentives led to an additional 3 percent decline in the statewide risk adjusted mortality rate (RAMR) between the pre- and post-report card periods. Incorporating estimates of the intrinsic response to information predicts changes in surgeon quality that accords well with the observed response to reporting. The intrinsic response to quality reporting is about four times as large as the response due to profit incentives induced by changes in demand for quality. Quality improvements cannot be explained by changes in underlying patient severity, either based on observables used in the risk adjustment scheme or "unused" observables collected, but not included in risk adjustment. There is evidence for selection efforts on "unused" observables. However, the impact of these severity measures on outcomes are far smaller than those included in the risk adjustment. Further sensitivity analysis finds that the observed intrinsic response cannot be explained by dynamic pecuniary incentives—either in volume or mix of insurance payers—or inter-temporal technological change.

This paper contributes to the debate in economics on the merit of reporting schemes. Quality reporting and disclosure policies have been implemented across a

variety of industries to address a multitude of market failures (Chatterji and Toffel 2007). Reviewing the evidence, Fung, Graham, and Weil (2007) find the welfare impact of information-based policy interventions varies dramatically depending on the market and regulatory environment. Their assessment, however, appraises disclosure through the lens of the standard profit-maximizing model. The same approach also underlies much of the evaluation of reporting for CABG surgery. Despite finding that additional quality improvement followed the introduction of reporting programs (Epstein 2006; Ghali et al. 1997; Peterson et al. 1998; Hannan et al. 2003), the debate on and analysis of reporting efforts has focused on consumers' ability to interpret and respond to the information supplied. If quality report cards deliver information on and to suppliers who care about performance intrinsically, then their impact (both positive and normative) is not solely mediated through changes in demand.

Decisions about the value and type of quality information that should be measured and publicly reported depend critically on the model of supplier behavior. If suppliers operate under a standard profit model, demand side incentives can produce quality improvements. On the other hand, to the extent that information about peers alters surgeons' intrinsic incentives, public release is of less relevance. In fact, contrary to current efforts to simplify provider report cards, it may be preferable to deliver data with more clinical detail.

This paper also contributes to a broad literature in economics on information and incentives. Work in behavioral game theory and experimental economics has demonstrated a potential role for reference-based utility in individual behavior and incentives (Fehr and Schmidt 2006; Heffetz and Frank 2011). To date, however, relatively few empirical studies have documented such incentives in practice (Bandiera, Barankay, and Rasul 2005; Sauermann and Cohen 2008). This is due in part to the difficulty of empirically identifying changes in information in a market. The CABG setting allows me to overcome this problem by observing an exogenous and measurable change—the release of quality report cards.

The paper proceeds as follows. Section I develops a model of quality choice with intrinsic motivation. Section II discusses the data and setting. Section III presents the econometric specifications and results. Section IV provides discussion and Section V concludes.

I. A Model of Surgeon Objectives with Uncertainty

I begin by considering a model of the equilibrium quality choices of surgeons who gain utility from income (profit) and from performing well. Because quality is also valued independently of earnings, surgeons are willing to forgo some profit to enhance quality. This willingness is a function of the ability to observe performance—determined by the information structure of the market. The intuition of the model is that a surgeon with little information on his own performance, and that of his peers, is unable to accurately observe both static levels of quality and improvements. In this way, increased uncertainty dilutes the intrinsic incentive for quality improvement because surgeons do not see the result of effort or may believe, in the absence of information, that they are performing as well as they should be. This information also impacts surgeon incentives through the standard channel—quality

information informs consumers and, subsequently, determines quality elasticity of demand and the profitability of quality improvement.

Consider the quality choice of a surgeon in a monopolistically competitive market with regulated prices above the marginal cost of production. Prices are fixed at a regulated level, p_{reg}, and surgeons maximize utility by selecting a quality level subject to a convex production technology, $c(\theta_i, q_i)$. To incorporate preferences that include both profit and intrinsic incentives, I express supplier utility as

$$(1) \qquad U_i = \Pi_i(\theta_i, \theta_{-i}, \Omega) + \Gamma_i(\theta_i, \theta_j, \Omega).$$

Allowing profit and intrinsic utility from quality to enter as additively separable terms can incorporate a range of intrinsic preferences and is a common feature of models of physician behavior (Fehr and Schmidt 2006; Harsanyi 1955; McGuire 2000; Segal and Sobel 2007). The term $\Gamma_i(\theta_i, \theta_j, \Omega)$ captures individual i's intrinsic utility from quality relative to the reference group $j \in J$.[1] Firm demand, $q_i(\theta_i, \theta_{-i}, \Omega)$, is determined by the quality of surgeon i as well as the quality choices of all competing surgeons. The reference group J is not necessarily the same as the set of all competing surgeons, indexed by $-i$. Information in the market is captured by the variable Ω, the effect of which I return to below. Surgeon i solves the following problem:

$$(2) \qquad \max_{\theta_i} U_i = \Pi_i(\theta_i, \theta_{-i}, \Omega) + \Gamma_i(\theta_i, \theta_j, \Omega)$$

$$= q_i(\theta_i, \theta_{-i} | \Omega) p_{reg} - c(\theta_i, q_i) + \Gamma_i(\theta_i, \theta_j | \Omega).$$

The argument that maximizes (2) (optimum quality) is reached when

$$(3) \qquad \left(\partial q_i(\theta_i, \theta_{-i}^* | \Omega) / \partial \theta_i \right) p_{reg} + \partial \Gamma_i(\theta_i, \theta_j^* | \Omega) / \partial \theta_i = \partial c(\theta_i, \theta_j^*) / \partial \theta_i,$$

where all partial derivatives are taken with respect to own quality, taking the best response of other surgeons as given. The optimum is simply the point at which the sum of the marginal revenue (determined by the price and demand elasticity of quality) and marginal intrinsic utility is equal to the marginal cost of quality.[2] Implicitly differentiating (3), optimal quality increases in demand for quality (determined jointly by consumers' willingness-to-pay for quality and their ability to observe it) and intrinsic utility from quality improvement. Quality declines with the marginal cost of quality. In the standard model (where a surgeon cares only about pecuniary rewards) $\partial \Gamma_i(\theta_i, \theta_j^* | \Omega) / \partial \theta_i = 0$ and the equilibrium condition is reduced to setting

---

[1] Intrinsic motivation has a strict definition in the psychology literature: utility from the activity must be derived from a stimulus within the individual (Sauermann and Cohen 2008 provide a useful synopsis). The model here captures intrinsic incentives under this definition as well as in a more general sense (typically used by economists). It fits a strict interpretation of intrinsic utility if reference utility is derived internally but, due to uncertainty, is altered by the outside information. However, if surgeons care directly about a ranking that is provided by the report card this is (strictly speaking) an extrinsic motive because it is initiated by an outside stimulus. For simplicity, in this paper I refer to all non-pecuniary rewards as intrinsic though I acknowledge this may not adhere to convention in some fields.

[2] One could also move the marginal intrinsic incentive in equation (3) to the right hand side. In this case the intrinsic incentive enters as a "reduction" in marginal cost. This interpretation is developed by Gaynor (2006) to model supplier quality choice with not-for-profit incentives.

marginal cost equal to marginal revenue.[3] Equilibrium quality is an increasing function of the residual elasticity of demand for quality.

I also assume that a surgeon determines his or her own quality relative to a reference group $\tilde{\theta}_j \in (\underline{\theta}_j, \bar{\theta}_j)$. Intrinsic utility from quality is captured in the model as a function that maps the deviation between an individual's quality and their reference point to a change in utility: $\Gamma_i(\theta_i^* - \tilde{\theta}_j)$. The precise relationship depends on the reference group and the shape of the intrinsic loss function.

The information structure of the market captures the impact of quality reporting in the model. Information in period t is indexed by the set $\Omega^t \in \{\mu, \varepsilon\}$ containing two elements. The first term, μ, captures information on the relative location of a surgeon in the distribution. The second term, ε, measures the "quality" of information or the precision of a surgeon's beliefs about the distribution of reference qualities.[4] From equation (1) and the first-order condition in (3), it is clear that changes to the information structure can alter both components of surgeon utility. First, improved information allows consumers to more easily observe the quality of their full choice set of surgeons. This change in demand alters the pecuniary returns to quality improvement. Second, a better signal provides surgeons with more precise information on the performance of the set J in the reference group. Improved knowledge about the reference group alters the shape of the intrinsic utility function because changes in performance are more easily measured and produce utility gains.

Figures 1 and 2 present a graphical example. In Figure 1, each surgeon has a convex marginal cost curve (labeled MC) and a marginal revenue curve that is increasing in quality (labeled MR). In this standard model a surgeon is solely extrinsically motivated. He selects the point at which the marginal cost of quality improvement equals the marginal pecuniary benefit (MR). Introducing information on performance alters quality by changing the slope of the marginal revenue from improving quality. This can be seen in the rotation from MR0 to MR1. Quality reporting leads to improvement in performance from θ^* to θ^{**} by altering demand for quality and, thus, the pecuniary reward.

Figure 2 introduces a mixed surgeon utility function. Panel A contains two possible intrinsic utility functions. Two surgeons (labeled 1 and 2) have the same concave intrinsic utility function, but compare themselves to the lowest and highest surgeon in their reference set respectively.[5] The deviation between profit and utility maximizing quality, visible in panel B of Figure 2, is determined by the shape of the MC, MR, and MB curves. For any monotonically increasing intrinsic utility function the marginal benefit curve (MB) is higher than MR. Graphically, this is captured in the

---

[3] Because prices are set by a regulator (Medicare), demand is equal to marginal revenue provided quality elasticity of demand is not a function of patient cost (i.e., raising quality does not lead to differential increases in demand from the most severe patients). I assume payments are sufficient to make marginal patients profitable. In cardiac surgery this condition is likely to hold. Huckman (2006) finds that cardiac surgical Diagnosis Related Groups (DRGs) are profitable on average and at the margin. Chernew, Gowrisankaran, and Scanlon (2001) also find evidence that reimbursement for cardiac surgery is greater than cost (though the degree of profitability varies by payer).

[4] In a general model, surgeons are Bayesian learners and new information induces a posterior distribution of reference quality. Updating alters both mean beliefs about relative quality and the precision of the posterior.

[5] Note that the curve for surgeon 2 is convex as presented due to the reference point. Utility, however, is still concave in that returns are diminishing for changes further from the reference quality, the "best" surgeon.
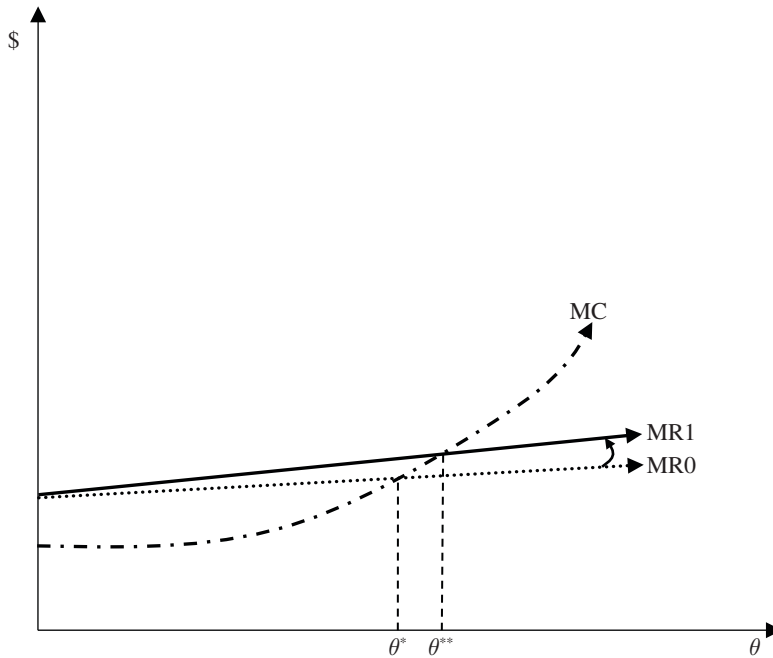
FIGURE 1. PROFIT MAXIMIZING QUALITY CHOICE

increase in equilibrium quality from θ^{**} to θ' or θ'', depending on surgeon's intrinsic utility function and reference point.[6]

Prior to quality reporting, without information on peers, any change in quality is indistinguishable from noise. In panel A of Figure 2, this is the flat intrinsic utility function $\Gamma(\theta_i - \underline{\theta} | \Omega^0)$. Without information, improving quality does not increase utility because it cannot be observed. Utility from quality need not, however, be set at zero if surgeons gain some level of static intrinsic utility—the "warm glow" from being a cardiac surgeon.

After report cards are released, new information is provided with the signal $\Omega^1 \in \{\mu'', \varepsilon''\}$. Information that alters a surgeon's perceived relative quality changes his utility either positively or negatively (e.g., they learn they are better or worse than expected). Marginal incentives are also altered by the quality of the signal and the ensuing shape of the utility function. In this case, because there was no information prior to quality reporting, the prior slope of intrinsic utility is zero so any signal that provides new information will unambiguously increase the slope of the intrinsic utility curve resulting in increased intrinsic incentives for all physicians.[7] This

---

[6] In Figure 2, equilibrium quality is relatively high even for a profit maximizing supplier so for surgeon 1 (where the reference point is defined by the lowest quality in the group) intrinsic incentives increase quality to θ', a small effect. On the other hand, for surgeon 2, who aspires to be "the best," intrinsic incentives are relatively strong, leading to an equilibrium quality choice of θ''.

