Ability, Learning and the Career Path of Cardiac Specialists

Erin M. Johnson*

April 2011

Abstract

Prior research suggests physicians respond to incentives in quality programs, yet little is known about incentives that arise in physician labor markets absent intervention. This paper investigates the role of market learning by referring doctors in promoting specialist quality. I motivate the empirical work with a model of specialist careers under learning. I then compare careers of higher and lower quality specialists, using the universe of Medicare claims from cardiac specialists from 1996-2005 and a new empirical quality measure that is robust to patient sorting. I find lower quality specialists are more likely to stop practicing and change markets over time.

*Massachusetts Institute of Technology, Sloan School of Management, 77 Massachusetts Avenue, E62-523, Cambridge, MA 02139 (e-mail: erinmj@mit.edu). I am especially grateful to the members of my dissertation committee: David Card, Patrick Kline, and Enrico Moretti. I would also like to thank the following individuals for helpful comments: Alan Auerbach, David Cutler, Joe Doyle, Bob Gibbons, Ashley Langer, Ronald Lee, Marit Rehavi, and Erin Syron. I am also grateful to Jean Roth and Mohan Ramanujan for their assistance in obtaining and managing the data. Drs. Kent Potts and James Hartsuck provided helpful discussions on cardiac careers and procedures. Seminar participants at BU School of Management, BU/Harvard/MIT Health Seminar, Carnegie Mellon, Chicago Harris, MIT Sloan, MIT, Stanford SIEPR, UC Berkeley, Wharton and the NBER Summer Institute Healthcare meeting provided valuable comments. Financial support for this work was provided by the Burch Center for Tax Policy and Public Finance and the Center for Labor Economics at UC Berkeley.

1 Introduction

Patients' uncertainty regarding the quality of health care providers has long been recognized as an information problem with potentially serious consequences (Arrow, 1963). In particular, consumer uncertainty is thought to lead to under-provision of quality care in equilibrium (Dranove and Satterthwaite, 1992). This issue has garnered a lot of interest in recent years, as empirical studies have documented differences in quality measures across providers domestically[1] and cross-country studies have shown the U.S. to lag other developed nations on several measures of quality.[2]

There are a number of programs aimed at enhancing incentives for quality care in the U.S., and there is a large empirical literature studying these programs. The physician "report card" literature measures the impact of the publication of physician quality measures (typically patient mortality rates in complex procedures).[3] Other recent studies address incentives arising from pay-for-performance programs.[4] However, there is relatively little work on one of the most important potential incentive mechanisms in professional service relationships: the role of market learning about the relative quality, or ability, of individual professionals.[5] This paper fits in this space and seeks to determine whether referring doctors, in their capacity as agents for uninformed patients, learn about specialist quality over time and exert quality incentives through their actions.

The paper draws on the large literature on employer learning to illuminate the incentives arising in the relationship between referring doctors and specialists. I then measure the strength of these incentives using the universe of Medicare claims filed by cardiac specialists in the U.S. from 1996-2005.

The model builds on the public learning model of Jovanovic (1979) and Farber and Gibbons (1996). In the model referring doctors observe patient outcomes and form expectations of specialists' ability over time. Referring doctors then use this information to allocate patients to specialists to improve patient survival. I also augment the basic

[1] For example, Hannan et al. (1990), O'Connor et al. (1991), Williams, Nash and Goldfarb (1991), and McClellan and Staiger (1999).

[2] WHO World Health Statistics 2010, Nolte and McKee (2008).

[3] See Hannan et al. (1994), Green and Wintfeld (1995), Schneider and Epstein (1996), Peterson et al. (1998), Cutler, Huckman and Landrum (2004), and Dranove et al. (2003).

[4] For example, Rosenthal, Frank and Epstein (2005), Campbell et al (2007) and Mullen, Frank and Rosenthal (2009), and Rosenthal and Frank (2006) provides a review.

[5] An exception is Fourrier and McInnes (2002), who find that surgeons with more malpractice claims receive fewer fee-for-service (FFS) patients over time, but not fewer HMO patients. They attribute this result to referring doctors being able to respond to the quality information in the malpractice signal for FFS but not HMO patients.

learning framework to allow for capacity constraints, which are an important institutional feature of cardiac specialty markets.

The model has three main predictions. First, under learning by referring doctors, lower quality specialists will be more likely to drop out of practice over time. Second, specialists will sort across markets based on quality - lower quality specialists can potentially avoid reductions in referrals by moving to markets that are capacity constrained. Third, referring doctors will seek to allocate more patients to higher quality doctors as quality becomes known.

The empirical analysis focuses on two types of cardiac specialists: interventional cardiologists (ICs), who perform percutaneous coronary interventions (PCI), such as angioplasty, and cardiothoracic (CT) surgeons, who perform coronary artery bypass graft surgery (CABG). For both of these specialties a large fraction of patients are over 65 and observable in Medicare claims data. Further, both practices are procedure-based, and technical skill is an important determinant of patient outcomes in PCI and CABG. The fact that I observe both specialties also allows me to exploit the proliferation of the bare metal stent during the sample period, which increased demand for PCI and decreased demand for CABG (Cutler and Huckman (2003)), to determine how incentives differ in growing and declining markets.

The labor market for medical specialists is unique, and ex ante the empirical magnitude of any learning effect is unclear. On the one hand, referring doctors have access to much of the quality information published in physician report cards – they observe their own patients' outcomes following specialist care, and they can also gather information from colleagues. On the other hand, licensing requirements and the long and arduous nature of specialty training may suggest a more limited role for learning than in labor markets for high school or college graduates. Further, the rarity of patient complications and the small numbers of patient procedures performed by some doctors may make it difficult for referring doctors to effectively determine specialist quality.

To empirically determine the importance of learning, one needs a measure of doctor quality. Unlike in the employer learning case where employee productivity is generally unobserved, I observe patient mortality outcomes following procedures with specific doctors. I develop a methodology for measuring quality using patient outcomes data that improves upon the hierarchical models at the forefront of the quality measurement literature.[6] The

[6]For examples of the hierarchical approach, see Localio et al. (1997), Burgess et al. (2000), Thomas, Longford and Rolph (1994), and Normand et al. (1997).

innovation is in modeling patient mortality using the correlated random effects design of Mundlak (1978) and Chamberlain (1982) to control for non-random sorting of patients to specialists on observables. I then construct quality measures as empirical Bayes estimates of the random intercepts in the model.

Using these measures, I first test the prediction that lower quality specialists should be more likely to stop performing PCI or CABG or to drop out of practice. Intuitively, as quality becomes known over time, lower quality specialists receive fewer referrals relative to higher quality specialists, and the outside option becomes relatively more attractive to them. I consider two outside options: one in which specialists stop performing interventional procedures and adopt a more clinical practice, and one in which specialists stop billing Medicare altogether. Logistic regressions of dropout indicators on quality measures reveal that lower quality ICs are in fact more likely to stop doing PCI. The effect is significant both statistically and economically: a one standard deviation decrease in doctor quality increases the likelihood of dropping out of performing PCI by one percentage point (a 10% effect). As predicted by the model, the effect is stronger in the declining market for CTs: a one standard deviation decrease in quality increases the likelihood of dropping out of CABG by two percentage points, and low quality CT surgeons are also more likely to stop billing Medicare altogether.

Next I turn to the prediction on specialist sorting. Because the impacts of learning on referral volumes are muted in capacity constrained markets, lower quality specialists may be able to increase their referral volumes through moving to markets with a relatively lower supply of specialists. For ICs I find a one standard deviation decrease in quality increases the likelihood of moving by two percentage points; for CT surgeons the effect is again slightly stronger, three percentage points. I also show that the last hospital referral region (HRR) of movers has lower capacity than their first HRR.[7]

Finally, I examine referral volumes of specialists over the career. For specialists who do not move or drop out of the labor market, I do not find convincing evidence that the growth rate of claims or of PCI/CABG claims diverges by quality. However, in most specifications I cannot rule out differences in volumes between 5 and 9 percent after ten years' experience for higher versus lower quality specialists (by one standard deviation). Plots of claims and PCI/CABG claims for dropouts do uncover sharp reductions several years prior to the dropout date compared with non-dropouts. Also consistent with the

[7] A hospital referral region is defined around hospitals performing both cardiovascular surgery and neurosurgery. It is the area in which the majority of patients are referred to the hospital. See Wennberg et al. (2008) for more detail.

model, movers experience reductions in claims prior to the moving date and strong claims recovery shortly thereafter.

The empirical evidence is consistent with modest learning by referring doctors, and it suggests specialists face a quality incentive even absent report card programs and pay-for-performance programs. Results also suggest learning impacts patient welfare: low quality specialists stopping doing PCI or stopping practicing altogether reduces patient mortality assuming adequate physician supply, and moving behavior may have distributional consequences.

The paper proceeds in seven sections. Section II reviews the empirical literature on provider quality incentives. Section III provides details on the practice setting and presents the model. Section IV describes the data and Section V the construction of doctor quality measures. Section VI presents the empirical evidence, and Section VII concludes.

2 Previous Literature

This section summarizes the literature on provider incentives, focusing on quality incentives.[8] The majority of research in this area has focused on two types of quality initiatives: pay-for-performance programs, which tie physician reimbursement directly to measurable outcomes, and report card programs, which make provider quality measures publicly available (typically risk-adjusted patient mortality rates). The evidence on pay-for-performance programs is mixed. Campbell et al. (2007) finds evidence that a government initiative in the U.K. significantly improved treatment of asthma and diabetes, but studies of smaller financial incentives have found little evidence of quality improvements (Rosenthal and Frank (2006), Mullen, Frank and Rosenthal (2009)).