[7] This also underscores the fact that for any intrinsic utility function unrelated to information, quality reporting should not alter non-pecuniary incentives for quality. Even if surgeons are purely altruistic and perfect agents for patients, quality reporting would not alter quality levels unless information enters utility (the argument parallels perfect agency in Ellis and McGuire 1986). For example, not-for-profit providers are assumed to more accurately reflect social preferences in their strategies but, in general, models have not considered the knowledge they have about quality and how this impacts their incentives (Arrow 1963; Newhouse 1970; Sloan 2000).

Panel A

$

$\Gamma_1(\theta_i - \underline{\theta}|\Omega^1)$

$\Gamma(\theta_i - \underline{\theta}|\Omega^0)$

$\Gamma_2(\theta_i - \bar{\theta}|\Omega^1)$

$\underline{\theta}_j$

$\bar{\theta}_j$

θ

Panel B

$

$MB1 = MR1 + \Gamma_1'$

MC

B

$MB1 = MR1 + \Gamma_2'$

MR1

A

C

MR0

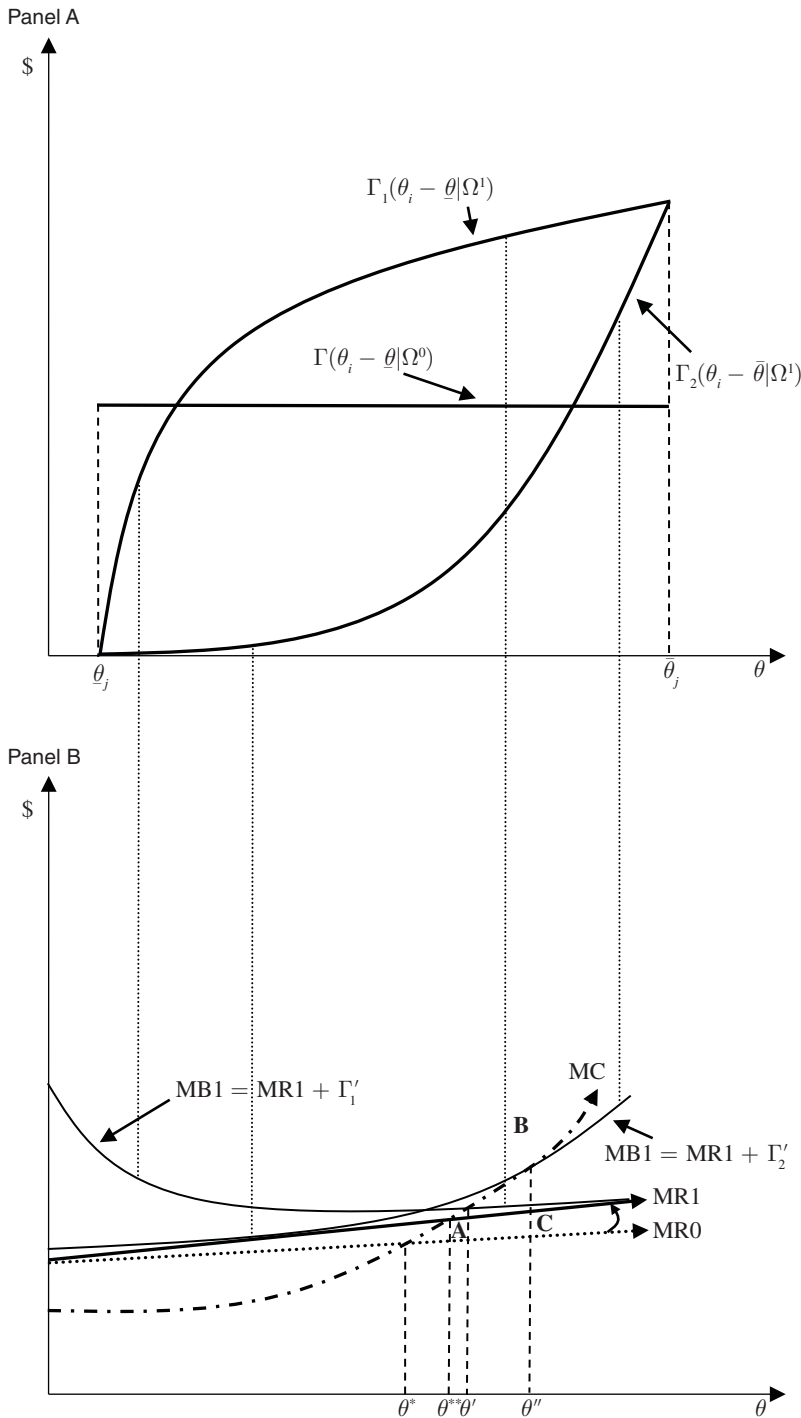θ^* $\theta^{**}\theta'$ θ''

θ

FIGURE 2. INTRINSIC UTILITY AND QUALITY CHOICE

need not be true, however. If surgeons have some information on the distribution of reference quality prior to formal reporting then the slope of the intrinsic utility function will not initially be zero and some surgeons can receive new information

that diminishes or leaves incentives unchanged. Rather than make explicit assumptions about priors that would allow unambiguous predictions, I note the potential for quality reporting to either increase or decrease intrinsic incentives and allow for both effects empirically.[8]

A measure of extrinsic and intrinsic utility can also be computed in this framework. In Figure 2, the area below the marginal cost curve between A and B and above the marginal revenue curve between A and C measures the additional cost in excess of revenue a surgeon is willing to expend in order to improve performance solely to gain intrinsic utility.[9] That is, for every quality investment beyond θ^{**} a surgeon loses money at the margin. Willingness to undertake such investments captures a non-pecuniary incentive in the utility function related to costly quality.

The primary goal of the remainder of the paper is to measure empirically the relative contribution of each of these incentive components in determining surgeon response to quality reporting in Pennsylvania. Viewed in Figure 2, this effort reduces to decomposing the observed quality improvement from θ^* to θ'' into the share due to the move from MR0 to MR1, and the share due to changes in intrinsic incentives from $\Gamma(\theta_i - \bar{\theta} \,|\, \Omega^0)$ to $\Gamma(\theta_i - \bar{\theta} \,|\, \Omega^1)$. Decomposing quality change in this manner is testing the predictions of the model that suppliers imperfectly observe quality, that they care about this quality independently of pecuniary rewards, and that they gain new information from the release of quality report cards.

II. Background and Setting

A. *Quality Reporting in Health Care*

Quality reporting programs have been implemented in many forms across a variety of markets for health insurance and for providers (see Kolstad and Chernew 2009 for a review of the evidence to date). As of 2006, 47 states had some form of quality reporting system in place for health care providers (37 are mandatory and 10 are voluntary) (Steinbrook 2006). The most studied within the provider context have been the CABG report card programs in New York and Pennsylvania.[10] Reporting of surgeon and hospitals' risk-adjusted mortality rate (RAMR) for CABG began in 1989 in New York State. Pennsylvania's experiment followed shortly thereafter and was led by the Pennsylvania Health Care Cost Containment Council (PHC4), a public/private partnership. They began collecting discharge data on outcomes and patient comorbidities in 1990. The first widely available report card was released in May of 1998 and included data from 1994–1995.[11]

---

[8] To make more explicit predictions in this setting would require me to define a set of beliefs for a given surgeon as well as the form of the signal being provided by the report card. The assumptions required would be very strong and could basically determine any set of outcomes based on those assumptions. Miller (2006) demonstrates the challenge posed by distributional assumptions on the form of competition in modeling provider response to quality reporting, though without incorporating intrinsic incentives.

[9] For any cost function and demand curve the dollar value of intrinsic utility is $\int_{\theta^{**}}^{\theta''} MC(\theta) \, d\theta - \int_{\theta^{**}}^{\theta''} MR(\theta) \, d\theta$.

[10] Similar report card programs for cardiac surgery are now in use in many states including California, Massachusetts, Florida, and New Jersey as well as at the country level in the United Kingdom (Steinbrook 2006).

[11] Reports based on 1990–1993 data were constructed and released between 1992 and 1995. However, these reports are no longer available and discussions with experts suggest that these data and the reports were not widely distributed. Furthermore, the risk-adjustment measures differed from later reports.

Some of the earliest evidence on these policies comes from surveys of market participants and case studies. The former suggest that quality reporting did not significantly alter consumer choice (Schneider and Epstein 1998; Hannan et al. 1994).[12] Dziuban et al. (1994) present a case study of the response of a large community teaching hospital to New York State's release of quality report cards. This description is particularly relevant as that experience maps very closely both to the theory developed in this study and to the empirical implementation that relies on risk adjustment as a measure of new information provided to surgeons. The hospital had comprehensive data capture and case review practices in place prior to quality reporting in order to improve performance. Despite this, new information was provided from the report card due to learning that their patient population was less severe than expected and, thus, their actual quality was low, despite their observed mortality rate. This new information led to a detailed analysis and practice changes that resulted in subsequent quality improvements (reductions in RAMR). Studies that rely on observed consumer behavior find more evidence for an effect of quality reporting on aggregate market share, though the effects remain relatively small (Cutler, Huckman, and Landrum 2004; Dranove and Sfekas 2008; Mukamel and Mushlin 1998).

Another line of research on supplier response to quality reporting has focused on selection against sicker patients. In their survey, Schneider and Epstein (1998) found that 63 percent of surgeons report reduced willingness to operate on severe patients and 59 percent of cardiologists report having difficulty finding a surgeon for their more severe patients. Dranove et al. (2003) compare outcomes for cardiac patients in the Medicare population in New York and Pennsylvania with those in locations without report cards. They find patients had better matches with providers—a gain from the release of information—as well as selection by surgeons against sicker patients, higher resources use, and worse outcomes.[13] Fong (2008) considers similar selection behavior in a theoretical setting. While this paper does not bear on the existence of such effects, her findings underscore the important role supplier objectives should play in determining optimal policy.

Survey evidence and, to a slightly lesser extent, revealed preference suggest little demand response. On the other hand, a review of the medical literature finds, "…there is evidence that the public disclosure of death rates associated with surgery in New York and other states has contributed to reductions in operative mortality…" (Steinbrook 2006, p. 1848). Applying the mixed incentive model in this paper can explain these findings.

B. *CABG Surgery*

CABG surgery is one of a range of possible treatments for coronary artery disease, a condition in which a patient's blood flow to the heart is compromised by narrowing of the coronary arteries. The severity and symptoms of the disease vary with the degree of obstruction. Cardiac catheterization, a process that allows a cardiologist to image the blockage(s), is used to assess the extent of the disease and determine

---

[12] The introduction of quality reporting was highly contentious in New York and Pennsylvania (Bumiller 1995). To the extent that surgeons or cardiologists had an opinion on the topic, there is a concern that their answers reflected their preference for or against quality reporting.

[13] The aggregate welfare effect of report cards during their study period was negative.

the appropriate treatment regime. Patients can be managed medically using drugs (beta-blockers, aspirin, ACE inhibitors, etc.) or surgically with either percutaneous transluminal coronary angioplasty (PTCA) or CABG. If a surgical intervention is decided upon, the patient must then choose between angioplasty and CABG and select a cardiac surgeon. All of these choices depend not only on patient characteristics but also on the incentives facing their agent in the choice, the cardiologist (e.g., Afendulis and Kessler 2007).

Cardiac bypass surgery is the most invasive treatment for cardiovascular disease. After opening the chest wall, the surgeon creates a bypass around the blocked coronary artery using either internal mammary arteries or arteries from the leg. The production function is complex and determined not only by the attending surgeon, but also by a team of physicians and support staff (Edmondson, Bohmer, and Pisano 2001). The physicians required for a CABG procedure include a cardiac anesthesiologist and the operating surgeon. In addition, a perfusionist, nursing, and other support staff play an important role in the procedure itself and the follow up care as the patient recovers. A widely documented effect in this market is the presence of a volume-outcome relationship. This is generally attributed to learning-by-doing, though the endogeneity of volume raises the alternate mechanism of selective referral (Arrow 1963; Gowrisankaran, Ho, and Town 2006; Gaynor 2006; Ramanarayanan 2008).

The model developed thus far has focused on the surgeon as the unit of analysis, not the hospital. This reflects the fact that, with a few specific exceptions, surgeons are not employed by the hospitals at which they perform surgery. Rather hospitals provide a space for surgeons to perform CABG as well as the necessary support staff and resources. The surgeons themselves act as freelance workers and gain profits directly from providing additional services (Cutler, Huckman, and Kolstad 2010). Underscoring the freelance nature of the relationship, many surgeons have privileges to perform CABG surgery across different hospitals at the same time (Huckman and Pisano 2006). The hospital and surgeon are compensated separately with a fee for the different services provided by each. The fee-for-service (FFS) compensation for CABG makes it one of the most profitable types of care for Medicare enrollees and most types of private insurance (Huckman 2006; Chernew, Scanlon, and Hayward 1998). Because reimbursement is above cost—both on average and at the margin—and payments are made directly to surgeons for each surgery, the quantity of procedures provided is a direct measure of profit.

C. *Data*

Data were obtained from the Pennsylvania Health Care Cost Containment Council (PHC4) and contain observations for 89,406 CABG surgeries performed in the state of Pennsylvania in 1994–1995, 2000, and 2002–2003 (PHC4, 1998, 2002, 2004, 2005). Each observation includes information on the surgeon performing the surgery, the hospital at which the surgery was performed, patient demographics, a set of patient comorbidities, the patient's home zip code, data on the payer type, and a set of outcome variables.[14] The outcome of interest in this paper is inpatient mortality.

---

[14] Patient characteristics include age, indicators for cardiogenic shock, concurrent angioplasty, complicated hypertension, dialysis, female gender, heart failure, and prior CABG or valve surgery.

In addition, I merge data on surgeon tenure in the Pennsylvania market.[15] This is intended to capture the life-cycle nature of returns to quality as well as costs that are associated with age of the surgeon. To account for the fact that academic surgeons may differ in their incentives, I also gather data on the number of publications for each surgeon by 1995 and the number of citations to those publications.

I compute a measure of risk adjusted mortality to capture surgeon performance using the standard approach used by PHC4. Each observation includes a dummy variable equal to one if a patient died in the hospital during or immediately following surgery. The log probability of death is computed as follows:

$$(4) \qquad \ln\left(\frac{\Pr(MORT_{i,s,h} = 1 \,|\, x_i)}{1 - \Pr(MORT_{i,s,h} = 1 \,|\, x_i)}\right) = \beta_0 + \beta_1 \cdot X_i + \varepsilon_{i,s,h},$$

where i indexes patient, s surgeon, and h hospital. *MORT* is the indicator variable that equals one if the patient died in the hospital. This model is estimated for each report card period (1994–1995, 2000, 2002, and 2003).[16] The fitted values are obtained for each patient to form a predicted probability of death—the Expected Mortality Rate (EMR). For each surgeon, I then compute a measure of risk adjusted performance (RAMR):

$$(5) \qquad RAMR_{s,h} = \left(\frac{OMR_{s,h}}{EMR_{s,h}}\right) OMR_{PA},$$

where the risk adjusted, expected, and observed mortality rates for each surgeon s or hospital h are *RAMR*, *EMR*, and *OMR* respectively. Risk adjustment is accomplished by dividing the actual number of fatalities by the expected number of deaths conditioning on the actual patients selecting surgeon s or hospital h. This ratio is then normalized by multiplying it by the statewide average mortality rate.

Table 1 contains summary statistics for key variables in the data. As angioplasty gains market share, substituting for CABG, the statewide volume declines (this also occurred in neighboring states, see Cutler, Huckman, and Kolstad 2010). The performance improvement over time is also apparent in the reductions in the mean *RAMR*. Between 1994 and 2003 the mean surgeon *RAMR* dropped 42 percent from a rate of 3.42 percent to 2 percent. Figure 3 plots quarterly mean *RAMR* and unadjusted mortality highlighting the drop following the release of quality report cards in 1998. Comparing performance in 1994–1995 to the period between 2000 and the second quarter of 2002 (quarters 24 to 33 in the graph) shows a decline in quarterly average *RAMR* and *OMR*, though these results are noisy from quarter to quarter.

Table 2 presents a matrix of transition probabilities between quintile measures of surgeon quality (5 indexes the highest quality and 1 the lowest in each period) in 1995 and 2000. The evidence suggests that it is feasible for surgeons to improve performance regardless of baseline quality. Of the worst performing surgeons in

---

[15] This is computed based on the date the surgeon was first licensed in the state. These data are gathered by matching surgeon names with the state database of license information available at http://www.dos.state.pa.us/bpoa/cwp/view.asp?a=1104&q=432785.

[16] The coefficient estimates and marginal effects for this regression for the 1994–1995 and year 2000 data are presented in Appendix A1.

TABLE 1—DESCRIPTIVE STATISTICS BY YEAR

Year	Observations	Surgeons	Hospitals	Mean RAMR[†]	Mean OMR[†]
1994–1995	18,351	201	43	3.42	3.23
2000	19,594	182	55	2.38	2.20
2002	15,999	187	62	2.02	1.82
2003	15,157	183	63	2.00	1.85

Note: [†] Surgeon weighted average.

TABLE 2—TRANSITION MATRIX FOR RAMR QUINTILE BETWEEN 1995 AND 2000

	2000 RAMR Quintile					
1995 RAMR Quintile	1	2	3	4	5	Exit
1	0.08	0.11	0.03	0.03	0.13	0.63
2	0.08	0.24	0.26	0.16	0.08	0.18
3	0.16	0.13	0.11	0.29	0.16	0.16
4	0.16	0.11	0.13	0.08	0.18	0.34
5	0.19	0.05	0.08	0.05	0.16	0.46
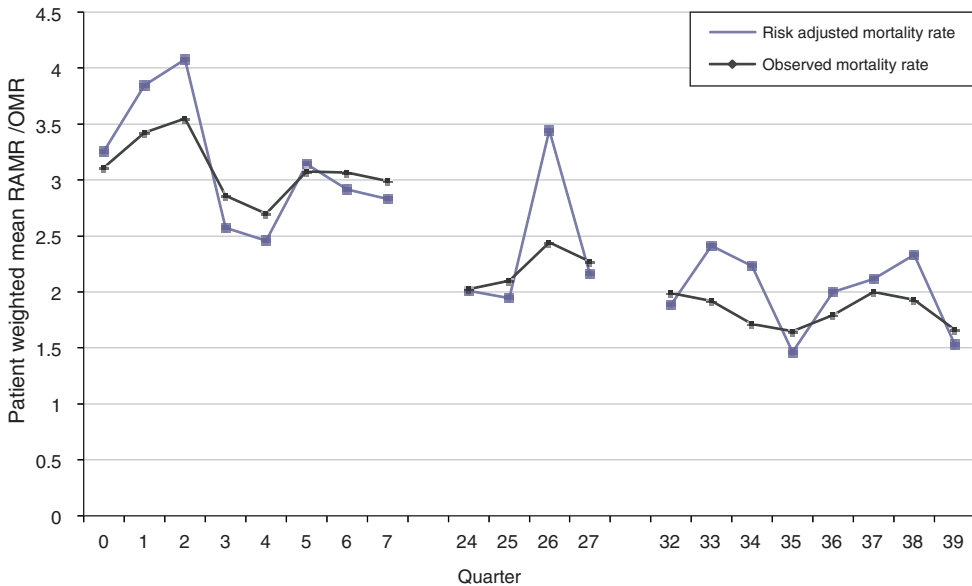
FIGURE 3. MEAN QUARTERLY PERFORMANCE

1995, 13 percent were in the top 20 percent in 2000, while 8 percent remained in the lowest performing group. Surgeons in the third quintile of performance in 1995 were equally likely to be in the highest quality as the lowest quality quintile by 2000 (16 percent in both cases). Transition probabilities are computed using the full sample of surgeons. Looking at the right most column of Table 2 it is also clear that exit is substantial, particularly in the highest and lowest quality (quintiles 5 and 1 respectively). For this reason subsequent analyses are limited to surgeons observed in both the pre- and post-reporting period.

III. Econometric Models and Results

A. *Identifying Intrinsic Incentives Using Information*

To measure the effect of new information on surgeon choices due to intrinsic incentives, I construct an estimate for the *new* information provided by quality reporting that was available only to surgeons, but not observed by consumers. The empirical question is whether surgeons who receive new information that is unrelated to demand differ in their subsequent quality improvement.

To construct the information measures I assume that, in the absence of data, a surgeon forms beliefs about performance by observing successes and failures—his or her own inpatient mortality rate. Surgeons do not, however, know with great certainty whether a patient is likely to have died given their underlying severity and the latest techniques and technologies.[17] The introduction of risk adjustment provides this information by giving surgeons the "true" difficulty of their cases—the expected outcome had the average surgeon in the state handled the case. The degree to which this confirms or differs from surgeons' priors is a measure of the amount of new information contained in the report for each surgeon.

I approximate this empirically as a nonlinear function of the difference between a surgeon's mean pre-report card *RAMR* and *OMR*: $f(RAMR_{pre} - OMR_{pre})$. A surgeon with a larger difference between his *RAMR* and his observed mortality, regardless of the level of each, is provided more new information by the introduction of quality reporting. Figure 4 presents a histogram of the frequency of this measure. Information contained in the risk-adjustment appears to be roughly normally distributed around zero, with a substantial share of surgeons who have a *RAMR* that differs from their *OMR*.

Viewed through the lens of the theory in Section I, this measure of information corresponds to a specific case in which surgeons care intrinsically about performing as well as possible given the average level of skill and technology in the market. More generally, this measure captures the magnitude of new information that might induce effort due to a comparison with many possible reference groups (e.g., very high quality surgeons for those learning they are doing very well and just getting to be average for those learning they are not doing as well). The intuition of the model holds in these cases, though the empirical expression allows the response to vary for good and bad news of different magnitudes without explicitly testing the precise reference group. I return to this issue with some sensitivity analysis in Section IVD which allows local peers to differentially impact effort.

I begin the analysis by plotting the difference between surgeon *RAMR* and *OMR* for 1994–1995 against both the percentage change in surgeon volume from 1995 to 2000 and the percentage change in *RAMR* over the same period. These results are presented in Figure 5. I fit the data using a Kernel-weighted local polynomial smoother (Fan and Gijbels 1996) to estimate a nonparametric surgeon response to new information.

---

[17] It is plausible physicians may have an indication of a given patient's severity. However, the evidence on physician difficulty in assessing probabilities, the dynamics of new technologies and techniques, and physician reliance on rules of thumb in treatment choices suggests that the objective measure of risk adjustment would provide new information relative to their existing assessments of patient severity (e.g., Frank and Zeckhauser 2007). Dziuban et al. (1994) present a case study of precisely this mechanism in response to New York State's quality report cards.
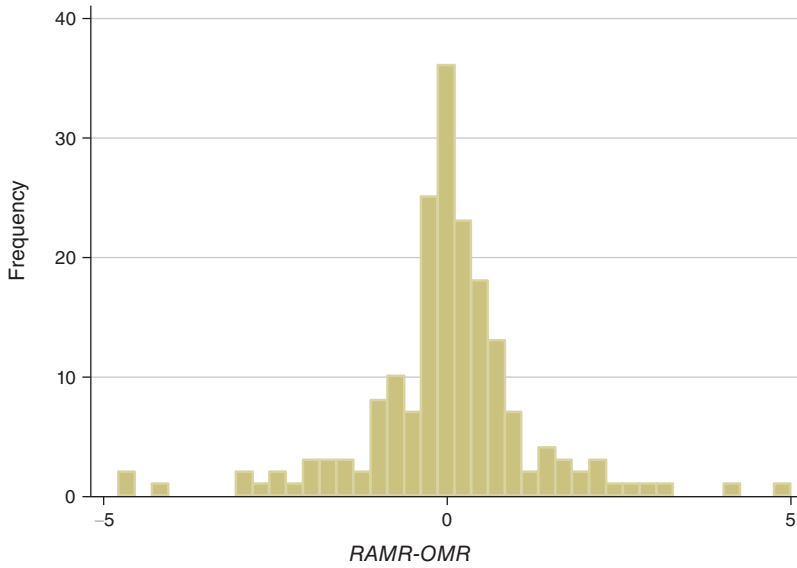
FIGURE 4. FREQUENCY OF NEW INFORMATION PROVIDED BY THE 1994–1995 REPORT CARD
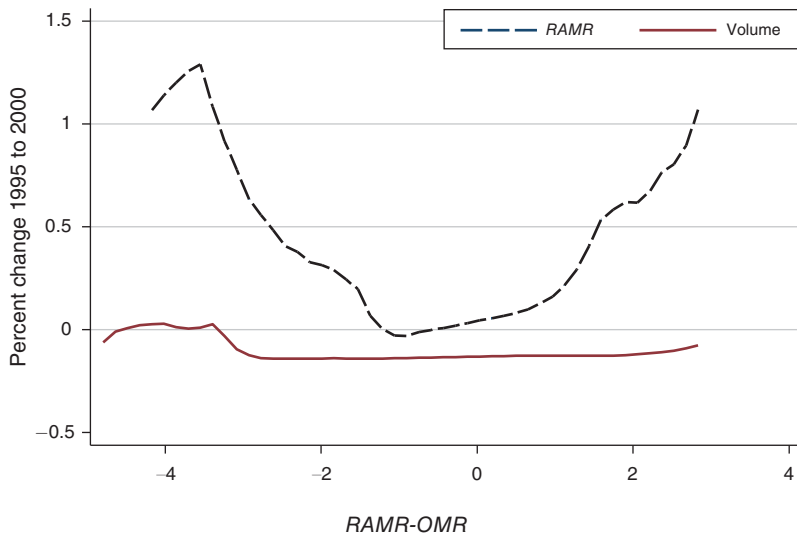
FIGURE 5. LOCAL POLYNOMIAL SMOOTHED ESTIMATES FOR THE RELATIONSHIP
BETWEEN NEW INFORMATION AND CHANGES IN VOLUME AND QUALITY

The lower curve plots the relationship of the difference between surgeon *RAMR* and *OMR* in 1994–1995, and the change in surgeon volume between 1995 and 2000. The curve is remarkably flat, suggesting that this measure of new information did not lead to large changes in demand.[18] The upper curve in Figure 5 plots the same information measure against surgeons' change in *RAMR* following reporting

---

[18] Formalizing this plot as a regression, I cannot reject the hypothesis that the five quartiles of the distribution of *RAMR-OMR* had the same change in volume between 1995 and 2000.

(computed by taking the difference relative to 1995 levels so positive values represent quality improvement). The curve is u-shaped centered near zero. These results suggest that *RAMR* remained roughly the same between 1995 and 2000 for surgeons receiving no new information. Moving away from zero in either direction, we see that increased information led to larger improvements in quality.

The effect increases monotonically in each direction over most information ranges. Enhanced incentives for surgeons whose performance was worse than expected (*RAMR* > *OMR*) fits easily into the model described in Section I. Performance improvement among surgeons learning they were at higher than expected quality is inconsistent with standard models of incentives, but accords with intrinsic incentives that are affected by quality information. If the shape of the intrinsic utility function is such that the posterior slope is greater than the prior, then improved information can enhance incentives, regardless of the sign of the change.[19] An example of this type of effect is a surgeon who desires to be the best. Learning that he is better than expected and closer to his objective increases the incentive to improve.