The literature on report cards is of particular relevance to the current study, since the primary goal of these programs is to promote selection of providers based on quality.[9] The majority of the report card literature has evaluated the CABG reporting programs in New York State and Pennsylvania. Early studies in New York found a large mortality reduction

[8] There is also a large literature on physicians' response to financial incentives. For example, Le Grand (1999) and Gaynor, Rebitzer and Taylor (2001) study physicians employed by Health Maintenance Organizations (the former in the UK, the latter in the US) and find evidence that physicians adjust behavior in response to incentive pay. Hemenway et al. (1990) and Barro and Beaulieu (2003) consider physicians employed by hospitals and find increases in productivity in response to changes from salary to profit-sharing pay. Gaynor and Pauly (1990) finds similar incentive responses in a study of physicians in a group practice setting.

[9] A secondary goal of report cards is to promote internal quality improvement through "intrinsic" incentives (Kolstad, 2009)).

and an increase in the risk factors of patients undergoing CABG after the program.[10] Research that has followed has attempted to determine the causality of and the mechanisms underlying this effect.[11] Green and Wintfeld (1995) finds evidence that coding of patient risk factors increased after implementation of the program. Dranove et al. (2003) finds that healthier patients receive CABG after the program compared with before, suggesting providers may engage in selecting patients based on risk factors out of concern for their report card. Cutler, Huckman and Landrum (2004) finds that hospitals with poor performance ratings lost healthy patients and experienced performance improvement relative to other hospitals, consistent with a demand incentive.[12]

While these studies have focused at the state or hospital level, there is also a literature studying impacts of physician report cards on physician practices. Mukamel and Mishlin (1998) and Mukamel et al. (2004) find that surgeons with good report cards increased market share following the program in New York. Kolstad (2009) finds slightly increased demand for highly rated surgeons after report cards are released in Pennsylvania. Wang et al. (2010) also study the Pennsylvania case and find reductions in patient volumes for poorly rated surgeons

There have also been several surveys of doctors aiming to uncover the impact of report cards. Schneider and Epstein (1996) reports that 87% of referring doctors in Pennsylvania say report cards had "minimal or no influence" on their referral recommendations. However, it is not clear from the survey the extent to which the market disciplines quality absent report cards.

3 Practice Setting & Model

3.1 Practice Setting

The empirical work focuses on two physician specialties: interventional cardiology and cardiac surgery. Interventional cardiologists are cardiologists who sub-specialize in performing interventions to open arteries in the heart that have been narrowed by coronary

[10] The most widely cited of these is Hannan et al. (1994)

[11] Epstein (2006) provides a summary of the empirical literature on hospital and physician report cards. Kolstad and Chernew (Forthcoming) review the literature on provider and health plan report cards.

[12] There are also literatures report cards in other contexts. Beaulieu (2002), Wedig and Tai-Seale (2002), Chernew, Gowrisankaran and Scanlon (2001) and Dafny and Dranove (2008) study health plan report cards. Bundorf et al. (2009) and Pope (2008) study clinic report cards and hospital rankings, respectively.

artery disease.[13] Cardiothoracic surgeons are surgical specialists who treat conditions of the heart and cardiovascular system.

In the empirical work I evaluate ICs and CT surgeons' skill at performing the primary procedures of each of their specialties. PCI are the primary procedures done by ICs. The two most common of these procedures are angioplasty and angioplasty with stent placement. In angioplasty a balloon-tipped catheter is threaded into the heart, and the balloon is inflated to clear arterial blockages. Since the introduction of stents in the U.S. in 1994 percutaneous interventions increasingly involve balloon inflation followed by the placement of a metal scaffold-like structure, called a stent, to keep vessels open.

CABG, also known as open heart surgery, is the primary procedure done by cardiac surgeons. In CABG, the patient's chest is opened, the heart is stopped, and blood is routed to a heart-lung machine for oxygenation. Arteries or veins are harvested from elsewhere in the patient's body and grafted to the heart to restore blood flow around diseased vessels.[14] Because CT surgeons do not typically perform percutaneous interventions, and ICs open arteries using only non-surgical, percutaneous techniques,[15] these two specialties are handled separately in the empirical work.

These specialties provide a number of advantages in studying referral relationships. For both, a large fraction of patients are over 65 and observable in Medicare claims data.[16] Further, both practices are procedure-based, and technical skill is an important determinant of patient outcomes in both PCI and CABG. In PCI, for example, it takes a high level of technical skill to thread the catheter into the heart and developed judgment to determine the amount of pressure used to inflate the balloon - too much pressure can

[13]There are three major sub-specialties of cardiology. Non-invasive cardiologists diagnose and provide medical management of patient conditions. For example, non-invasive cardiologists perform stress tests, EKGs, echocardiograms and see patients in a clinical setting. Invasive cardiologists do everything non-invasive cardiologists do plus diagnostic angiography. In diagnostic angiography, also known as heart catheterization, a catheter is threaded into the heart and used to inject contrast agent allowing for X-ray photography of the heart vessels and assessment of heart function. Interventional cardiologists do everything invasive cardiologists do plus perform interventions, such as angioplasty, to open arteries in the heart that have been narrowed by plaque.

[14]CABG is generally indicated instead of PCI for patients with more severe coronary artery disease. However, indications for PCI have been expanding to more risky patient groups with more severe disease over time (Bohmer, Christensen and Kenagy (2000)).

[15]The training of ICs and CT surgeons diverge after medical school. ICs do a 3 year internal medicine residency followed by a 2-3 year cardiology fellowship and 1-2 additional years of training in PCI as fellows in interventional cardiology. CT surgeons complete a 5 year residency program in general surgery and then a 2-3 year fellowship in cardiac surgery.

[16]Analysis of cardiac report card data from the state of New Jersey from 2001-2005 shows 52 percent of patients undergoing angioplasty and 60 percent of patients undergoing CABG are over 65.

rupture the vessel, but not enough pressure can result in re-narrowing of the artery. A less skilled technician might also have more patient complications because the procedure takes longer or requires the injection of larger amounts of contrast agent. In CABG, a less skilled technician might have increased risk of bleeding or require the patient to be on a heart-lung machine for longer.

These specialties have also been at the center of the quality measurement or "report card" movement. While CT surgery was the original focus, several states and agencies have recently also begun reporting quality information for ICs. The use of data on the two specialties also allows me to exploit the proliferation of the bare metal stent during the sample period. This innovation increased demand for PCI and decreased demand for CABG (Cutler and Huckman (2003)), allowing me to study how incentives differ in growing and declining markets.

In the U.S. the majority of cardiac specialists (58% of cardiologists, 78% of CT surgeons) are in private practice[17] and therefore dependent on flows of patients into their practice to cover office overhead and earn a profit. The majority of patients in an IC practice are referred from a primary care physician (PCP) after the PCP has determined the patient needs cardiac care, and CT surgeons receive most referrals from cardiologists who have determined patients need surgery. While some patients in both specialties are self-referred, most patients rely exclusively on referrals in specialist choice (Center for Studying Health System Change, 2007).

3.2 Theoretical Framework

In this section I present a simple model to illustrate the effects of learning by referring doctors on specialist careers. The model builds on the public learning models of Jovanovic (1979) and Farber and Gibbons (1996).[18] To begin I assume there are no constraints on the number of patients a specialists can see. This provides the simplest treatment and illuminates the basic implications of learning for specialist careers. In the next section I add capacity constraints.

3.2.1 Modeling Referrals Under Learning

To begin let $i = 1, ..., N$ denote referring doctors (RDs) and $j = 1, ..., M$ denote specialists. Assume specialists have different ability levels, η_j, and assume that individual specialist

[17] Bruno & Ridgway Research Associates, Inc. (2002); Shemin et al. (2002)

[18] The treatment is also closely related to the model with no social learning of Moretti (forthcoming).

skill is not observed by RDs. RDs only have prior knowledge of the distribution of specialist skill:

$$\eta_j \sim N\left(X_j^{'}\beta, \frac{1}{H}\right) \tag{1}$$

where X_j is a vector of doctor characteristics observable to RDs at the beginning of their careers, for example the prestige of medical school attended, and H is the precision of the prior.

I also assume that RDs observe a signal of ability, y_{ijt}, in each period. This might be the patient survival rate for specialist j or other observed quality information. It is made up of two components: true specialist skill and a mean zero normally distributed error term, ϵ_{ijt}.

$$y_{ijt} = \eta_j + \epsilon_{ijt}, \qquad \epsilon_{ijt} \sim N\left(0, \frac{1}{h}\right), \text{ iid} \tag{2}$$

The error term is assumed to vary across RDs to allow different RDs to have different information sets.

At each time t, RDs form expectations of specialist ability based on the information available to them. At the beginning of a specialist's career, this expectation is simply the mean of the skill distribution, $X_j^{'}\beta$. But in the second period RD i also observes the specialist's first period signal, y_{ij1}, and incorporates this into his expectation:

$$E_{i2}[\eta_j | X_j^{'}\beta, y_{ij1}] = \frac{H}{H+h}X_j^{'}\beta + \frac{h}{H+h}y_{ij1} = w_2 X_j^{'}\beta + (1-w_2)y_{ij1} \tag{3}$$

where $w_2 = \frac{H}{H+h}$. The expectation is a weighted average of the prior and the signal, with the weight on the signal increasing in its precision.

Iterating on the learning model gives RD i's expectation in time t given the prior and the signals up until time t, $y_{ij1}, ..., y_{ij(t-1)}$:

$$\begin{aligned} E_{it}[\eta_j | X_j^{'}\beta, y_{ij1}, ..., y_{ij(t-1)}] &= \frac{H}{H+(t-1)h}X_j^{'}\beta + \frac{h}{H+(t-1)h}\sum_{s=1}^{s=(t-1)} y_{ijs} \\ &= w_t X_j^{'}\beta + (1-w_t)\eta_j + \frac{h}{H+(t-1)h}\sum_{s=1}^{s=(t-1)} \epsilon_{ijs} \end{aligned} \tag{4}$$

where $w_t = \frac{H}{H+(t-1)h}$. From this equation it is evident that, as t approaches infinity,

the prior belief becomes less important and true quality becomes more important in the expectation. To see this note that the weight on the prior, w_t, is decreasing in t.