To implement this identification strategy in a testable model requires additional parametric assumptions. I saturate the model by separating the magnitude of new information into quintiles of the difference between *RAMR* and *OMR*, indexed by *n*. Surgeons in the bottom two quintiles (groups 1 and 2) receive information that they are worse than they thought they were (i.e., *RAMR* is higher than their *OMR*). Surgeons in the middle quintile (percentiles 40 to 59 of the distribution and the omitted category in estimation) receive no new information (*RAMR* roughly equal to *OMR*) and the surgeons in the top two quintiles (groups 4 and 5) receive new information that suggests they are better than expected. It is possible that surgeons whose *RAMR* is similar to their *OMR* also learn from the introduction of report cards. To the extent this is true, we expect some response from quintile three if they are intrinsically motivated. In this case the estimated effect of intrinsic incentives is a lower bound.[20] Using this measure of information, the primary estimating equation is as follows:

$$(6) \quad \Delta\theta_s = \alpha + \lambda\Delta\Pi_s + \xi_n \sum_{n=1}^{5} I_{n,s}(RAMR_{pre} - OMR_{pre}) + X_s + X_h + \varepsilon_{s,h},$$

where the dependent variable is the change in surgeon quality (*RAMR*) between the pre- and post-reporting periods. X_s is a set of surgeon level observables, X_h is a set of hospital level controls, and ε_s is an i.i.d., mean zero error term. If information alters quality due to intrinsic incentives, additional data from report cards should produce performance improvement. The hypothesis $H_0 : \xi_n = \xi_3$ is a test for intrinsic incentives associated with the information contained in each group *n*. With quality as the dependent variable, changes in profit incentives due to quality reporting also enter the model, captured by $\Delta\Pi_s$. Estimating (6) thus requires constructing a measure of

[19] While I do not formally rely on this as a test, responding to information that performance is better than expected by improving performance is inconsistent with the behavior predicted by a standard model of profit maximization.

[20] Another reasonable possibility is that higher performing surgeons have a better idea of their underlying patient severity than those at the bottom of the distribution. In this case it is also true that the estimated impact for groups 1 and 2 will be biased downwards. I thank an anonymous referee for this helpful extension.

the change in extrinsic incentives (the relative slope of MR in Figure 1) induced by quality reporting. This is the focus of the next section of the paper.

One important concern in estimating (6) is the potential for mean reversion that could bias the estimated effect of both intrinsic and extrinsic incentives on changes in surgeon quality. To account for this possibility, I include a surgeon's average *RAMR* in 1994–1995 in the vector X_s. This eliminates mean reversion in the estimated effect of information and controls for the mechanical relationship between a surgeon's baseline quality and the implied returns to quality improvement.[21] Despite these empirical advantages, I also present results without the pre-reporting *RAMR* because any correlation between *RAMR* and the intrinsic information measure will bias the ξ_n coefficients toward zero.

Finally, technological change over time could affect the model. I account for this in two ways. First, risk adjustment is computed for each period and, subsequently, incorporates changes in the average ability to treat a patient with a given comorbidity. That is, a surgeon's risk adjusted performance in period t is defined in terms of the expected mortality conditional on patient covariates given the current period technology. This controls for the majority of inter-temporal improvement in technology. A second concern is that technological change alters quality due to unobserved factors uncorrelated with the risk adjusters. I assume that this effect is equal across the panel of surgeons, conditional on surgeon and market observables. Under this assumption, the intercept α captures any additional unobserved technological change. Furthermore, specifications that include a surgeon's baseline *RAMR* allow the impact of technology to enter flexibly across the ex ante quality distribution. If, for example, a new technique or device were introduced that is more effective for lower performing surgeons (or they adopt this later than surgeons observed to be high quality) it is captured in the relationship between mean 1994–1995 *RAMR* and changes in quality.

B. *Surgeon Quality and Patient Demand*

In order to more precisely relate demand side factors to surgeon quality choices following reporting, I estimate a structural model of consumer demand. Relying on parameter estimates for patients' utility, I can simulate alternate information environments. I model patients' discrete selection of a surgeon allowing for factors that are unobserved (to the econometrician) but alter choice. These enter as random coefficients. (Berry, Levinsohn, and Pakes 1995; Nevo 2000; Train 2003). I present a brief discussion of the key variables and refer the reader to Kolstad (2012) for a detailed description of the structural model.

Each patient selects from the set of surgeons in his or her Hospital Referral Region (HRR). The utility for patient i from choosing a given surgeon s is a function of

---

[21] I also note that, even in the absence of 1994–1995 *RAMR* control, if I have accurate measures of quality, concerns about mean reversion will be minimized (i.e., the error term on the estimated quality in each period is small relative to the treatment effect). Because I have observations from a relatively long pre- and post-period (I observe an average of 240 surgeries per surgeon in the pre- and 160 in the post-reporting period), I expect the variance in estimated quality in each period to be small relative to the magnitude of the effect I am trying to identify (*RAMR* change).

travel cost, expected health improvement (capturing all components of quality), and an error term. Indirect utility to consumer i who selects surgeon s is

$$(7) \qquad u_{i,s,h} = g(X_i, \eta_i, Z_{s,h}, \theta_{s,h}; \rho) + \varepsilon_{i,s,h},$$

where X_i and η_i are vectors of observed and unobserved patient characteristics all of which lead to differences in taste. $Z_{s,h}$ is a K-dimensional vector of hospital and surgeon characteristics not directly related to expected health. $\theta_{s,h}$ is the expected quality (gains in health) for surgeon s at hospital h. Finally, $\varepsilon_{i,s,h}$ is an i.i.d. error term with a type-1 extreme value distribution and ρ is a vector of parameters. A patient selects surgeon s at hospital h if and only if $u_{i,s} > u_{i,-s} \forall -s \neq s$. Indirect utility in (7) is derived directly from a quasilinear utility function without wealth effects or prices. Prices do not enter choice because patients are generally well insured (55 percent by Medicare) so the out-of-pocket cost is unlikely to vary in any meaningful way between surgeons.[22]

Information (and quality reporting) enters the model by altering beliefs about expected health gains from choosing a given surgeon, s. Patients are assumed to develop beliefs based on all available information on surgeon performance (from both formal and informal information sources). One important feature of patient choice is the likely influence of an experienced agent (the patient's cardiologist). To account for the role of agency in determining demand, I include a measure of the patient-surgeon match assumed to be mediated through agency. This is captured by the deviation between a patient's severity and the lagged average severity seen by surgeon s in the prior quarter ($EMR_i - \overline{EMR}_{s,t-1}$ where EMR is computed using the fitted values from equation (4)). Incorporating this measure into the demand systems assumes that patients who have more comorbidities are sicker on average and would benefit more from being treated by a surgeon who has the training and background to treat relatively sick patients. I also assume that this match value is not known to a patient. While he may be aware of his own severity he would not know the appropriate surgeon given previous patient flows. This information is, however, known by a referring cardiologist. Conditioning demand on agent specific information allows patient flows to respond to reporting directly, taking potentially important agent preferences into account, and any agent (cardiologist) response to quality reporting that changes preferences over patient-surgeon matching on severity.

Incorporating these features into the choice model, the expected quality if patient i chooses surgeon s at hospital h is

$$(8) \qquad \theta_{i,s,h} = (X_i + \eta_i) RC_{s,h}.$$

RC is a vector of surgeon and surgeon-hospital characteristics observed by the patient. Included in RC is a continuous measure of surgeon performance—prior quarter $RAMR$—and dummies for discrete quality ratings included in the report card

(i.e., better, worse or as expected mortality based on patient severity). Substituting back into (7) the patient's utility function is

$$(9) \qquad u_{i,s,h} = X_i Z_{s,h} + (X_i + \eta_i) RC_{s,h} + \varepsilon_{i,s,h}.$$

Individual choice is thus a function of observed (to the econometrician) patient and surgeon attributes as well as unobserved factors that alter patient response to quality information.[23] The probability that patient i chooses surgeon s at hospital h is

$$(10) \qquad P_{i,s,h} = \int \frac{e^{X_i Z_{s,h} + (X_i + \eta_i) RC_{s,h}}}{\displaystyle\sum_{-s \in S} e^{X_i Z_{-s,h} + (X_i + \eta_i) RC_{-s,h}}} \, \phi(\eta_i \,|\, b, W) \, d\eta_i,$$

where the unobserved components of utility are distributed according to the distribution $\phi(\eta_i \,|\, b, W)$ that is known up to a mean and covariance, b and W, to be estimated. Using this expression, I fit the data using simulated maximum likelihood (Train 2003). Estimates for the demand system are computed by solving analytically for the logit choice probabilities and integrating out the random taste distribution by taking draws from the joint distribution of unobserved terms.

Results are presented in Table 3. I estimate the model with time period interactions for the period before (1994–1995) and following the release of the 1998 report card, but prior to the release of the ensuing report cards (2000 and the first two quarters of 2002). The estimates are generally consistent with expectations. Travel distance enters choice significantly, as does surgeon quality. The effect of the introduction of quality reporting on demand can be seen in the interactions of the quality variables with the dummy for post-quality reporting. Columns 3 and 4 include controls for agency induced surgeon-patient matching. Because these effects are significant in all specifications, I include these controls in all subsequent analysis.

C. *Calibrating Report Card Related Extrinsic Incentives*

To calibrate the magnitude of the demand side incentives facing each surgeon, I simulate a series of counter-factual scenarios. Extrinsic incentives are captured by the measured returns to quality identified by the demand system.

The impact of information on profit is measured empirically by computing a predicted choice probability under alternate information scenarios—with and without quality reporting. The estimate for the report card induced change in profit is

$$(11) \quad \Delta\Pi_s = \lambda_1 \left[\sum_{i=1}^{I} \hat{P}_{i,s,h,t}(X_i^t, Z_{s,h}^t, \theta_{s,t} \,|\, \Omega^{pre}) - \sum_{i=1}^{I} \hat{P}_{i,s,h}(X_i^t, Z_{s,h}^t, \theta_{s,t} \,|\, \Omega^{post}) \right]$$

$$+ \lambda_2 I_s^{Dem\uparrow} + \lambda_3 I_s^{Dem\uparrow} \times \left[\sum_{i=1}^{I} \hat{P}_{i,s,h,t}(X_i^t, Z_{s,h}^t, \theta_{s,t} \,|\, \Omega^{pre}) \right.$$

$$\left. - \sum_{i=1}^{I} \hat{P}_{i,s,h}(X_i^t, Z_{s,h}^t, \theta_{s,t} \,|\, \Omega^{post}) \right] + I_{hrr},$$

---

[23] Contained in such unobserved influences are agency, insurance network constraints, and patient-surgeon matching.

TABLE 3—DEMAND SYSTEM PARAMETER ESTIMATES

	log probability patient i selects surgeon s			
	Mean		SD		Mean		SD	
Mean surgeon RAMR $s, t-1$	−0.001	(0.002)	−0.001	(0.003)	−0.001	(0.002)	0.000	(0.003)
Mean surgeon RAMR $s, t-1$ × post RC	−0.008**	(0.004)	−0.001	(0.008)	−0.008**	(0.004)	−0.001	(0.008)
Distance (miles) i, s	−0.216***	(0.006)	0.014	(0.011)	−0.212***	(0.006)	−0.007	(0.005)
Distance (miles) i, s × post RC	0.072***	(0.010)	0.007	(0.013)	0.073***	(0.010)	−0.004	(0.008)
Distance squared (miles) i, s	−0.001***	(0.000)	0.005***	(0.000)	−0.001***	(0.000)	0.005***	(0.000)
Distance squared (miles) i, s × post RC	−0.002***	(0.000)	0.000*	(0.000)	−0.002***	(0.000)	0.000	(0.000)
EMR i−EMR $s, t-1$					1.084**	(0.516)	0.092	(0.886)
EMR i−EMR $s, t-1$ × post RC					−3.616***	(1.173)	0.208	(2.145)
EMR i−EMR $s, t-1^2$					−17.322***	(3.483)	−0.308	(4.919)
EMR i−EMR $s, t-1^2$ × post RC					−17.107*	(9.839)	−5.032	(14.717)
Low quality RC score 1994–1995 s	−1.059***	(0.121)	1.540***	(0.154)	−0.499***	(0.075)	0.601***	(0.188)
Low quality RC score 1994–1995 s × post RC	−0.165*	(0.087)	0.238	(0.225)	−0.002	(0.068)	−0.046	(0.238)
High quality RC score 1994–1995 s	0.283***	(0.024)	0.017	(0.112)	0.271***	(0.024)	−0.033	(0.095)
High quality RC score 1994–1995 s × post RC	−0.130	(0.105)	0.277	(0.511)	−0.097*	(0.055)	−0.087	(0.377)
Not in report card 1994–1995 s	0.137***	(0.023)	0.043	(0.091)	0.131***	(0.023)	0.022	(0.090)
Not in report card 1994–1995 s × post RC	0.079*	(0.042)	0.064	(0.192)	0.067	(0.049)	0.193	(0.243)
Observations	720,364					
log likelihood	−86606.20					
Sample	1994–1995, 2000, 2002 (Q1–2)					

(Right-hand block: Observations 720,364; log likelihood −86570.20; Sample 1994–1995, 2000, 2002 (Q1–2))

Notes: Distance is computed from the center of patients' zip codes to the hospital at which the surgeon performed surgery. EMR $s, t-1$ is the mean expected mortality rate (severity) of patients treated by surgeon s in the prior quarter. EMR i − EMR $s, t-1$ is the absolute value of the difference. Post RC takes a value of 1 if period t is in the post-report card period (2000, 2002 (Q1–2)). Low and High quality scores are dummy variables for surgeons who were flagged as having higher (lower) mortality than expected. Not in report cards takes a value of 1 for surgeons not included in the quality report card.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

where $\sum_{i=1}^{I} \hat{P}_{i,s,h}(X_i^t, Z_{s,h}^t, \theta_{s,t} | \Omega^{RC})$ is the sum of the fitted choice probabilities for all patients i receiving surgery in the HRR in which surgeon s practiced in period t, a measure of expected demand. Ω^{RC} indexes the information environment (pre- or post-report cards) that defines the consumer utility parameters. $I_s^{Dem\uparrow}$ is a dummy variable equal to one if surgeon s has an expected demand using post-reporting demand parameters that is greater than expected volume using pre-report card demand (a windfall profit from quality reporting).[24] Including $I_s^{Dem\uparrow}$ and interacting it with the change in predicted volume allows the model to flexibly capture potentially discontinuous changes in incentives at a surgeon's current volume due to income effects (an issue discussed in more detail below). I also include market fixed effects to control for all time invariant market level factors that influence the profitability of quality. In this model, differences in the profitability of quality for surgeons are identified by differences in geographic locations, the distribution of patients' tastes, the competitive structure of the market, and (in models without a control for 1994–1995 *RAMR*) baseline quality. For example, a surgeon who is at a low quality level in the pre-reporting period and faces few high quality competitors

---

[24] The difference in volume is measured with respect to the predicted volume pre-quality reporting. Surgeons expecting to gain volume after quality reporting ($I_s^{Dem\uparrow} = 1$) have *negative* estimates for the volume difference $\left(\sum_{i=1}^{I} \hat{P}_{i,s,h,t}(X_i^t, Z_{s,h}^t, \theta_{s,t} | \Omega^{pre}) - \sum_{i=1}^{I} \hat{P}_{i,s,h}(X_i^t, Z_{s,h}^t, \theta_{s,t} | \Omega^{post}) < 0 \right)$.

and patients who (given their locations and attributes) respond strongly to quality reporting has a greater return to improving performance.