The RD then uses this information to allocate patients to specialists. I model the RD's action as follows: in each period RD i decides to refer his patients to specialist j if j's expected patient survival is above some threshold, q. The RD then randomizes his patients among those specialists he has decided to refer to. In this decision rule, RDs care that doctor quality is above a threshold, perhaps the local standard of care; they do not care about doctor quality above q. While I cannot observe the decision rule generating referral flows, anecdotal evidence suggests this rule is consistent with RD behavior. RDs generally maintain relationships with several specialists and allocate patients fairly evenly among those specialists. I have also considered an alternative rule in which RDs maximize expected patient survival by referring exclusively to the specialist whose expected ability is highest in his estimation. This rule makes similar predictions for the career dynamics I consider.[19]

Under my rule, RD i refers to j in period t if:

$$E_{it}[\eta_j | X_j^{'}\beta, y_{ij}, ...y_{ij(t-1)}] > q \tag{5}$$

From this expression I derive the probability that specialist j is above the threshold for RD i:

$$\begin{aligned} P_{jt} &= Pr\left\{E_{it}[\eta_j | X_j^{'}\beta, y_{ij1}, ..., y_{ij(t-1)}] > q\right\} \\ &= Pr\left\{w_t X_j^{'}\beta + (1-w_t)\eta_j + \frac{h}{H+(t-1)h}\sum_{s=1}^{s=(t-1)}\epsilon_{ijs} > q\right\} \\ &= Pr\left\{w_t X_j^{'}\beta + (1-w_t)\eta_j - q > -\frac{h}{H+(t-1)h}\sum_{s=1}^{s=(t-1)}\epsilon_{ijs}\right\} \\ &= \Phi\left(\frac{w_t X_j^{'}\beta + (1-w_t)\eta_j - q}{\sigma_t}\right) \end{aligned} \tag{6}$$

where Φ is the normal cumulative distribution function and $\sigma_t = \sqrt{(\frac{h}{H+(t-1)h})^2 \frac{1}{h}(t-1)}$. Note that each RD i has the same probability of sending patients to specialist j before

[19]This model is not tractable when there is a continuum of specialist types. With two types, the model predicts divergence by type in dropout behavior, moving behavior and claims over the career.

draws of ϵ_{ijt} are realized (P_{jt} does not vary with i, because RDs' information differs only in their realizations of ϵ_{ijt}).

Next consider the number of patient referrals specialist j expects to receive. Assuming, without loss of generality, that each RD has one patient to refer, specialist j expects referrals from RD i equal to the probability RD i refers to him times one over the expected number of specialists RD i refers to. The second term is equal to the expected number of patients RD i refers to each specialist above the threshold, and it arises from the assumption that he is randomizing patients among this group. Summing this expression across RDs gives j's expected referrals:

$$\sum_{i=1}^{N} \frac{P_{jt}}{\sum_{j=1}^{M} P_{jt}} \tag{7}$$

Note that the expected number of specialists above the threshold (the denominator) does not vary with j. For purposes of comparing doctors of different ability, then, the denominator is simply a constant:[20]

$$\sum_{i=1}^{N} \frac{P_{jt}}{\sum_{j=1}^{M} P_{jt}} = kP_{jt} \tag{8}$$

where $k = \frac{N}{\sum_{j=1}^{M} P_{jt}}$. The constant simply acts as a scale factor to ensure the number of patients referred does not exceed the number of patients needing care in expectation.[21]

The first thing to note about this model is that the expected number of referrals to each specialist, kP_{jt}, is increasing in specialist quality, η_j, all else equal. This is evident from equation 6 – if RDs are learning about specialist quality then the term in η_j is positively weighted in the numerator; absent learning (as $h \to \infty$), this term has a zero weight.

The model also predicts that P_{jt} and therefore referrals will diverge over time for higher and lower quality doctors. Consider first the case where $x_j^{'}\beta = q$. This is the case where all doctors are believed to perform at the local standard of care at the beginning

[20]Note here I am also assuming that the expected number of specialists above the threshold is constant over time. This is true in the first case considered below, where $X_j^{'}\beta = q$ for any exogenous q. For this to be true in the more general case we require that q is equal to the average expected ability in each time period (and in this case $q = q_t$). I discuss the comparative statics both under this assumption and in the case where q is exogenous below.

[21]To see this note that the expected number of patient seen, $\sum_{j=1}^{M}\sum_{i=1}^{N} \frac{P_{jt}}{\sum_{j=1}^{M} P_{jt}} = \sum_{i=1}^{N}\sum_{j=1}^{M} \frac{P_{jt}}{\sum_{j=1}^{M} P_{jt}} = N$, the number of patients needing to be seen.

of their careers. This case may provide a good approximation of learning in specialist markets, because of long training periods and licensing requirements. Moreover, specialists are rarely exposed to the RDs they will interact with in private practice during residencies and fellowships. Under this assumption, referrals increase over time for doctors whose true ability is above the prior (high ability doctors) and decrease over time for doctors whose true ability is below the prior (low ability doctors):[22]

$$\frac{dP_{jt}}{dt} > 0 \text{ for } \eta_j > X_j^{'}\beta = q,$$
$$\frac{dP_{jt}}{dt} < 0 \text{ for } \eta_j < X_j^{'}\beta = q$$

For the more general case where priors differ across doctors the condition is more complicated, but the intuition is similar. There is divergence in referrals so long as η_j is high (low) enough relative to the prior for $\eta_j > \overline{\eta}$ ($\eta_j < \overline{\eta}$). Essentially, learning must be positive enough for the best doctors and negative enough for the worst doctors.[23] Alternatively, one can think of the condition in terms of movements above and below the threshold as true ability becomes known. Divergence requires only that more specialists in the top half of the quality distribution move above the threshold than below and vice versa for specialists in the bottom half of the quality distribution. This is essentially a requirement that learning is productive.

Next consider career decisions of the specialist. It is logical to assume that specialists with an outside option will drop out of the labor market if their expected referrals drop below some threshold (if W is the threshold the condition is: $kP_{jt} < W$).[24] Because lower

[22]The expression for the change in referrals over time in this case is as follows:

$$\frac{dP_{jt}}{dt} = \phi\left(\frac{w_t X_j^{'}\beta + (1-w_t)\eta_j - q}{\sigma_t}\right)\frac{H(t-1)^{1/2}}{h^{1/2}}(\eta_j - X_j^{'}\beta)$$

This expression is positive for $\eta_j > X_j^{'}\beta$ and negative for $\eta_j < X_j^{'}\beta$.

[23]The condition is as follows for the case where q equals the average expected ability in each period:

$$\eta_j > \frac{w_t}{(1-w_t)}(X_j^{'}\beta - q) + q \text{ for } \eta_j > \overline{\eta}$$
$$\eta_j < \frac{w_t}{(1-w_t)}(X_j^{'}\beta - q) + q \text{ for } \eta_j < \overline{\eta}$$

In the case where q is exogenous, the condition for divergence in referrals by quality is more complicated: the effect of learning through the impact on expected ability must be stronger than the effect arising from the change in expected referrals per doctor above the threshold.

[24]Through rescaling of the dropout point and given information on income per referral, this can be rewritten as a condition on the net present value of income from referral flows.

quality specialists receive fewer referrals over time, this condition implies that lower quality specialists should be more likely to drop out of the labor market over time.[25]

3.2.2 Capacity Constraints

Up to this point, I have not put any constraints on the number of patients any single specialist can see. However, it is likely that specialists face capacity constraints. I impose capacity constraints by requiring each specialist's expected referrals to be less than or equal to $(1+\gamma)\frac{N}{M}$ for $\gamma > 0$. This is equivalent to assuming there is $\gamma * 100$ percent excess capacity in the market.[26]

$$\sum_{i=1}^{N} \frac{P_{jt}}{\sum_{j=1}^{M} P_{jt}} \leq (1+\gamma)\frac{N}{M} \quad \forall j \tag{9}$$

As before, referrals for specialist j are determined from the definition of P_{jt} (equation 6), the probability doctor j is above the threshold. But now, in addition to this condition, there is a system of j inequalities which must be satisfied. Specialists whose expected referrals under equation 6 exceed the capacity constraint must have P_{jt} adjusted downward until the constraint just binds. Intuitively, some referring doctors remove a full capacity specialist from their referral list, lowering P_{jt} for that specialist. This in turn reduces the expected number of specialists above the threshold, $\sum_{j=1}^{M} P_{jt}$, which increases expected referrals for doctors below the constraint. When one doctor is dropped off a RD's list, the RD then randomizes patients among a smaller group of remaining specialists.

The end result is that referrals are equal for doctors at the constraint; for doctors below the constraint referrals are increased by a multiplicative factor that is constant across j. Thus, with capacity constraints and some excess supply, quality will not matter for referrals for doctors above some expected quality level, but it will continue to matter as in the original model with no capacity constraints for other specialists. This implies that we should expect to see stronger effects of quality on careers in markets with more excess capacity, and it implies that specialists can avoid reputation effects by moving to markets

[25]This is true so long as the outside option W is not substantially lower for lower quality doctors, which is likely to hold in the setting under study - low technical skill at performing angioplasty does not imply low skill at medical management and certainly not in employment outside of medicine.