Surgeon response to pecuniary rewards for quality depends on the strength of income and substitution effects (McGuire and Pauly 1991). If income effects dominate, reductions in demand at a surgeon's current level of quality should alter extrinsic incentives more than the opportunity to gain additional patients by improving performance. Losing volume (profit) increases the marginal utility of an additional patient, mediated through the marginal utility of income.[25] On the other hand, if income effects are not sufficiently strong, demand side incentives are better measured by computing the ceteris paribus returns to a reduction in *RAMR*. Equation (11) models both effects by allowing incentives to differ for surgeons expecting to gain and expecting to lose patients after quality reporting (captured by the coefficients λ_2 and λ_3).

Using the approach in (11) and the full first stage demand system (including agency controls), the sample mean predicted quarterly change in volume is 0.14 with a standard deviation of 3.09. The small average effect (statistically indistinguishable from zero) is offset by a large variance across surgeons. Figure 6 plots the frequency of quarterly differences between surgeons' pre- and post-report card predicted volumes. Breaking the impact down by positive and negative demand effects, the mean predicted change in demand is 1.3 patients for surgeons losing volume in a given quarter, and a gain of 2.5 patients for surgeons gaining volume. Compared to the average quarterly volume of 30 surgeries in the sample, these effects are economically significant.

To put the financial returns in perspective, Huckman (2006) finds that the marginal profit from a CABG surgery is $6,900.[26] Thus, surgeons facing the average decline in demand expected a reduction in annual profit (joint with the hospital at which they practice) of $35,000 and surgeons gaining volume expected an annual profit increase of $69,000. To get an idea of the share of this profit that might accrue to a surgeon, we can compare the total profit to the Medicare fees provided to surgeons performing a single CABG (the dollar converted Relative Value Units associated with CABG with a single arterial bypass). The surgeon fee was $2,126 in 1994, $2,359 in 1995, and $2,009 in 2000 (Federal Register 1993; 1994; 1999).[27] Comparing these fees to the estimated marginal profit suggests that surgeons retained roughly one third of the profits associated with an additional CABG procedure in direct compensation.[28]

D. *Mixed Incentives and Quality Improvement*

Using this measure of profit incentives and the information contained in the risk adjustment, I estimate equation (6). The coefficients λ and ξ_n capture the effect of a

[25] The extreme version of this argument is consistent with the classic target income hypothesis (see McGuire 2000 for a review of a long literature).

[26] This is computed using data from New York and is a measure of total marginal profit to the hospital and the surgeon. Average profit is $2,000 for a Medicare beneficiary and $3,900 across all payer types.

[27] I thank Josh Gottlieb for sharing the CABG specific RVU fees.

[28] If the freelance surgeon bills the patient directly in addition to the estimated marginal profit the full fee would go directly to the surgeon in addition to the profit to the hospital. In this case, the share of the marginal profit would be roughly 24 percent ($= 2,200/9,100$). In both cases, the marginal incentives are substantial.
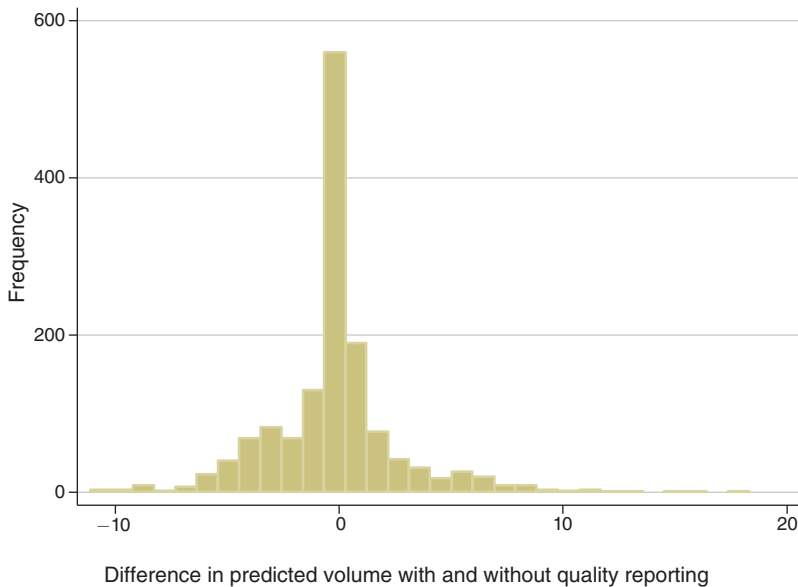
FIGURE 6. FREQUENCY OF QUARTERLY DIFFERENCES BETWEEN SURGEONS'
PRE- AND POST-REPORT CARD PREDICTED DEMAND

change in profit and a change in information respectively on the change in surgeon *RAMR* between 1994–1995 and 2000. I account for measurement error from the first stage demand estimates by using bootstrap simulation to compute the standard errors. Table 4 presents results.

Parameter estimates for ξ_n are negative for all surgeon information groups. The effects are large and significant for information groups 1 and 2 (surgeons learning their performance was lower than expected). I fail to reject the hypothesis H_0—that information increased intrinsic incentives for quality—for surgeons in groups 1 and 2. Interpreting the coefficient estimates in Table 5 suggests that, after controlling for demand side incentives and mean reversion, surgeons who learn significantly more about their own quality, and that of their peers, and find that they are not performing as well as expected (group 1), improve quality by an average of 0.57 to 0.62 percentage points more than surgeons receiving no information from reporting. In columns 1 and 3, those without controls for 1994–1995 *RAMR*, the impact of learning is much larger for the same group, around 2.5 percentage points. The same learning, but for those only slightly worse than expected, produced an average improvement of between 0.71 and 0.68 percentage points of *RAMR* with and without market fixed effects respectively. The magnitude of the response is relatively similar including a control for pre-reporting quality.[29] Finally, surgeons learning they were much better than expected improved by an average additional 1 percentage point

---

[29] These results are robust to alternate specifications of baseline *RAMR*. In unreported regressions, I allow for quadratic and nonparametric (quintiles) specifications of mean *RAMR* in 1994–1995. The results are qualitatively unchanged relative to a linear control.

TABLE 4—EFFECT OF REPORT CARD INDUCED INFORMATION AND DEMAND ON SURGEON QUALITY

	Change RAMR s 1994–1995 to 2000							
	(1)		(2)		(3)		(4)	
Intrinsic incentives								
1994–1995 report card info (RAMR-OMR) group								
Much better than expected (0–20%)	−1.016**	(0.407)	−0.213	(0.350)	−0.988***	(0.345)	−0.198	(0.286)
Slightly better than expected (20–40%)	−0.332	(0.334)	−0.241	(0.320)	−0.417	(0.374)	−0.305	(0.302)
Slightly worse than expected (60–80%)	−0.855***	(0.304)	−0.709**	(0.280)	−0.854***	(0.292)	−0.681**	(0.288)
Much worse than expected (80–100%)	−2.460***	(0.373)	−0.827**	(0.385)	−2.672***	(0.391)	−0.923***	(0.343)
Extrinsic incentives								
Pred Vol No RC − Pred Vol RC	−0.037	(0.028)	−0.015	(0.035)	−0.023	(0.039)	−0.007	(0.030)
Increased demand with RC (I[RCDem > 0])	0.262	(0.190)	0.066	(0.161)	0.279	(0.202)	0.084	(0.183)
I[RCDem > 0] × Pred Vol No RC − Pred Vol RC	−0.137**	(0.063)	−0.150**	(0.067)	−0.118*	(0.070)	−0.143**	(0.071)
Controls								
Mean RAMR 1994–1995			−0.695***	(0.068)			−0.699***	(0.057)
Surgeon license year (PA)	0.287***	(0.087)	0.292***	(0.067)	0.288***	(0.090)	0.294***	(0.073)
Surgeon license year (PA)^2	−0.007***	(0.002)	−0.006***	(0.001)	−0.007***	(0.002)	−0.006***	(0.002)
Publications	0.221***	(0.067)	0.175***	(0.074)	0.226***	(0.066)	0.180***	(0.063)
Market fixed effects?	No		No		Yes		Yes	
Observations (surgeon/quarter)	1,572		1,572		1,572		1,572	
R^2	0.1729		0.3491		0.1896		0.3641	

Notes: Observations are at the surgeon quarter level. Changes in RAMR are computed such that negative coefficient represent lower mortality (improvements in quality). Pred Vol No RC − Pred Vol RC is positive for surgeons who expect to lose volume due to reporting. For surgeons who expect to gain volume, the difference in volume is multiplied by −1 so positive values represent quality improvements. I[RCDem > 0] if equal a dummy equal to 1 for surgeons expecting increased volume with report cards. Block bootstrap standard errors clustered at the surgeon year level in parentheses.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

compared to surgeons gaining no new information, though this effect is ameliorated by including pre-reporting average *RAMR*.

I next turn to the estimates of the λ coefficients that capture surgeon response to profit incentives. Profit incentives enter with the "right" sign—surgeons facing more responsive patients improve quality by more. The coefficient estimates of λ_1 suggest that for every additional patient a surgeon expected to lose per quarter under quality reporting (given ex ante surgeon, competitor, and consumer characteristics) they reduced *RAMR* by an average of 0.04 to 0.01 additional percentage points. Estimates of λ_2 are insignificant, suggesting that the response does not drop discontinuously when surgeons expect to maintain their pre-report card volume (e.g., income effects are relatively small). Consistent with a stronger substitution effect, surgeons expecting to gain patients are also observed to respond to greater report card-induced demand response. This can be seen in the significant estimates of λ_3 across all specifications. In Table 5, the extrinsic response effects are similar even after controlling for baseline quality, suggesting the differences are driven by surgeons responding to patients who are more quality elastic, not merely by the mechanical relationship between current quality and demand. Overall, surgeons

TABLE 5—EFFECT OF REPORT CARD INDUCED INFORMATION AND DEMAND ON CHANGES IN PATIENT SEVERITY

	Change in share of patients with specific comorbidities/risk adjusters 1994–1995 to 2000							
	"Used" risk adjusters		"Used" risk adjusters		"Unused" risk adjusters		"Unused" risk adjusters	
Intrinsic incentives								
1994–1995 report card info (RAMR-OMR) group								
Much better than expected (0–20%)	0.032	(0.025)	0.041*	(0.024)	0.009	(0.018)	−0.014	(0.016)
Slightly better than expected (20–40%)	0.049**	(0.019)	0.050**	(0.023)	−0.006	(0.019)	−0.010	(0.013)
Slightly worse than expected (60–80%)	0.036*	(0.021)	0.040*	(0.022)	0.033	(0.020)	0.021	(0.013)
Much worse than expected (80–100%)	0.042**	(0.019)	0.060**	(0.029)	−0.037**	(0.018)	−0.082***	(0.024)
Extrinsic incentives								
Pred Vol No RC − Pred Vol RC	−0.002	(0.002)	−0.002	(0.002)	−0.002	(0.002)	−0.002	(0.002)
I[RCDem > 0]	0.034***	(0.009)	0.031***	(0.008)	−0.007	(0.009)	0.002	(0.009)
I[RCDem > 0] × Pred Vol No RC − Pred Vol RC	−0.006*	(0.004)	−0.007	(0.004)	0.003	(0.004)	0.004	(0.004)
Controls								
Mean RAMR 1994–1995			−0.007	(0.006)			0.019**	(0.007)
Surgeon license year (PA)	0.001	(0.007)	0.001	(0.006)	0.018***	(0.006)	0.018***	(0.004)
Surgeon license year (PA)^2	0.000	(0.000)	0.000	(0.000)	0.000***	(0.000)	0.000***	(0.000)
Publications	0.005	(0.005)	0.004	(0.004)	0.005	(0.004)	0.007**	(0.003)
Observations (surgeon/quarter)	1,590		1,590		1,590		1,590	
R^2	0.0353		0.0475		0.0760		0.1545	

Notes: Observations are at the surgeon quarter level. Block bootstrap standard errors in parentheses are clustered at the surgeon year level. Shares are computed on a 0 to 1 scale. Pred Vol No RC − Pred Vol RC is positive for surgeons who expect to lose volume due to reporting. For surgeons who expect to gain volume, the difference in volume is multiplied by −1 so positive (negative) values represent increases (decreases) in the share of patients with used or unused observables. "Used" risk adjusters are those patient characteristics included in the final risk adjustment in the report card that were available in both 1994–1995 and 2000. They include cardiogenic shock, complicated hypertension, dialysis, concurrent PTCA, and gender. "Unused" observables are patient characteristics that were gathered in 1994–1995 and 2000 but not used in the risk adjustment scheme for the report card. They include cardiomyopathy, diabetes, and race.

   ***Significant at the 1 percent level.
   **Significant at the 5 percent level.
   *Significant at the 10 percent level.

expecting to gain more patients due to quality reporting improve by more, regardless of their current volume.

IV. Discussion

A. *Comparing Intrinsic and Extrinsic Response to Quality Reporting*

The evidence thus far is consistent with a role for both extrinsic and intrinsic incentives in determining surgeons' response to quality reporting. I use the estimates from equation (6) to calibrate the relative magnitudes of these incentives by fitting the model under alternative incentive scenarios (i.e., with and without extrinsic and intrinsic incentives). For brevity, I only report results from the analysis using full controls (the relevant coefficient estimates are in column 4 of Table 4).

Overall, the model predicts changes in *RAMR* relatively well. The coefficient of correlation between the predicted and actual change in *RAMR* for each surgeon is 0.89. The full model predicts a mean improvement of 0.64 percentage points of *RAMR* compared to the actual average of 0.72. Despite the relatively good fit, I use predicted improvement to compare the role of extrinsic and intrinsic incentives. As

long as the error is not correlated with the intrinsic or extrinsic incentive measures, the predicted changes are consistent.

I begin by computing the predicted change in quality had report cards not been implemented. Constraining $\lambda = 0$ and $\xi_n = \xi_3$ the average predicted change in *RAMR* between 1994–1995 and 2000 is 0.07 percentage points. Allowing for extrinsic incentives only (incorporating λ and continuing to set $\xi_n = \xi_3$), the predicted quality improvement is 0.19 percentage points of *RAMR*. Surgeons' response to profit incentives led to an additional decline in *RAMR* of 0.12 ($= 0.19 - 0.07$) percentage points. Compared to the mean statewide *RAMR* in 1994–1995 (3.42 percent), this constitutes an additional 3.5 percent decline in mortality.

Allowing only information-induced intrinsic incentives to enter the model (i.e., setting $\lambda = 0$) the model predicts a market wide average *RAMR* that is an average of 0.52 percentage points lower due to quality reporting. The intrinsic response to quality reporting alone led to an additional 0.45 ($= 0.52 - 0.07$) percentage point decline in *RAMR*, or 13 percent lower statewide *RAMR*. Comparing the two effects, the incremental impact of quality reporting mediated through intrinsic response is roughly four times as large as the response to profit incentives. Taken together, it is clear that report cards had an important role in quality improvement for CABG surgery in Pennsylvania. The bulk of that impact, though, is driven by information observed by surgeons and not by incentives generated by consumer response to reporting.

B. *Incentives for Selecting Healthier Patients*

A concern in any effort to measure performance is the potential for efforts to game the rating system. This has been widely discussed in the context of quality reporting in health care generally and in CABG surgery specifically (e.g., Dranove et al. 2003), raising the question of whether the results in this paper reflect patient selection rather than true quality improvement. This section focuses on understanding whether the observed improvements following quality reporting can be explained by selection efforts by surgeons.

The detailed data on underlying patient characteristics allow me to address the question of selection directly as a potential explanation for the observed quality improvement in Table 4. To do so, I rely on a test that is similar to the test for selection in insurance markets using "unused observables" (Finkelstein and Poterba 2006). Specifically, I model changes in surgeon's share of patients with risk factors that were used to model patient severity ("used" patient observables) as well as candidate variables for risk adjustment that were gathered but not used in the final model ("unused" patient observables). That is, the PHC4 data contain a set of patient comorbidites that were thought at the outset to be correlated with mortality, but were not ultimately used because they were found not to be sufficiently predictive.[30] If surgeons were selecting on patient characteristics that they believed

---

[30] The unused observables in 1994–1995 and 2000 include whether a patient had cardiomyopathy, diabetes, COPD, or peripheral vascular disease as well as race and obesity. Unfortunately, data on peripheral vascular disease, obesity, and COPD were not collected in the 1994–1995 data. Thus, the variables used in computing patient shares with "unused" observables are based on cardiomyopathy, diabetes, and race.

were correlated with risk but not captured by risk adjustment these are precisely the measures we would expect to change for surgeons whose *RAMR* declined due to seeing healthier patients. To do so, I estimate a variant of equation (6) that replaces the dependent variable with the change in share of patients treated by surgeon s who had at least one comorbidity r between the pre- and post-reporting periods. The estimating equation of interest is thus

$$(12) \qquad \Delta Share_{s,r} = \alpha + \lambda \Delta \Pi_s + \mu_n \sum_{n=1}^{5} I_{n,s} (RAMR_{pre} - OMR_{pre})$$

$$+ X_s + X_h + \varepsilon_{s,h},$$

where the independent variables are as in equation (6). The null hypothesis $H_0 : \xi_n = \xi_3$ is a test that the groups observed to improve due to intrinsic incentives did not differentially change the observable or unobservable severity of patients, they treated depending on the outcome variable. Similarly, we expect the coefficient λ to equal zero if differences in profit incentives did not lead surgeons to differentially change the composition of patients they treated.[31]

The results are presented in Table 5. The left two columns present results from estimating equation (12) using the share of patients with at least one comorbidity used to risk adjust CABG outcomes. The coefficients on the measures of new information due to reporting that drive intrinsic incentives are positive for all measures of new information. Depending on the specification, the coefficient estimates for ξ_n are statistically significant different from zero. This suggests that, if anything, surgeons who had more new information from reporting saw measurably *more* severe patients. The coefficient estimates of λ_2 and λ_3 (the impact of facing larger extrinsic incentives) are statistically significant. The coefficient on the dummy variable for facing higher demand with quality reporting is 0.034 with and 0.031 without controlling for baseline *RAMR*, suggesting that surgeons who stood to gain patient volume from quality reporting differentially saw more severe patients after report cards were released. This effect is declining in the magnitude of the expected gain in market share. The estimates of λ_3 are -0.006 and -0.007, depending on the specification, though only the estimate in column 1 is significant at conventional levels. Thus, for every additional patient expected surgeons gaining market share actually lowered the share of patients with an observable risk adjuster. Recall that the average surgeon gaining patients expected to increase volume by 2.5 surgeries per quarter. Scaling the λ_3 coefficients by 2.5, the net effect for the average surgeon expecting to gain volume on the change in the share of patients with an observable risk adjuster is a reduction in share of 1.9 ($= 0.034 - (2.5 \times 0.006)$) to 1.3 ($= 0.031 - (2.5 \times 0.007)$) percentage points. In the pre-reporting period the average share of patients with at least one observable risk adjuster for surgeons who stood to gain volume was 48 percent. Thus, the differential change in share of

---

[31] One thing to note is that this is not a test for the existence of any selection due to quality reporting overall. Rather, this test asks whether the main results in the paper (those in Table 5) can be explained by selection. If the entire market is shifting the types of patients receiving CABG due to selection incentives but surgeons with differing incentives do not do so differentially we will not find evidence for selection in this test.

patients with at least one risk adjuster for surgeons expecting to gain volume due to reporting was a decline of 2.7 to 4 percent.[32]

The right half of Table 5 extends the analysis to test not only for changes in the observable risk adjusters—where the risk adjustment scheme should mitigate selection incentives—but also for changes in the share of patients with "unused" risk adjusters. If surgeons have the ability to select patients who are healthier in unobservable (to the regulatory) ways, we expect to see changes in the share of patients with these comorbidities for those surgeons who improved performance due to either extrinsic or intrinsic incentives. The majority of coefficient estimates are small and statistically insignificant, suggesting that the improvements in quality observed are not due to significant changes in patient severity not included in risk adjustment. There are, however, some significant effects. Specifically, it appears that surgeons learning their performance was much worse than expected saw significantly fewer patients with at least one "unused" risk adjuster. The coefficient estimates of -0.037 and -0.082 with and without controls for baseline *RAMR* are both statistically and economically significant in terms of patient composition. Comparing this to the surgeon average of 31 percent of patients with at least one unused risk adjuster in 1994–1995 suggests this group saw a reduction of between 12 and 26 percent in the number of patients with an unused risk adjuster.

These results suggest efforts to select patients, one potential channel through which this group of surgeons improved performance. Incorporating the actual contribution of these measures of severity to outcomes, however, suggests that this large shift in patients was unlikely to explain the improvement in quality found in Section IVB. In Appendix A1, I replicate the PHC4 risk adjustment model with and without "unused" risk adjusters. These results make clear that the marginal contribution of avoiding a patient with any of the unused risk adjusters had a far smaller impact on expected mortality than the used risk adjusters and was unlikely to materially affect *RAMR*. Appendix Table A1 presents the model used in the 1994–1995 risk adjustment model. The fourth column presents the marginal effects of the full model with both used and unused risk adjusters. The unused risk adjusters are all an order of magnitude smaller than the used risk measures and none are statistically significant. If we take the upper bound on the marginal effect of any of the unused risk adjusters, a 0.6 percentage point increase in the probability of inpatient mortality for patients with cardiomyopathy, a reduction of 8 percentage points in the share of patients with this risk adjuster would translate to a reduction in *RAMR* of 0.05 percentage points. Compared to the estimated differential improvement of this group of surgeons of 0.92 percentage points (column 4 of Table 4), selection can only explain 5 percent of the observed improvement due to intrinsic incentives. Furthermore, this is a conservative estimate as the impact of cardiomyopathy is approximately six times larger than the marginal impact of the other unused risk

---

[32] In an additional specification (not reported), I reestimate the same model using only the share of patients with cardiogenic shock as the dependent variable. Conversations with CABG surgeons suggests they believe this is a key factor that would make a patient more difficult and, therefore, a characteristic that we would expect to see avoided if possible. The coefficient estimates for intrinsic and extrinsic incentives are not significantly different from zero.

adjusters, and it is unlikely the full share of patients who shifted away from surgeons learning they were worse than expected had this comorbidity.[33]

Taken together, these results suggest that surgeons with differing intrinsic incentives to improve quality do not appear to have accomplished reductions in mortality by changing their patient mix either based on observable or unobservable (to the regulator) patient severity. It does, however, appear that surgeons learning they were performing poorly did try to select unobservably healthier patients. Despite this, the actual contribution of this selection to quality improvement was negligible. If anything, this finding suggests selection could be an issue, but the risk adjustment scheme employed by PHC4 was sufficient to account for measures that can contribute significantly to mortality risk.

C. *Long Run/Dynamic Pecuniary Incentives*

Perhaps the most relevant threat to the validity of the model is the potential for dynamic pecuniary returns to quality improvement. If current period investments are related to future demand, what we are calling intrinsic incentives may in fact be quality improvement in response to expected future compensation. To provide a simple test for such long run pecuniary incentives, I assume that surgeons make decisions based on rational expectations of future returns to quality improvements.[34] I estimate the following model:

$$(13) \quad q_{s,t} = \alpha + \gamma_1 \Delta\theta_{s,t-y,t-n} + \gamma_2 q_{s,t-z} + \gamma_3 \theta_{s,t-n} + X_s + X_h + \varepsilon_{s,h},$$

where $q_{s,t}$ is the quantity of patients seen by surgeon s in period t and $\Delta\theta_{s,t-y,t-n}$ is the change in surgeon quality between period $t-n$ and $t-y$. z, n, and y are positive integers satisfying: $n > y$. Including the quantity of surgeries by surgeon s in all periods, $t-z$ controls for the full stock of demand and quality effects prior to period t. Because I am primarily interested in surgeons' response to quality reporting introduced in 1998, I estimate (13) using surgeon volume in 2003 as the dependent variable and measure the change in quality as the difference between a surgeon's 1994–1995 and 2000 *RAMR*. I control for baseline quality and surgeon volume in 1994, 1995, 2000, and 2002.

The estimated coefficient on $\Delta\theta_{s,2000,1994-1995}$ is -1.98 (s.e. $= 1.52$). The negative and statistically insignificant estimate suggests that surgeons were unlikely to expect changes in quality to produce pecuniary returns in the future. If anything, it appears that greater quality improvement led to a slightly smaller number of patients in the long run. I also estimate (12) with the change in quality between 1994 and 1995 as an independent variable. The estimated coefficient on $\Delta\theta_{s,1995,1994}$ is 0.43 (s.e. $= 0.6$), again an insignificant effect of the change in quality on subsequent

---

[33] Table A2 presents a similar risk adjustment model with and without unused risk adjusters using data from the post-report card period. The coefficient estimates for the unused risk adjusters remain largely the same with only the estimate for diabetes without complications changing. Since the marginal effect for this unused risk adjuster was actually to make patients *less* severe, avoiding those patients would not explain the observed improvement.

[34] This is a simplifying assumption, though standard in traditional models of incentives and investment. However, models that incorporate risk or ambiguity aversion are not being tested directly in equation (13). Appendix A3 discusses the different incentives, both extrinsic and intrinsic, that could be at play and the identifying variation and assumptions for each.

annual volume; in this case eight years hence. These results are consistent with Johnson (2010) who studies the impacts of market learning (as opposed to report card-based) about quality among cardiac surgeons on their career paths over time. She also finds no impact of quality (measured by mortality) on long run surgery volume in the Medicare market.