[26]It is unlikely there is no excess capacity in the labor market for specialists. For example, we might think specialists can adjust practice styles or hours worked to increase capacity in response to increased referral flows. Alternatively, we might think patients will lower demand for procedures if only lower quality specialists are available, thereby increasing capacity.

that are more capacity constrained. Considering again the career decisions of specialists, we have a third prediction: lower quality specialists will be more willing to incur costs of moving to avoid reputation effects and therefore will be more likely to move. Further, their moves should be to areas that are more capacity constrained or areas with lower excess capacity.[27]

4 Empirical Approach

The primary data used in this analysis is an extract from the confidential Medicare Part B claims file for years 1996-2005. The extract is a 100 percent sample of claims submitted by ICs and CT surgeons that was created specifically for this project.[28] The 100 percent sample ensures that I observe enough patients for each doctor both to measure quality and to measure career paths of specialists.

The data includes physician identifiers, beneficiary identifiers, CPT procedure codes, up to 8 diagnosis codes (icd9), charges submitted to Medicare, and Medicare payment information, including the amount allowed and the amount paid by Medicare. I link this file to the Medicare Physician Identification and Eligibility Registry (MPIER) file to obtain information on physicians, including self-reported specialty and medical school name and graduation date. I also link to the Medicare Denominator file to get patient demographic information including date of death, date of birth, sex, ZIP code, and Medicare eligibility reason codes.

The empirical approach involves several steps. I first create doctor quality measures using patient mortality outcomes and risk factors. I do this by running patient-level regressions using the first four sample years of data (1996-1999), as described in the next section. I then summarize claims by specialist each year, creating measures of dropout behavior, moving behavior and claims volumes at the doctor level. Finally, I run doctor-level regressions to test the predictions of the model. Specifically, I measure the extent to which dropout behavior, moving behavior and claims volume diverge over time by quality.[29]

[27] It is also interesting to note that specialists dropping out of the labor market and moving has an effect on capacity - one specialist dropping out adjusts total market capacity downward by a factor of $\frac{M-1}{M}$. Thus, dropouts may have a dampening effect on the relationship between referrals and quality similar to the effect of shrinking excess capacity discussed above.

[28] The appendix contains detailed information on the construction of this extract.

[29] It is also important to note, in tying the empirical work to the theory, that I am implicitly assuming in the research design that the quality measure contains new information as of the beginning of the sample period. If the quality information contained in the quality measure is already known at this point, then we should not expect divergence in career measures by this measure even if learning is taking place. For

I limit the analysis to ICs and CT surgeons who have performed at least 30 PCI or CABG surgeries by 1999 to ensure quality measures contain real quality information.[30] This results in a sample of 4,417 ICs and 3,011 CT surgeons. I drop 2 ICs and 1 CT surgeon for whom information was not available in the MPIER file, and I drop 178 ICs and 172 CT surgeons for having zero or fewer years of experience as determined from their year of medical school graduation. Finally, I drop 9 ICs and 23 CT surgeons with no billing zipcode and 7 ICs and 1 CT surgeon with blank graduation year. The resulting sample is 4,228 ICs and 2,814 CT surgeons.

Table 1 provides summary information for this sample. The top panel summarizes physician characteristics. ICs are more likely to have attended a foreign medical school and less likely to have attended a prestigious medical school.[31] IC practices are also located in hospital referral regions (HRRs)[32] that are slightly smaller in terms of beneficiaries and Medicare spending per beneficiary.[33] Experience is imputed as current year minus the medical school graduation year minus expected time in specialty training (6 years for ICs, 7 years for CT surgeons). ICs are notably less experienced than CT surgeons, most likely because the IC sub-specialty is relatively new.

The second panel summarizes measures of claim volume. Claims is total annual claims, and PCI/CABG reports annual PCI claims for ICs and CABG claims for CT surgeons.[34] Non-emergency PCI/CABG claims are the subset of PCI/CABG claims that could be identified as non-emergency admissions by linking with the inpatient claim record.[35] Al-

young cohorts, this is likely an innocuous assumption. For older cohorts, it is likely that current patient survival rates contain new information if learning is slow and/or if quality is not fixed over the career. For this reason I also conduct analyses separately by age group.

[30]States which have implemented report card programs generally limit inclusion in the sample to physicians performing a minimum number of procedures in the sample period. New York and New Jersey require 100, Pennsylvania 30. Defining specialists in this way also reduces the likelihood that I have included non-cardiac specialists in the analysis who are erroneously appearing as performing doctors on claims. Analysis of the case mix and self-reported specialties of doctors performing under 30 procedures suggests the large majority are not cardiac specialists.

[31]Prestigious medical schools are identified as reprinted from US News and World Report in Hartz, Kuhn and Pulido (1999).

[32]A hospital referral region is defined around hospitals performing both cardiovascular surgery and neurosurgery. It is the area in which the majority of patients are referred to the hospital. See Wennberg et al. (2008) for more detail.

[33]The practice location is determined from the billing zip code on the physician claim.

[34]Claims with codes (92980-92982, 92984, 92995-92996) are considered PCI claims, and claims with codes (33510-33516, 33500, 33508, 33572, 33517-33522, 33530, 33533-33536) are considered CABG claims.

[35]Approximately 90 percent of Part B PCI/CABG claims were linked to the inpatient claim using the beneficiary ID, procedure codes, and dates of service. The linkage was undertaken for claims from years 1997-2005, as inpatient data was unavailable for 1996.

lowed charges is the amount of the physician's submitted charges allowed under the CMS contract (the amount paid by CMS and any amount that is the patient's responsibility), and patients is the number of unique Medicare beneficiaries seen by a physician each year. The average IC submits 5,452 claims each year for 1,155 unique patients, 63 of which are for PCI, and his allowed charges are just over $400,000. The average CT surgeon has a more procedural practice. He submits fewer claims (529) for fewer patients (216), but a higher number of CABG claims (76). Looking at case mix, just 4 percent of ICs claims are for PCI, but these account for 20 percent of charges - they are by far the most highly reimbursed claims. Procedures account for a significantly higher fraction of CT surgeon case mix, both by claims and charges.

Finally, the bottom panel includes measures of dropout and moving behavior. 3 percent of ICs and 16 percent of CT surgeons drop out of the sample over time, where dropouts are identified as having zero claims in each year after the year in which they dropout. 10 percent of ICs and 22 percent of CT surgeons stop doing PCI or CABG during the sample period. The higher dropout rates for CT surgeons likely result from declining CABG demand in the sample period and the fact that CT surgeons are closer to retirement on average. Moving is also quite common in the sample. 46 percent of ICs and 45 percent CT surgeons change billing zip codes during the sample period, and 16 percent of ICs and 24 percent of CT surgeons change HRRs.[36]

Figure 1 plots the number of Medicare claims for patients undergoing PCI, PCI with stent placement and CABG over the sample period. PCI and CABG are common procedures: by 2005 roughly 400,000 Medicare patients undergo PCI, and 200,000 patients undergo CABG. It is also clear from the figure that stents are steadily gaining market share from CABG over time; they increased their percentage of PCIs by over 200 percent in the ten year period, while the CABG market declined both in absolute terms and in market share. In the empirical work I exploit this difference to compare learning effects in growing and declining markets.

5 Constructing Quality Measures for Specialists

To measure specialist quality I model patient outcomes as a function of patient and procedure characteristics in a correlated random effects framework. In doing this I am drawing

[36]Specialists are determined to have changed zip codes if the billing zip code on the majority of their claims changes across years.

from the work of Mundlak (1978) and Chamberlain (1982) to improve upon the methods at the front of the quality measurement literature. The innovation is in controlling for sorting of patients to higher and lower quality doctors based on observable characteristics where the hierarchical models in the literature have not. Failing to control for patient sorting of this type results in bias: if higher risk patients sort to higher quality doctors, the model underestimates the effects of risk characteristics on mortality and penalizes high quality doctors for seeing high risk patients.

I evaluate ICs' quality based on the outcomes of their patients undergoing PCI, and I evaluate CT surgeons' quality based on the outcomes of their CABG patients. In both cases I use mortality in-hospital, defined as death within 2 days of PCI and within 7 days of CABG, as the outcome measure.[37] Note that these measures most likely capture technical skill - for example skills associated with successfully maneuvering the catheter into a heart vessel and repairing a blockage. I cannot observe other dimensions of specialist skill that are also likely important for patient satisfaction or survival, such as bedside manner or clinical diagnostic capabilities.

More formally, assume $y_{ijt} \in (1, 0)$ is the binary mortality outcome for patient j seeing doctor i in year t. Further assume the underlying equation for the model is:

$$
\begin{aligned}
y^*_{ijt} &= x_{ijt}\beta + \sigma_u u_i + v_{ijt} \\
u_i &\sim N(0,1) \\
y_{ijt} &= 1\left\{y^*_{ijt} > 0\right\} \\
E[v_{ijt}|u_i, x_{ijt}] &= \overline{x_i}\gamma
\end{aligned}
$$

where y^*_{ijt} is an unobservable latent variable determining patient mortality, and the logistic distribution is specified for v_{ijt}. Here x_{ijt} is a vector of patient and procedure characteristics, and $\overline{x_i}$ is the mean of these characteristics taken at the doctor level. Under the correlated random effects assumption, that $E[v_{ijt}|u_i, x_{ijt}] = \overline{x_i}\gamma$, inclusion of $\overline{x_i}$ in the fitted model controls for effects of patient sorting on $\hat{\beta}$ and ultimately on doctor quality measures.

First I obtain estimates of $\widehat{\beta}$, $\widehat{\gamma}$, and $\widehat{\sigma}_u$ via maximum likelihood estimation. I then

[37] For both types of procedures patient mortality may indicate a failed intervention. For example, technician skill affects the likelihood of vessel damage and bleeding, and lower skill technicians may also require more time to complete the procedure, putting the patient at increased risk. Longer procedure time is associated with increased stress on body systems, increased time on the heart-lung machine for CABG, and the use of more potentially harmful contrast agent in PCI.

apply empirical Bayes inference to estimate the random intercepts in the model. Specifically, I construct $\hat{\sigma}_u \hat{u}_i$ as the expectation of the posterior distribution of $\widehat{\sigma}_u u_i$ taking $\widehat{\beta}$, $\widehat{\gamma}$, and $\widehat{\sigma}_u$ as given.[38] I evaluate this integral using Gaussian quadrature methods.[39] The empirical Bayes approach applies a shrinkage factor to measures to account for estimation error (Morris, 1983).