It is also plausible that the performance improvements do not change patient volume, but allow surgeons to gain more patients whose insurance provides better remuneration through surgical fees for CABG. The detailed data on patients' insurance allows us to test for this directly by estimating another variant of equation (13). I find little evidence for changes in payor share as a result of quality improvements. For brevity, results are presented in Appendix A2.

D. *Local Peers and Learning*

I turn next to a sensitivity test for the specific model that information alters surgeons' behavior by altering their knowledge about a reference group. The base specification (equation (6)) models surgeons' learning using the risk adjustment based on statewide performance. If surgeon incentives are related to their knowledge about the performance of their peers I also expect that new learning about other peer groups will alter the intrinsic incentives. The most easily defined peer group is the set of other surgeons practicing at the same hospital. To test for an effect of learning about within-hospital peer performance, I reestimate (6) and include a measure of the difference between surgeon *s' RAMR* in the pre-reporting period and the top (lowest *RAMR*) and bottom (highest *RAMR*) surgeon at the hospital at which they practiced in 1994–1995. The model is

$$(14) \quad \Delta\theta_s = \alpha + \lambda\Delta\Pi_s + \xi_n \sum_{n=1}^{5} I_{n,s}(RAMR_{pre} - OMR_{pre})$$

$$+ \eta(RAMR_{pre} - \min RAMR_{h,s,pre}) + \mu(RAMR_{pre} - \max RAMR_{h,s,pre})$$

$$+ X_s + X_h + \varepsilon_{s,h}.$$

Results are presented in Table 6 for models including baseline *RAMR*, with and without market fixed effects.

In both models, coefficient estimates for both η and μ are negative and significant, suggesting a surgeon learning he is further from either the best—opportunity to improve—or the worst surgeon at his hospital improves *RAMR* by relatively more. These results support the earlier findings, that report card learning alters incentives by improving knowledge about a reference group. Moreover, adding these controls produces larger coefficient estimates for suggesting that surgeons compare performance both to their peers at their own hospital and to the full statewide population.

V. **Conclusion**

The impact of information on equilibrium quality is mediated through demand insofar as suppliers choose quality levels to maximize profit. In this paper, I present

TABLE 6—EFFECT OF REPORT CARD INDUCED INFORMATION AND DEMAND ON SURGEON QUALITY
INCLUDING LEARNING ABOUT WITHIN HOSPITAL REFERENCE SURGEONS

	Change RAMR s 1994–1995 to 2000[1]			
	(1)		(2)	
Intrinsic incentives				
1994–1995 report card info (RAMR-OMR) spline				
Much better than expected (0–20%)	−0.360	(0.314)	−0.362	(0.278)
Slightly better than expected (20–40%)	−0.346	(0.308)	−0.396	(0.286)
Slightly worse than expected (60–80%)	−0.632**	(0.263)	−0.646**	(0.250)
Much worse than expected (80–100%)	−0.961***	(0.304)	−1.007***	(0.339)
RAMR s − Best RAMR h	−0.313**	(0.136)	−0.310**	(0.129)
RAMR s − Worst RAMR h	−0.041***	(0.008)	−0.038***	(0.008)
Extrinsic incentives				
Pred Vol No RC − Pred Vol RC	−0.040	(0.043)	−0.039	(0.037)
Increased demand with RC (I[RCDem > 0])	0.233	(0.191)	0.221	(0.172)
I[RCDem>0] × Pred Vol No RC − Pred Vol RC	−0.004	(0.067)	−0.013	(0.057)
Controls				
Mean RAMR 1994–1995	−0.406***	(0.152)	−0.417***	(0.134)
Surgeon license year (PA)	0.249***	(0.075)	0.259***	(0.076)
Surgeon license year (PA)^2	−0.005***	(0.002)	−0.005***	(0.002)
Publications	0.145**	(0.062)	0.148***	(0.056)
Publications^2	−0.007***	(0.002)	−0.007***	(0.002)
Market fixed effects?	No		Yes	
Observations (surgeon/quarter)	1,572		1,572	
R^2	0.3830		0.3873	

Notes: Observations are at the surgeon quarter level. Block bootstrap standard errors (in parentheses) are clustered at the surgeon year level. Changes in RAMR are computed such that negative coefficient represent lower mortality (improvements in quality). Pred Vol No RC − Pred Vol RC is positive for surgeons who expect to lose volume due to reporting. For surgeons who expect to gain volume, the difference in volume is multiplied by −1 so positive values represent quality improvements. I[RCDem > 0] if equal a dummy equal to 1 for surgeons expecting increased volume with report cards. RAMR s − Worst RAMR h multiplied by −1 so positive values represent quality improvement.

  ***Significant at the 1 percent level.
   **Significant at the 5 percent level.
    *Significant at the 10 percent level.

an alternate model in which reference intrinsic utility also determines surgeons' willingness to make costly quality improvements. Information (quality reporting) alters a surgeon's beliefs about their own quality level relative to a reference set of peers. If intrinsically motivated suppliers update their beliefs in a Bayesian fashion, the improved posterior beliefs alter the shape of the intrinsic utility function and thus a surgeon's marginal incentive to improve quality.

The risk-adjustment model that underlies quality report cards provides a simple way of identifying the magnitude of the new information provided to surgeons and its effect on performance. Surgeons who gain more information about their performance relative to their peers (from risk adjustment) improve significantly more. A structural patient choice model allows me to estimate the profit incentives from quality reporting. Conditioning on extrinsic incentives, the intrinsic response to information leads to significant declines in *RAMR* and is large relative to the response to profit motives. Improvements cannot be explained by surgeon efforts to shift the composition of the patients they treat, either on observable risk measures used in risk adjustment or on "unused" observables that could be used to game the quality measure.

The results of this paper add to the literature on the behavioral response to improved information in markets. However, I note some shortcomings and directions for future work. I consider a specific setting—cardiac surgery—that may not generalize to other markets or professions. The field of medicine, perhaps more than any other, relies on not-for-profit incentives to correct market failures. As trainees, physicians also become accustomed to evaluation mechanisms based on relative performance (grades and MCAT for medical school, board scores for residency and licensure, etc). For these reasons, the relative magnitude of intrinsic versus extrinsic effects I find among cardiac surgeons may be larger than in other settings.

Nevertheless, these results provide an empirical first step and guidance on the potential role for mixed incentives in determining skilled professionals' effort and investment. The findings also contribute to our understanding of the effects of quality reporting in health care and inform policy making in this market.

Appendix A1: Risk Adjustment Model

In this section, I replicate the risk adjustment model used by the Pennsylvania Health Care Cost Containment Council (PHC4) to predict inpatient mortality following CABG surgery in 1994–1995. The details of the risk adjustment scheme are covered in PHC4 (1998). To replicate the results, I estimate a logistic regression following equation (4). Results are presented in Table A1. The first column presents coefficients for the logistic regression. The second column presents the marginal effects of each of the observables on the probability of inpatient mortality. The third and fourth columns present the same model but also incorporate the unused risk adjusters used in the analysis in Section IVB. In addition, I reestimate the same model, used in the 1994–1995 report card, using data from the year 2000. This demonstrates the relative impact of the same comorbidities in the post-reporting period. The models are not exactly comparable as PHC4 switched from using ASG severity scores, coded in discrete levels between zero and four in Table A1, to using a continuous severity measure provided by Mediqual. Nevertheless, the estimates appear relatively similar. These results are presented in Table A2.

Appendix A2: Payer Mix and Reimbursement

In this section, I test the long run impact of quality improvement on the mix of insurance coverage a surgeon treats. To do so, I estimate a variant of equation (13). In the new specification the dependent variable is the probability a surgeon in each group sees a patient with insurance plan type z.

Results are presented in Table A3. Each column presents one regression in which changes in the share of patients seen by each surgeon in 2003 is regressed against improvements in quality between 1994–1995 and 2000, controlling for the payer shares in those earlier periods and baseline *RAMR* in 1994–1995. The coefficients on most types of coverage are small and statistically insignificant. The only changes in coverage type that are statistically significant are for the share being treated with either Medicare HMO or Medicaid. Quality improvements in earlier periods seem to lead to slightly larger share of Medicare HMO enrollees and a smaller share of Medicaid patients. These effects are, however, small in magnitude. Taken together

TABLE A1—RISK ADJUSTMENT MODEL USED IN THE 1994–1995 REPORT CARD
WITH AND WITHOUT "UNUSED OBSERVABLES" USING YEAR 1994–1995 OBSERVATIONS

	Inpatient mortality			
	logit (1)	Marginal effect (2)	logit (3)	Marginal effect (4)
ASG score 1 (d)	0.656 (1.02)	0.011 (0.02)	0.644 (1.02)	0.011 (0.02)
ASG score 2 (d)	1.323 (1.02)	0.029 (0.03)	1.300 (1.02)	0.028 (0.03)
ASG score 3 (d)	2.015** (1.02)	0.088 (0.09)	1.989* (1.02)	0.086 (0.09)
ASG score 4 (d)	3.767*** (1.03)	0.417* (0.25)	3.739*** (1.03)	0.409 (0.25)
Age	−0.061* (0.04)	−0.001* (0.00)	−0.062* (0.04)	−0.001* (0.00)
Age squared	0.066** (0.03)	0.001** (0.00)	0.068*** (0.03)	0.001** (0.00)
Cardiogenic shock (d)	1.758*** (0.14)	0.076*** (0.01)	1.754*** (0.14)	0.076*** (0.01)
Concurrent PTCA (d)	0.436*** (0.13)	0.009*** (0.00)	0.450*** (0.13)	0.010*** (0.00)
Complicated hypertension (d)	0.533*** (0.14)	0.012*** (0.00)	0.507*** (0.14)	0.011*** (0.00)
Dialysis (d)	1.814*** (0.14)	0.081*** (0.01)	1.799*** (0.14)	0.079*** (0.01)
Female (d)	0.448*** (0.06)	0.009*** (0.00)	0.436*** (0.06)	0.008*** (0.00)
Heart failure (d)	0.797*** (0.07)	0.018*** (0.00)	0.778*** (0.07)	0.018*** (0.00)
Prior CABG or valve surgery (d)	1.294*** (0.08)	0.041*** (0.00)	1.299*** (0.08)	0.041*** (0.00)
Non-white (d)			0.066 (0.11)	0.001 (0.00)
Diabetes without complications (d)			0.068 (0.07)	0.001 (0.00)
Diabetes with complications (d)			0.124 (0.13)	0.002 (0.00)
Cardiomyopathy (d)			0.316* (0.18)	0.006 (0.00)
Observations	38,656	38,656	38,656	38,656

Notes: Includes all observations from 1994–1995. logit models present coefficients estimates from a logistic regressions with inpatient mortality as the outcome variable. Even numbered columns present marginal impact of each risk adjuster on the probability of inpatient mortality. ASG risk scores are supplied by PHC4 directly. (d) for discrete change of dummy variable from zero to one.

   ***Significant at the 1 percent level.
   **Significant at the 5 percent level.
   *Significant at the 10 percent level.

TABLE A2—RISK ADJUSTMENT MODEL USED IN THE 1994–1995 REPORT CARD
WITH AND WITHOUT "UNUSED OBSERVABLES" USING YEAR 2000 OBSERVATIONS

	Inpatient mortality			
	logit (1)	Marginal effect (2)	logit (3)	Marginal effect (4)
MediQual CABG severity	6.993*** (0.53)	0.152*** (0.01)	6.875*** (0.53)	0.148*** (0.01)
Age	0.011 (0.04)	0.000 (0.00)	0.021 (0.04)	0.000 (0.00)
Age squared	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)
Cardiogenic shock (d)	2.006*** (0.20)	0.121*** (0.02)	1.991*** (0.20)	0.118*** (0.02)
Concurrent PTCA (d)	0.264 (0.32)	0.007 (0.01)	0.281 (0.32)	0.007 (0.01)
Complicated hypertension (d)	0.333** (0.15)	0.008* (0.00)	0.314** (0.16)	0.008* (0.00)
Dialysis (d)	1.621*** (0.18)	0.080*** (0.02)	1.610*** (0.18)	0.078*** (0.02)
Female (d)	0.223*** (0.08)	0.005*** (0.00)	0.245*** (0.08)	0.006*** (0.00)
Heart failure (d)	0.807*** (0.08)	0.022*** (0.00)	0.805*** (0.08)	0.022*** (0.00)
Prior CABG or valve surgery (d)	0.296** (0.14)	0.007* (0.00)	0.304** (0.14)	0.007** (0.00)
Non-white (d)			0.175* (0.10)	0.004* (0.00)
Diabetes without complications (d)			−0.325*** (0.10)	−0.007*** (0.00)
Diabetes with complications (d)			−0.031 (0.15)	−0.001 (0.00)
Cardiomyopathy (d)			0.225 (0.20)	0.005 (0.01)
Observations	22,856	22,856	22,856	22,856

Notes: Includes all observations from 2000. logit models present coefficients estimates from a logistic regressions with inpatient mortality as the outcome variable. Even numbered columns present marginal impact of each risk adjuster on the probability of inpatient mortality. Risk adjustment model based on the model used by PHC4 in 1994–1995 but includes a continuous measure of severity, "MediQual CABG Severity" score, instead of the ASG score. The severity score replaced the discrete ASG scores in data provided by PHC4. (d) for discrete change of dummy variable from 0 to 1.

   *** Significant at the 1 percent level.
    ** Significant at the 5 percent level.
     * Significant at the 10 percent level.

these results do not suggest substantial long run impacts on insurance composition for those surgeons who improved quality between 1994–1995 and 2000.