This approach differs from that employed by most states in constructing report cards. In report cards patient outcomes are typically modeled in a logit framework with no fixed or random effects. The doctor's quality measure is then his actual patient mortality divided by his predicted mortality, where mortality is predicted from estimated logit coefficients and patient risk factors. The approach I take is closer to that in the literature on quality measurement, in which outcomes are generally measured in a random effects or a random coefficients framework.[40] Fixed effects methods offer an alternative approach and some authors have suggested fixed effects models for measuring hospital quality (e.g., Glance et al. (2006)). However, a problem arises when this approach is applied to measuring physician rather than hospital quality. It fails to provide a measure of quality for doctors with no variation in the dependent variable, which is common given low mortality rates.[41] By modeling patient outcomes in a correlated random effects framework, I provide a quality measure for each doctor that is robust to patient sorting on observables.[42]

[38]The posterior distribution can be expressed as follows:

$$\begin{aligned}
\hat{\sigma}_u \hat{u}_i &= E[\hat{\sigma}_u u_i | y_i, x_i; \hat{\sigma}_u, \hat{\beta}, \hat{\gamma}] \\
&= \int \hat{\sigma}_u u_i f(u_i | y_i, x_i; \hat{\sigma}_u, \hat{\beta}, \hat{\gamma}) du_i \\
&= \frac{\int \hat{\sigma}_u u_i f(y_i | u_i, x_i; \hat{\sigma}_u, \hat{\beta}, \hat{\gamma}) f(u_i) du_i}{\int f(y_i | u_i, x_i; \hat{\sigma}_u, \hat{\beta}, \hat{\gamma}) f(u_i) du_i} \\
&= \frac{\int \hat{\sigma}_u u_i \prod_{jt=1}^{JT} \left[\Lambda(x_{ijt}\hat{\beta} + \overline{x_i}\hat{\gamma} + \hat{\sigma}_u u_i)^{y_{ijt}} \left(1 - \Lambda(x_{ijt}\hat{\beta} + \overline{x_i}\hat{\gamma} + \hat{\sigma}_u u_i)\right)^{(1-y_{ijt})} \right] \phi(\frac{u_i}{\hat{\sigma}_u}) du_i}{\int \prod_{jt=1}^{JT} \left[\Lambda(x_{ijt}\hat{\beta} + \overline{x_i}\hat{\gamma} + \hat{\sigma}_u u_i)^{y_{ijt}} \left(1 - \Lambda(x_{ijt}\hat{\beta} + \overline{x_i}\hat{\gamma} + \hat{\sigma}_u u_i)\right)^{(1-y_{ijt})} \right] \phi(\frac{u_i}{\hat{\sigma}_u}) du_i}
\end{aligned}$$

where $\Lambda(x) = \left(\frac{e^x}{1+e^x}\right)$ and $y_i = (y_{ij1}, ..., y_{ijt})'$ and $x_i = (x_{ij1}, ..., x_{ijt})'$.

[39]Note that this method ignores the fact that $\hat{\sigma}_u$, $\hat{\beta}$, and $\hat{\gamma}$ are estimated in deriving the conditional distribution of y. Implementing corrections for this variation is unlikely to make much difference for quality measures given the small standard errors on these estimates.

[40]See, for example, Thomas et al. (1994), Normand et al. (1997), Localio et al. (1997), TenHave and Localio (1999), and Burgess et al. (2000).

[41]These models may also suffer from an incidental parameters problem (Neyman and Scott (1948)).

[42]Linear probability models are an alternative solution (see McClellan and Staiger (1999), but this is unattractive given low patient mortality rates.

I implement my measure using all non-denied and non-duplicate claims for PCI or CABG filed from 1996-1999. I use only the first four years of data to ensure that I can observe doctors' careers for a substantial period of time even after discarding the years used in constructing quality measures.[43] This results in a sample of 1,067,018 patients undergoing PCI and 1,019,770 undergoing CABG. The sample is summarized in Appendix Table A.1. 1.4 percent of PCI patients die in-hospital, and mortality is higher for CABG patients (3 percent).

Regressions include the full set of Charlson comorbidities as well as procedure information (for example, the number of vessels and the location of the blockage). They also include interactions between age, race and sex, and between these variables and indicators for acute myocardial infarction (AMI) and heart failure. As indicated in the model, I also included doctor-level means for each of the patient-level variables. Appendix TableA.2 displays coefficients from the maximum likelihood regressions. Finally, Appendix Table A.3 compares my quality measures with measures constructed without controlling for patient sorting and demonstrates that risk measures and ordinal ranks of doctors are significantly affected.

Table 2 summarizes the distribution of the resulting measures. First note that higher measures are associated with better quality (lower risk-adjusted patient mortality).[44] Next note that there is substantial variation in the quality measures - moving from the 10th to 90th percentile in IC skill is associated with a 26 percent change in the log odds of dying in the hospital (49 percent for CT surgeons). For ICs this is an effect about half as large as the effect of heart failure on hospital mortality.

6 Empirical Evidence

In this section I use doctor quality measures from the previous section to test the predictions of the model.[45] The first prediction I consider is that lower quality specialists should be more likely to drop out of the labor market over time. To test this I use two measures of

[43]This ensures doctor-level regressions are not affected if my quality measures include some component that is specific to the doctor's experience in 1996-1999.

[44]I have followed convention by predicting patient mortality in the logit regressions. However, because most readers, especially those familiar with the teacher quality literature in labor economics, are accustomed to higher quality measures being associated with higher quality, I use $-\hat{\sigma_u}\hat{u_i}$ as my doctor quality measure in the analysis that follows.

[45]Results are similar if I use quality measures based on death within three months of PCI/CABG or death within three months, but not in-hospital.

dropout status: stopping performing PCI or CABG and stopping billing Medicare. I use the first measure, because the best outside option for ICs who perform PCI poorly is likely a career within cardiology focused on non-interventional treatments (for example, diagnostic angiography or medical management of cardiovascular disease). Similarly, CTs who stop performing CABG may focus on other thoracic surgeries or general surgery. Dropping out of Medicare is the best available proxy for dropping out of the labor market entirely, though physicians who stop billing Medicare are likely only dropping out of private practice.[46]

I begin by analyzing the careers of ICs. To do this I estimate models of the form:

$$\begin{aligned} y_j^* &= \beta_0 + \beta_1 Q_j + \beta_3 X_j + v_j \\ y_j &= 1\left\{y_j^* > 0\right\} \end{aligned}$$

where y_j^* is an unobservable latent variable determining dropout behavior, Q_j is the quality measure for doctor j, X_j is a vector of doctor characteristics, and the logistic distribution is specified for v_j.[47] Table 3 contains results. For non-binary independent variables, the table displays effects of a one standard deviation change in the variable on the probability of dropping out, expressed in percentage points. The top panel shows results for dropping out of performing PCI, and the bottom panel shows results for stopping billing Medicare. Column 1 displays results from the basic specification with no fixed effects, and Column 2 includes cohort fixed effects to control for differences in dropout behavior across cohorts that may be correlated with quality.[48] Column 3 excludes doctors aged 65 years or older by 2005 to determine whether results are driven by differential retirement behavior by quality.

In the top panel, estimates are negative in all specifications and statistically significant at the 5 percent level in all but one. Point estimates suggest a one standard deviation decrease in doctor quality increases the likelihood of dropping out of PCI by approximately one percentage point. Compared with the overall fraction of ICs dropping out of PCI, this is a sizable effect (a 10 percent change in the probability of dropping out). The point estimates

[46] A limitation of the data is the inability to distinguish true dropout behavior from physicians stopping accepting Medicare patients. For the subsample of CT surgeons practicing in New Jersey, I was able to investigate this assumption by merging the CMS data with the confidential open heart surgery registry data from the New Jersey Department of Health and Senior Services. The ten New Jersey doctors who dropped out of the CMS data also stopped appearing in the cardiac registry data, which includes all patients undergoing CABG in the state, during the sample period.

[47] Results are substantively the same if doctor characteristics are omitted from the regressions.

[48] Small cohorts of ICs (less than 50 doctors) are grouped together to prevent dropping of observations in the fixed effects regressions. Specifically, cohorts with graduation years between 1967-1971, and 1990-1994 are grouped together.

for effects of quality on stopping doing PCI are only slightly smaller when specialists near retirement are excluded, approximately .8 of a percentage point. In the bottom panel, effects are negative in all specifications but smaller in magnitude, and an effect of minus one percentage point can be ruled out. Thus, while low quality ICs may choose to stop doing PCI, focusing instead on medical management or diagnostic angiography, they do not stop billing for services entirely.[49]

Next I analyze careers of CT surgeons. The model suggests that dropout behavior should be more pronounced in the declining labor market for CT surgeons. Coefficients on quality in dropout regressions for CTs, displayed in Columns 3-6 of Table 3, are in fact larger in magnitude. Coefficients on quality in all specifications are significant at the 1 percent level, and estimates suggest a one standard deviation increase in doctor quality reduces the likelihood of dropping out of performing CABG by between 2 and 3 percentage points (a 9 to 14 percent effect). Higher quality CT surgeons are also less likely to stop billing Medicare. The size of this effect ranges across specifications from 1.24 percentage points to 2.66 percentage points for a one standard deviation change in doctor quality. Again results are quite similar when doctors near retirement are excluded.[50]

Since we are interested in determining not only whether learning occurs, but also the importance of learning in the labor market, it is helpful to translate the effects on dropout behavior to impacts on patient mortality. Assuming dropout behavior has no effect other than to lower the average patient mortality rate for doctors continuing to practice, the -2.45 point estimate in Column 6 of Table 3 suggests learning reduces patient mortality by approximately .67 percent. Though small, this translates into approximately 476 Medicare beneficiaries saved over the ten year period. For ICs, using the -.79 point estimate in Column 3 of Table 3, approximately 63 Medicare beneficiaries are saved over the ten year period. Thus, while effects on patient mortality are modest, the large number of patients undergoing these procedures means they are meaningful from a social welfare perspective.