APPENDIX A3: TYPES OF INCENTIVES AND IDENTIFICATION

The approach taken both in the theoretical and empirical model of intrinsic response to quality reporting is relatively general. Both focus on identifying and separating the response of surgeons to new information that is unrelated to changes in patient demand from the change due to changes in patient demand. Underlying

TABLE A3—EFFECT OF CHANGES IN QUALITY ON LONG RUN (2003) PAYOR MIX

| | Share of patients with payor type 2003 | | | | | |
	Private FFS	Private managed care	Medicare FFS	Medicare HMO	Medicaid	Uninsured
Change RAMR 1994–1995 to 2000	−0.001 (0.001)	0.002 (0.001)	−0.002 (0.002)	0.003** (0.001)	−0.001* (0.001)	0.000 (0.000)
Share payor type 1994–1995	0.202 (0.131)	−0.133 (0.162)	0.267 (0.195)	0.244 (0.186)	0.478** (0.220)	−0.003 (0.038)
Share payor type 2000	0.254** (0.124)	0.334 (0.244)	0.109 (0.235)	1.211*** (0.379)	0.143 (0.245)	0.009 (0.204)
Mean RAMR 1994–1995	0.000 (0.001)	0.000 (0.002)	−0.004 (0.003)	0.000 (0.002)	0.002** (0.001)	0.000 (0.000)
Observations (surgeon/quarter)	1,242	1,242	1,242	1,242	1,242	1,242
R^2	0.102	0.022	0.060	0.173	0.064	0.003

Notes: Observations are at the surgeon quarter level. Bootstrap standard errors (in parentheses) are clustered at the surgeon level. Shares are computed on a 0 to 1 scale. Each column is the change in share for the payor type listed in the header in 2003.

*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

TABLE A4—COMPONENTS OF EXTRINSIC AND INTRINSIC INCENTIVES AND IDENTIFICATION

Extrinsic	Identification	Intrinsic	Identification
Profit	Change in volume with and without quality reporting given market structure. Assume CABG is profitable on average and at the margin.	Performance relative to peers	Difference between prior beliefs (OMR) and information on average peers' performance had they treated the same patients (RAMR). Assume surgeons do not know RAMR perfectly.
Long run profit	Relationship between RAMR change from 1994–1995 to 2000 and 1994 to 1995 and change in volume and payor mix in 2003. Condition on short run changes due to the release of report cards.	Ability to observe performance relative to peers	Release of report cards with information. Assume performance of peers, given the set of patients they treat, is imperfectly observed. Specification check controls for local peer performance.
Risk aversion	Compare surgeons expecting to gain volume to those expecting to lose volume.		
Beliefs about long run incentives	Assume rational expectations among surgeons. Assume no long run risk or ambiguity aversion.		

this general question, different components enter into intrinsic and extrinsic incentives in theory. Those components are identified with different sources of variation and assumptions. To help clarify each of these components, Table A4 enumerates each of the features that enter into extrinsic (column 1) and intrinsic (column 3) motivation in the surgeon setting. Columns 2 and 4 then present a summary of the key identifying assumptions and variation for each of these features.

REFERENCES

Afendulis, Christopher C., and Daniel P. Kessler. 2007. "Tradeoffs from Integrating Diagnosis and Treatment in Markets for Health Care." *American Economic Review* 97 (3): 1013–20.

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (3): 488–500.

Arrow, Kenneth J. 1962. "The Economic Implications of Learning by Doing." *Review of Economic Studies* 29 (3): 155–73.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53 (5): 941–73.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2005. "Social Preferences and the Response to Incentives: Evidence from Personnel Data." *Quarterly Journal of Economics* 120 (3): 917–62.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–90.

Bumiller, Elisabeth. 1995. "Death Rate Rankings Shake New York Cardiac Surgeons." *New York Times*, September 6.

Chatterji, Aaron K., and Michael W. Toffel. 2007. "Shamed and Able: How Firms Respond to Information Disclosure." Corporate Social Responsibility Initiative Working Paper 39.

Chernew, Michael, Gautam Gowrisankaran, and Dennis P. Scanlon. 2001. "Learning and the Value of Information: The Case of Health Plan Report Cards." National Bureau of Economic Research Working Paper 8589.

Chernew, Michael, Dennis Scanlon, and Rod Hayward. 1998. "Insurance Type and Choice of Hospital for Coronary Artery Bypass Graft Surgery." *Health Services Research* 33 (3): 447–66.

Cutler, David M., Robert S. Huckman, and Jonathan T. Kolstad. 2010. "Input Constraints and the Efficiency of Entry: Lessons from Cardiac Surgery." *American Economic Journal: Economic Policy* 2 (1): 51–76.

Cutler, David M., Robert S. Huckman, and Mary Beth Landrum. 2004. "The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery." *American Economic Review* 94 (2): 342–46.

Dranove, David, Daniel Kessler, Mark McClellan, and Mark Satterthwaite. 2003. "Is More Information Better? The Effects of "Report Cards" on Health Care Providers." *Journal of Political Economy* 111 (3): 555–88.

Dranove, David, and Mark A. Satterthwaite. 1992. "Monopolistic Competition When Price and Quality Are Imperfectly Observable." *RAND Journal of Economics* 23 (4): 518–34.

Dranove, David, and Andrew Sfekas. 2008. "Start Spreading the News: A Structural Estimate of the Effects of New York Hospital Report Cards." *Journal of Health Economics* 27 (5): 1201–07.

Dziuban, Stanley W., Joseph B. McIlduff, Stuart J. Miller, and Richard H. Dal Col. 1994. "How a New York Cardiac Surgery Program Uses Outcomes Data." *The Annals of Thoracic Surgery* 58 (6): 1871–1876.

Edmondson, Amy C., Richard M. Bohmer, and Gary P. Pisano. 2001. "Disrupted Routines: Team Learning and New Technology Implementation in Hospitals." *Administrative Science Quarterly* 46 (4): 685–716.

Ellis, Randall P., and Thomas G. McGuire. 1986. "Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply." *Journal of Health Economics* 5 (2): 129–51.

Epstein, Andrew. 2006. "Do Cardiac Surgery Report Cards Reduce Mortality? Assessing the Evidence." *Medical Care Research and Review* 63 (4): 403–426.

Fan, Jianqing, and Irene Gijbels. 1996. *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Federal Register. 1993. Revisions to Payment Policies and Adjustments to the Relative Value Unity Under the Physician Fee Schedule for Calendar Year 1994. 58 Federal Registrar 63626 (December 2).

Federal Register. 1994. Physician Fee Schedule for Calendar Year 1995. 59 Federal Registrar 235 (December 8).

Federal Register. 1999. Revisions to Payment Policies Under the Physician Fee Schedule for Calendar Year 2000. 64 Federal Registrar 2011 (November 2).

Fehr, Ernst, and Klaus M. Schmidt. 2006. "The Economics of Fairness, Reciprocity and Altruism—Experimental Evidence and New Theories." In *Handbook of the Economics of Giving, Altruism and Reciprocity*, edited by Serge-Christophe Kolm and Jean Mercier Ythier, 615–91. Amsterdam: Elsevier.

Finkelstein, Amy, and James Poterba. 2006. "Testing for Selection Using "Unused Observables."" National Bureau of Economic Research Working Paper 12112.

Fong, Kyna G. 2008. "Evaluating Skilled Experts: Optimal Scoring Rules for Surgeons." Unpublished.

Frank, Richard G., and Richard J. Zeckhauser. 2007. "Custom-Made versus Ready-to-Wear Treatments: Behavioral Propensities in Physicians' Choices." *Journal of Health Economics* 26 (6): 1101–27.

Fung, Archon, Mary Graham, and David Weil. 2007. *Full Disclosure: The Perils and Promise of Transparency*. New York: Cambridge University Press.

Gaynor, Martin. 2006. "What Do We Know About Competition and Quality in Health Care Markets?" National Bureau of Economic Research Working Paper 12301.

Ghali, William, Arlene Ash, Ruth Hall, Mark Moskowitz. 1997. "Statewide Quality Improvement Initiatives and Mortality After Cardiac Surgery." *Journal of the American Medical Association* 277 (5): 379–82.

Gowrisankaran, Gautam, Vivian Ho, and Robert J. Town. 2006. "Causality, Learning and Forgetting in Surgery." Unpublished.

Hannan, Edward L., Harold Kilburn, Michael Racz, Eileen Shields, and Mark R. Chassin. 1994. "Improving the Outcomes of Coronary Artery Bypass Surgery in New York State." *Journal of the American Medical Association* 271 (10): 761–6.

Hannan, Edward L., Chuntao Wu, Thomas J. Ryan, Edward Bennett, Alfred T. Culliford, Jeffrey P. Gold, Alan Hartman, et al. 2003. "Do Hospitals and Surgeons With Higher Coronary Artery Bypass Graft Surgery Volumes Still Have Lower Risk-Adjusted Mortality Rates?" *Circulation* 108 (7): 795–801.

Harsanyi, J. C. 1955. "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility." *Journal of Political Economy* 63 (4): 309–21.

Heffetz, Ori, and Robert H. Frank. 2011. "Preferences for Status: Evidence and Economic Implications." In *Handbook of Social Economics*, edited by Jess Benhabib, Alberto Bisin, Matthew Jackson, 69–91. Amsterdam: Elsevier, North-Holland.

Huckman, Robert S. 2006. "Hospital Integration and Vertical Consolidation: An Analysis of Acquisitions in New York State." *Journal of Health Economics* 25 (1): 58–80.

Huckman, Robert S., and Gary P. Pisano. 2006. "The Firm Specificity of Individual Performance: Evidence from Cardiac Surgery." *Management Science* 52 (4): 473–88.

Johnson, Erin M. 2010. "Ability, Learning and the Career Path of Cardiac Specialists." Unpublished.

Kolstad, Jonathan. 2012. "The Effects of Public and Private Quality Information on Consumer Choice in Health Care Markets." In *The Value of Information,* edited by Ramanan Laxminarayan and Molly K. Macauley, 143–64. New York: Springer.

Kolstad, Jonathan T. 2013. "Information and Quality When Motivation is Intrinsic: Evidence from Surgeon Report Cards: Dataset." *American Economic Review*. http://dx.doi.org/10.1257/aer.103.7.2875.

Kolstad, Jonathan T., and Michael E. Chernew. 2009. "Quality and Consumer Decision Making in the Market for Health Insurance and Health Care Services." *Medical Care Research and Review* 66 (1): 28S–52S.

McGuire, Thomas G. 2000. "Physician Agency." In *Handbook of Health Economics*. Vol. 1A, edited by Anthony J. Culyer and Joseph P. Newhouse, 461–536. Amsterdam: Elsevier, North-Holland.

McGuire, Thomas G., and Mark V. Pauly. 1991. "Physician Response to Fee Changes with Multiple Payers." *Journal of Health Economics* 10 (4): 385–410.

Miller, Nolan. 2006. "Report Cards, Incentives and Quality Competition in Health Care." Unpublished.

Mukamel, Dana B., and Alvin I. Mushlin. 1998. "Quality of Care Information Makes a Difference: An Analysis of Market Share and Price Change After Publication of the New York State Cardiac Surgery Mortality Reports." *Medical Care* 36 (7): 945–54.

Nevo, Aviv. 2000. "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand." *Journal of Economics and Management Strategy* 9 (4): 513–48.

Newhouse, Joseph P. 1970. "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital." *American Economic Review* 60 (1): 64–74.

Pennsylvania Health Care Cost Containment Council (**PHC4**)**.** 1998. *Coronary Artery Bypass Graft Surgery—1994–95 Data: Research Methods and Results*. Harrisburg: PHC4.

Pennsylvania Health Care Cost Containment Council (**PHC4**)**.** 2002. *Coronary Artery Bypass Graft Surgery—2000 Data: Research Methods and Results*. Harrisburg: PHC4.

Pennsylvania Health Care Cost Containment Council (**PHC4**)**.** 2004. *Coronary Artery Bypass Graft Surgery—2002 Data: Research Methods and Results*. Harrisburg: PHC4.

Pennsylvania Health Care Cost Containment Council (**PHC4**)**.** 2005. *Coronary Artery Bypass Graft Surgery—2003 Data: Research Methods and Results*. Harrisburg: PHC4.

Peterson, Eric D., Elizabeth R. DeLong, James G. Jollis, Lawrence H. Muhlbaier, and Daniel B. Mark. 1998. "The Effects of New York's Bypass Surgery Provider Profiling on Access to Care and Patient Outcomes in the Elderly." *Journal of the American College of Cardiology* 32 (4): 993–9.

Ramanarayanan, Subramaniam. 2008. "Does Practice Make Perfect: An Empirical Analysis of Learning-by-Doing in Cardiac Surgery." Unpublished.

Salop, Steven, and Joseph E. Stiglitz. 1977. "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion." *Review of Economic Studies* 44 (3): 493–510.

Sauermann, Henry, and Wesley M. Cohen. 2008. "What Makes Them Tick? Employee Motives and Firm Innovation." National Bureau of Economic Research Working Paper 14443.

Segal, Uzi, and Joel Sobel. 2007. "Tit for Tat: Foundations of Preferences for Reciprocity in Strategic Settings." *Journal of Economic Theory* 136 (1): 197–216.

Schneider, Eric C., and Arnold M. Epstein. 1998. "Use of Public Performance Reports: A Survey of Patients Undergoing Cardiac Surgery." *Journal of the American Medical Association* 279 (20): 1638–42.

Sloan, Frank A. 2000. "Not-for-Profit Ownership and Hospital Behavior." In *Handbook of Health Economics*. Vol. 1B, edited by Anthony J. Culyer and Joseph P. Newhouse, 1141–74. Amsterdam: Elsevier Science, North-Holland.

Steinbrook, Robert. 2006. "Public Report Cards—Cardiac Surgery and Beyond." *New England Journal of Medicine* 355 (18): 1847–49.

Train, Kenneth E. 2003. *Discrete Choice Methods with Simulation*. New York: Cambridge University Press.