[49]Results by experience level further suggest differential dropout behavior occurs mostly among cohorts that are older, but not only among those nearing retirement. (Appendix Table A.4). This result contrasts with Chevalier and Ellison (1997), which finds that termination of mutual fund managers is more sensitive to returns for younger managers. This may reflect the fact that ability is not constant across careers: in this context learning by referring doctors about the (perhaps changing) quality of older specialists appears to be important. Results comparing dropout behavior by quality of doctors within HRRs are also consistent with the main results. Appendix Table A.6 displays results from regressions with HRR fixed effects for HRRs with at least 15 ICs. (15 was selected as the cutoff to ensure competition within HRRs and to avoid incidental parameters problems (Katz, 2001).)

[50]Regressions by experience level for CT surgeons suggest most of the action is again in older cohorts (those aged 48-55 in 1996) (see Appendix Table A.4). Looking at dropout behavior within HRRs produces similar results when using the in hospital quality measure (Appendix Table A.6).

Next I test the model's predictions for specialists' moving behavior by quality. The model suggests that lower quality specialists should be more likely to move, because they can increase their referral volumes through moving to markets with a relatively lower supply of specialists. Again I run logit regressions controlling for quality and doctor characteristics; this time with indicators for moving as outcome variables. Negative coefficients on quality again suggest a role for learning.

Results for ICs and CT surgeons are in Table 4. The top panel displays results for changing zip codes. Columns 1 and 3 again present results from the basic specification, Columns 2 and 4 include cohort fixed effects, and now Columns 3 and 6 exclude dropouts. For ICs coefficients are significantly different from zero at the ten percent level in all specifications, but only the coefficient in column 1 is significant at the five percent level. Point estimates suggest lower quality ICs are roughly 1.5 percentage points more likely to move. Results are again stronger in the declining market: lower quality CT surgeons are significantly more likely to move, with point estimates ranging from 1.8 to 2.4 percentage points. For ICs this is a 3 percent effect, and for CTs it is a 4-5 percent effect.

The second panel of Table 4 displays results for changing HRRs. While changing zip codes may not reflect a change in labor markets, a change in HRR is a larger move and necessarily a change in hospitals and in the local patient and PCP populations. The sample size for this analysis is reduced by 79 ICs and 37 CT surgeons with billing zip codes that do not match to HRRs. The point estimates are slightly smaller than before, but the quality effect now accounts for 10-14 of moves across HRRs for CT surgeons, respectively. Results by experience level suggest differential moving is most important in middle cohorts for ICs and in young and middle cohorts for CT surgeons (Appendix Table A.5).

To further investigate the nature of moves, Table 5 summarizes the characteristics of movers' first and last HRR. For the 666 ICs who change HRRs during the sample period, the HRR they are last observed in has fewer Medicare beneficiaries, lower Medicare reimbursement per beneficiary, and a lower ratio of ICs to Medicare beneficiaries. The latter can be thought of as a measure of local capacity, as we would expect areas with a lower number of doctors compared with potential patients to be more capacity constrained. A paired t-test rejects the null of equality between old and new HRRs on each of these characteristics. For CTs, the decline in the number of beneficiaries and the reimbursement per beneficiary are not statistically significant, but we can reject the null that the change in capacity is zero.[51] The table also summarizes the malpractice environment of the first

[51] Within the groups of movers, lower quality movers are not associated with larger reductions in capacity,

and last state for ICs and CT surgeons who move across state lines during the sample period. The number of claims per doctor is from the National Practitioner Data Bank for 2006, and state-level premiums are rates for general surgery reported in the 2000 Medical Liability Monitor. For these specialists, moving behavior does not seem to be driven by malpractice concerns.

Results thus far are weakly consistent with learning by referring doctors. To further investigate whether learning by referring doctors is driving dropout and moving behavior, Figure 2 plots claims over time for dropouts by the year of dropout and for non-dropouts. Under learning, we should expect to observe declines in practice volumes for doctors who drop out prior to their career change. The top panel includes results for ICs and the bottom for CT surgeons. ICs who drop out of the labor market experience declines in total procedures for 3 to 4 years prior to their dropout date. ICs who drop out of PCI experience similar declines in PCI before they refocus their careers on non-interventional procedures. Patterns are similar for CT surgeons, but the reduction in referrals appears to be for a longer period of time prior to the dropout date. This is consistent with dropout behavior being driven by reductions in referrals. It is also informative to look at referral volumes for specialists who move across HRRs. Figure 3 plots claims for movers by the year they move. The top panel shows results for ICs and the bottom for CT surgeons. For both specialties, claims drop prior to the move and recover quickly after the move.

Next I test whether lower quality specialists have lower growth rates in referrals over time compared with higher quality specialists. To do this I estimate models of the form:

$$y_{jt} = \beta_0 + \beta_1 f(exp_{jt}) + \beta_2 Q_j + \beta_3 (exp_{jt} \times Q_j) + \epsilon_{jt} \tag{10}$$

where y_{jt} is the outcome for doctor j in year t, Q_j is the quality measure for doctor j, exp_{jt} is the years experience of doctor j in year t, and ϵ_{jt} is a mean zero normally distributed error term. The regression includes a second degree polynomial in experience, with experience defined as the current year minus the year of graduation from medical school minus 6 for ICs and 7 for CT surgeons (the youngest IC cohort, graduating in 1989, thus has zero years experience in 1996), and the usual characteristics of the medical school and HRR. The coefficient on the interaction between experience and quality, β_3, is the effect of doctor quality on the growth rate of referrals over the career. $\beta_3 > 0$ indicates higher referral growth for higher quality doctors and is consistent with learning by RDs.

Estimates are in Table 6. The main outcomes examined are total claims (top panel)

HRR size, or HRR benefits generosity.

and total PCI / CABG claims (middle panel).[52] I also show results for non-emergency PCI / CABG claims (bottom panel) since referrals may be less important for emergency procedures.[53] For each quality measure, I run the basic model (first column), a specification with cohort fixed effects, and a version with cohort fixed effects excluding dropouts and movers (third column). Standard errors are clustered at the doctor level.

For ICs, β_3 is not significantly different from zero in the interaction with the quality measure. After ten years' experience we can rule out a difference in claims for higher quality doctors (by one standard deviation in the quality measure) larger than 3 percent. In the second and third panel β_3 is not significantly different from zero in any specification, but we do not rule out effects on PCI claims smaller than 5 percent (0.53=0.13+2*.20) after ten years. Point estimates and standard errors are larger in the bottom panel, and we cannot rule out differences in non-emergency PCI claims as large as 8-10 percent.[54]

Results are similar for CT surgeons. β_3 is not significantly different from zero in any specification. For claims, we can rule out effects after ten years experience greater than 14 percent using estimates in column 3. For CABG claims and non-emergency CABG claims we can rule out effects larger than 4.7 percent and 6.6 percent (both using column 3), respectively. Higher and lower quality CT surgeons do appear to differ significantly in claims at the beginning of the career, but these effects go away when dropouts and movers are excluded.

Table 7 displays results from several robustness checks. In columns 1 and 4, observations from 1996-1999 are dropped, since these are used in constructing the quality measures. Results are not substantively affected. Columns 2 and 5 drop doctors in large physician groups, as financial incentives (rather than quality) might drive referral patterns within these groups.[55] Columns 3 and 6 limit the sample to HRRs with at least 15 ICs or CT surgeons and include HRR fixed effects.[56] Throwing out large groups and accounting

[52]Results using unique patients, allowed charges, and patient risk factors, in Appendix Table A.8, are similar.

[53]Non-emergency procedures are identified by merging the Part B claim file with the inpatient claim using beneficiary ID, procedure codes, and dates of service. Approximately 90 percent of Part B claims were successfully matched to an inpatient claim containing information on emergency admission.

[54]Results from specifications with dummies indicating doctors with quality measures below the 33rd percentile are substantively similar

[55]I construct a measure of a physician's group size using the tax numbers physicians bill from. I construct the group size variable using a 20% random sample of the national Medicare Part B claims for 2000. I summarize the number of unique doctors in any specialty submitting claims from each tax number. Physicians billing from the same tax number are considered to be part of the same group.

[56]I select 15 as the cutoff point to ensure competition within HRRs and to avoid the incidental parameters problem (Katz, 2001).

for sorting of specialists across HRRs, we still cannot reject that referral growth is on the same trajectory for high and low quality specialists.[57]

To summarize, the results suggest lower quality specialists are more likely to drop out of the labor market and to move compared with high quality peers. While there is suggestive evidence that this behavior is driven by reductions in referrals, because movers and dropouts experience claims reductions before taking action, it is perhaps puzzling that I do not find strong evidence of divergence in referrals by quality for those remaining locally in practice. This could arise simply because the empirical approach does not have the power to pick up the divergence in claims. In fact, I cannot rule out divergence as large as the divergence in workers' wages by ability uncovered by Altonji and Pierret (1999).

Alternatively, the results could suggest there is another mechanism, other than adjustment of referrals, driving dropout and moving behavior of low quality specialists. While I cannot investigate the mechanism directly due to the data limitation that the referring doctor is unobserved,[58] here I consider several alternative mechanisms. First, consider the possibility that patients and not referring doctors are learning about specialist quality. The policy implications are similar if patients or referring doctors discipline quality through learning. However, learning by patients is less likely, because patients report largely relying on the advice of referring doctors for the types of treatments studied here (Center for Studying Health System Change, 2007). Moreover, we expect patients to have less repeated interaction with specialists and lower access to quality information. Alternatively, it is possible that specialists are themselves learning about their ability and dropping out accordingly, but this does not explain moving behavior or reductions in claims volumes prior to dropout and moving dates.

It is also possible that external oversight by medical licensure boards or hospitals might be driving results. However, oversight by medical boards is unlikely to explain moving behavior, and both types of action are quite rare.[59] Finally, it is possible that low

[57] As an additional robustness check, I analyze the full set of CABG cases, including non-Medicare cases, for the subsample of CT surgeons practicing in New Jersey. While the large majority (approximately 90%) of physicians in the U.S. are contracted with Medicare, it may be the case that doctors are able to differentially select patients on insurance status over the course of their career. By merging the CMS data with confidential report card data from the New Jersey Department of Health and Senior Services, I can observe the full sample of patients treated for CABG by each of these doctors. Results, available upon request, are similar. The coefficient of interest is not significantly different from zero in any specification, though large effects cannot be excluded due to large standard errors.

[58] While the data does include the referring physician as a field, the data is not of high quality.

[59] According to the National Practitioner Data Bank in 2006 there were 7,044 adverse action reports (including state licensure, clinical privileges, professional society membership, and DEA actions) for all physicians and dentists. Of these 63.2 percent were state licensure actions, and 11.9 percent were clinical

quality doctors are driven out of practice when they have malpractice claims filed against them. And a similar argument would suggest low quality specialists might move to areas with lower malpractice claims rates or lower premiums if they are afraid of liability or having difficulty obtaining coverage. However, movers do not seem to be moving to areas that are less litigious or with lower premiums (Table 5).[60]

7 Conclusion

This paper examines the the role of market learning about relative quality (or ability) in the labor market for medical specialists. Adapting models of employer learning, I model the referral relationship between referring doctors and specialists. The model illuminates incentives arising from this relationship and makes predictions for specialist careers. I then test the predictions of the model using confidential Medicare claims data from the universe of cardiac specialists in the U.S.

The empirical evidence is consistent with some degree of learning by referring doctors. Lower quality interventional cardiologists are more likely to stop doing PCI. They are also more likely to move their practice across geographic markets to avoid the effects of reputation on referrals. Moreover, the effects of quality on moving and dropout behavior are stronger for CT surgeons, for whom labor demand was declining over the sample period. Lower quality CT surgeons are more likely to stop doing CABG, to change geographic markets, and even to drop out of the labor market altogether.

The paper also demonstrates that specialists who drop out of the labor market experience declines in claims volumes for several years prior to dropping out. And specialists who change HRR on average experience declines in claims prior to moving and claims recovery thereafter. This evidence, combined with the fact that movers do not appear to be moving out of malpractice concerns, suggest that reputation impacts are driving dropout and moving behavior. However, the empirical strategy does not have enough power to pick up divergence in claims volumes by quality for specialists who remain locally in practice. I cannot rule out differences in claims and PCI/CABG claims after 10 years experience smaller than 5-10 percent for a one standard deviation difference in quality.

The empirical evidence is consistent with modest learning by referring doctors, and

privileges reports.

[60] Merging in the Florida Medical Malpractice Closed Claims data I determined that, for cardiac specialists in Florida, movers are actually less likely to have had a malpractice claim filed against them compared with non-movers.

it suggests specialists face a quality incentive even absent report card programs and pay-for-performance programs. But results are not consistent with strong quality incentives for specialists who remain locally in practice. Results also suggest learning impacts patient welfare: low quality specialists stopping doing PCI or stopping practicing altogether reduces patient mortality assuming adequate physician supply, and moving behavior may have distributional consequences.

This paper also presents a new methodology for measuring doctor quality using data on patient outcomes. My quality measure controls for nonrandom sorting of patients to higher and lower quality specialists based on risk characteristics. Failing to control for this type of sorting substantially penalizes specialists seeing relatively risky patients. This finding has implications for report card design and other programs aimed at measuring and evaluating providers based on quality. A similar methodology could also be applied to measure quality in other contexts. For example, one could construct a measure of teacher quality based on the likelihood that students drop out, controlling for sorting of students to teachers on observable characteristics.

References

[1] Altonji, Joseph G. and Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics*, 116:313-350.

[2] Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review*, 53(5):941-973.

[3] Barro, Jason and Nancy Beaulieu. 2003. "Selection and Improvement: Physician Responses to Financial Incentives." National Bureau of Economic Research Working Paper 10017.

[4] Beaulieu, Nancy. 2002. "Quality Information and Consumer Health Plan Choices," *Journal of Health Economics*, 21:43-63.

[5] Christensen, Clayton M., Bohmer, Richard and Kenagy, John. 2000. "Will disruptive innovations cure health care?" *Harvard Business Review*, 78:102-112.

[6] Bruno and Ridgway Research Associates, Inc. 2002. "ACC Cardiology Workforce Study conducted for the American College of Cardiology."

[7] Bundorf, M. Kate, Natalie Chun, Gopi Shah Goda, and Daniel P. Kessler. 2009. "Do Markets Respond to Quality Information? The Case of Fertility Clinics," *Journal of Health Economics*, 28(3):718-727.

[8] Burgess, James F., Cindy L. Christiansen, Sarah E. Michalak, and Carl N. Morris. 2000. "Medical Profiling: Improving Standards and Risk Adjustments Using Hierarchical Models." *Journal of Health Economics*, 19(3):291-309.

[9] Campbell, Stephen, David Reeves, Evangelos Kontopantelis, Elizabeth Middleton, Bonnie Sibbald and Martin Roland. 2007. "Quality of Primary Care in England with the Introduction of Pay for Performance." *New England Journal of Medicine*, 357:181-190.

[10] Chernew, Michael, Gautam Gowrisankaran, and Dennis P. Scanlon. 2001. "Learning and the Value of Information: the Case of Health Plan Report Cards," *Journal of Econometrics*, 144(1):156-174.

[11] Center for Studying Health System Change. Health Tracking Household Survey, 2007. [United States] [Computer file]. ICPSR26001-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-10-16.

[12] Chamberlain, Gary. 1982. "Multivariate Regression Models for Panel Data." *Journal of Econometrics*, 18(1):5-46.

[13] Chevalier, Judith and Glenn Ellison. 1999. "Career Concerns of Mutual Fund Managers." *Quarterly Journal of Economics*, 114(2):389-432.

[14] 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):342-346.

[15] Cutler, David M. and Robert S. Huckman. 2003. "Technology Development and Medical Productivity: The Diffusion of Angioplasty in New York State." *Journal of Health Economics*, 22:187-217.

[16] Dafny, Leemore and David Dranove. 2008. "Do Report Cards Tell Consumers Anything They Don't Already Know? The Case of Medicare HMOs." *The RAND Journal of Economics*, 39(3):790-821.

[17] Dranove, David, Daniel Kessler, Mark McClellan, and Mark A. Satterthwaite. 2003. "Is More Information Better? The Effects of Report Cards on Health Care Providers." *Journal of Political Economy*, 111(3):555-588.

[18] Dranove, David and Mark A. Satterthwaite. 1992. "Monopolistic Competition when Price and Quantity are not Perfectly Observable." *RAND Journal of Economics*, 23(4):518-534.

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

[20] Farber, Henry S. and Robert Gibbons. 1996. "Learning and Wage Dynamics." *Quarterly Journal of Economics*, 111(4):1007-1047.

[21] Fournier, Gary M. and Melayne Morgan McInnes. 2002. "The Effects of Managed Care on Medical Referrals and the Quality of Specialty Care." *Journal of Industrial Economics*, 50(4):457-473.

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

[23] Gaynor, Martin, James B. Rebitzer, and Lowell J. Taylor. 2004. "Incentives in HMOs." *Journal of Political Economy*, 112(4):915-931.

[24] Glance, Laurent G., Andrew Dick, Turner M. Osler, Yue Li, and Dana B. Mukamel. 2006. "Impact of Changing the Statistical Methodology on Hospital and Surgeon Ranking: The Case of the New York State Cardiac Surgery Report Card." *Medical Care*, 44(4):311-319.

[25] Green, Jesse, and Neil Wintfeld. 1995. "Report Cards on Cardiac Surgeons - Assessing New York State's Approach." *New England Journal of Medicine*, 332:1229-32.

[26] Hemenway, David, Alice Killen, Suzanne B. Cashman, Cindy Lou Parks, and William J. Bicknell. 1990. "Physicians' Response to Financial Incentives: Evidence from a for-Profit Ambulatory Care Center." *The New England Journal of Medicine*, 322:1059-1063.

[27] Hannan, Edward L., Harold Kilburn, Joseph F. O'Donnell, Gary Lukacik, and Eileen P. Shields, E. 1990. "Adult Open Heart Surgery in New York State. An Analysis of Risk Factors and Hospital Mortality Rates." *Journal of the American Medical Association*, 264:2768-2774.

[28] Hannan, Edward L., Dinesh Kumar, Michael Racz, Albert L. Siu, and Mark R. Chassin. 1994. "New York State's Cardiac Surgery Reporting System: Four Years Later." *Annals of Thoracic Surgery*, 58(6):1852-57.

[29] Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy*, 87(5):972-990.

[30] Katz, Ethan. 2001. "Bias in Conditional and Unconditional Fixed Effects Logit Estimation." *Political Analysis*, 9(4):379-384.

[31] Kolstad, Jonathan. 2010. "Information and Quality when Motivation is Intrinsic: Evidence from Surgeon Report Cards." University of York Health Econometrics Group Working Paper 10/18.

[32] Kolstad, Jonathan T. and Michael E. Chernew. Forthcoming. "Quality and Consumer Decision Making in the Market for Health Insurance and Health Care Services." *Medical Care Research and Review*.

[33] Localio, A. Russell, Bruce H. Hamory, Alicia Cocks Fisher, and Thomas R. TenHave. 1997. "The Public Release of Hospital and Physician Mortality Data in Pennsylvania: A Case Study." *Medical Care*, 35(3):272-286.

[34] LeGrand, Julian. 1999. "Competition, Cooperation or Control? Tales from the British National Health Service." *Health Affairs*, 18(3):27-39.

[35] McClellan, Mark and Douglas Staiger. 2000. "Comparing the Quality of Health Care Providers." In *Frontiers in Health Policy Research*, Volume 3, ed. Alan Garber, 113-136. Cambridge: The MIT Press.

[36] Moretti, Enrico. Forthcoming. "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales." *Review of Economic Studies*.

[37] Morris, Carl N. 1983. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association*, 78(381):47-55.

[38] Mukamel, Dana B., and Alvin I. Mushlin. 1998. "Quality of Care Information Makes a Difference: An Analysis of Market Share and Price Changes Following Publication of the New York State Cardiac Surgery Mortality Reports. *Medical Care* 36(7):945-954.

[39] Mukamel, Dana B., David L. Weimer, Jack Zwanziger, Huang Gorthy Shih-Fang, Alvin A. Mushlin. 2004. "Quality Report Cards, Selection of Cardiac Surgeons, and Racial Disparities: A Study of the Publication of the New York State Cardiac Surgery Reports." *Inquiry*, 41(4):435-446.

[40] Mullen, Kathleen J., Richard G. Frank, and Meredith B. Rosenthal. 2010. "Can You Get What You Pay For? Pay-for-Performance and the Quality of Healthcare Providers." *RAND Journal of Economics*, 41(1):64-91.

[41] Mundlak, Yair. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica*, 46(1):69-85.

[42] Nolte, Ellen and C. Martin McKee. 2008. "Measuring the Health of Nations: Updating an Earlier Analysis." *Health Affairs*, 27(1): 58-71.

[43] Normand, Sharon-Lise T., Mark E. Glickman, and Constantine Gatsonis. 1997. "Statistical Methods for Profiling Providers of Medical Care: Issues and Applications." *Journal of the American Statistical Association*, 92(439):803-814.

[44] O'Connor, Gerald T., Stephen K. Plume, Elaine M. Olbstead, Laurence H. Coffin, Jeremy R. Morton, Christopher T. Maloney, Edward R. Nowicki, Joan F. Tryzelaar, Felix Hernandez, Lawrence Adrian, Kevin J. Casey, David N. Soule, and Charles A. S. Marrin, William C. Nugent, David C. Charlesworth, Robert Clough, Saul Katz, Bruce J. Leavitt, and John E. Wennberg. 1991. "A Regional Prospective Study of In-hospital Mortality Associated with Coronary Artery Bypass Grafting." *Journal of the American Medical Association*, 266(6):803-809.

[45] 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):993-999.

[46] Pope, Devin G. 2008. "Reacting to Rankings: Evidence from 'America's Best Hospitals."' *Journal of Health Economics*, 28(6): 1154-1165.

[47] Rosenthal, Meredith B. and Richard G. Frank. 2006. "What Is the Empirical Basis for Paying for Quality in Health Care?" *Medical Care Research and Review*, 63:135-157.

[48] Rosenthal, Meredith B., Richard G. Frank, Zhonghe Li, and Arnold M. Epstein. 2005. "Early Experience with Pay-for-Performance: From Concept to Practice." *Journal of the American Medical Association*, 294(14):1788-1793.

[49] Schneider, Eric C. and Arnold M. Epstein. 1996. "Influence of Cardiac-Surgery Performance Reports on Referral Practices and Access to Care: A Survey of Cardiovascular Specialists." *New England Journal of Medicine*, 335:251-256.

[50] Shemin, Richard J., Stanley W. Cziuban, Larry R. Kaiser, James E. Lowe, William C. Nugent, Mehmet C. Oz, Donald A. Turney, and Jamie K. Wallace. 2002. "Thoracic Surgery Workforce: Snapshot at the End of the Twentieth Century and Implications for the New Mellinium." *Annals of Thoracic Surgery*, 73:2014-32.

[51] TenHave, Thomas R. and A. Russell Localio. 1999. "Empirical Bayes Estimation of Random Effects Parameters in Mixed Effects Logistic Regression Models." *Biometrics*, 55(4):1022-1029.

[52] Thomas, Neal, Nicholas T. Longford, and John E. Rolph. 1994. "Empirical Bayes Methods for Estimating Hospital-Specific Mortality Rates." *Statistics in Medicine*, 13(9):889-903.

[53] Wang, Justin, Jason Hockenberry, Shin-Yi Chou, and Muzhe Yang. Forthcoming. "Do Bad Report Cards Have Consequences? Impacts of Publicly Reported Provider Quality Information on the CABG Market in Pennsylvania." *Journal of Health Economics*.

[54] Wedig, Gerard J. and M. Tai-Seale (2002), "The Effect of Report Cards on Consumer Choice in the Health Insurance Market." *Journal of Health Economics*, 21:1031-1048.

[55] Wennberg, John E., Elliott S. Fisher, David C. Goodman, and Jonathan S. Skinner. 2008. "Tracking the Care of Patients with Severe Chronic Illness - The Dartmouth Atlas of Health Care 2008."

[56] Williams, Sankey V., David B. Nash and Neil Goldfarb. 1991. "Differences in Mortality from Coronary Artery Bypass Graft Surgery at Five Teaching Hospitals." *Journal of the American Medical Association*, 266(6):810-815.

[57] World Health Organization. 2010. "WHO World Health Statistics." http://www.who.int/whosis/whostat/2010/en/index.html (accessed January 12, 2011).

Table 1: Physician Summary Statistics

	ICs		CT Surgeons	
	Mean	Std. Dev.	Mean	Std. Dev.
Physician characteristics				
Foreign medical school (%)	23	42	17	37
Prestigious medical school (%)	15	36	20	40
HRR beneficiaries	148,083	105,283	149,152	107,199
HRR spending per beneficiary	6,438	1,358	6,496	1,382
Experience	15.7	6.7	18.5	8.5
Practice volume				
Claims	5,452	3,642	529	734
PCI/CABG claims	63	46	76	45
Non-emergency PCI / CABG claims	19	21	28	21
Allowed charges	407,504	265,160	247,259	152,386
Patients	1,155	678	216	203
Case mix - claims (%)				
PCI/CABG	3.9	5.6	28.1	21.6
Other procedures	6.5	7.7	26.6	17.5
Testing	33.4	15.7	5.4	13
Lab	5.8	7.9	0.42	3.7
Evaluation and management	36.2	13.5	39.4	21.9
Case mix - charges (%)				
PCI/CABG	19.7	13.1	43.6	21.7
Other procedures	15.0	8.7	33.9	19.8
Testing	21.7	12.5	2.5	8.2
Lab	1.0	1.8	0.11	2.11
Evaluation and management	23.2	10.4	20	21.4
Dropout and moving behavior				
PCI/CABG dropouts	10.2	30.3	22.2	41.6
Medicare dropouts	2.7	16.1	16.0	36.7
Change in zipcode	45.5	49.8	45.1	49.8
Change in HRR	16.1	36.7	23.8	42.6
N	4,228		2,814	

Sample includes ICs and CT surgeons performing at least 30 procedures (PCI / CABG) by 1999. Specialists without records in the MPIER file, and specialists who have zero or fewer years of experience, no billing zip code or no graduation year are dropped. Statistics are calculated from annual measures for years 1996-2005.

Table 2: Doctor Quality Measures

	ICs	CT Surgeons
Distribution of measures:		
10th percentile	-0.14	-0.25
25th percentile	-0.062	-0.13
50th percentile	0.013	0.0057
75th percentile	0.073	0.13
90th percentile	0.12	0.24
90-10 gap	0.26	0.49
Mean	0.00029	-0.00066
Std. deviation	0.10	0.19
N	4,228	2,814

Sample includes ICs and CT surgeons performing at least 30 procedures (PCI/CABG) by 1999. Column (1) summarizes doctor the quality measure based on the death in-hospital patient outcome for ICs, Column (2) summarizes the quality measure for CT surgeons. Quality measures are constructed in a two-step procedure as described in Section V.

Table 3: Analysis of Dropout Behavior

	ICs			CT Surgeons		
PCI/CABG dropout	(1)	(2)	(3)	(4)	(5)	(6)
Doctor Quality	-0.99	-0.84	-0.79	-3.24	-2.01	-2.45
	(0.44)	(0.40)	(0.39)	(0.78)	(0.74)	(0.68)
Foreign School	-2.58	-3.64	-3.77	-2.58	-3.64	-1.44
	(1.07)	(0.88)	(0.86)	(1.07)	(0.88)	(1.94)
Prestigious School	0.01	-1.29	-2.16	0.01	-1.29	0.16
	(1.29)	(1.04)	(1.01)	(1.29)	(1.04)	(1.79)
HRR size (beneficiaries)	-0.20	-0.10	0.01	-0.39	0.58	-0.40
	(0.54)	(0.49)	(0.48)	(0.93)	(0.87)	(0.84)
HRR size (reimbursement)	-0.39	-0.18	0.11	-2.83	-3.61	2.40
	(0.55)	(0.51)	(0.48)	(0.96)	(0.96)	(0.89)
Cohort fixed effects		Yes	Yes		Yes	Yes
N	4228	4228	3987	2814	2814	2377
Pseudo R-squared	0.0047	0.081	0.040	0.012	0.14	0.073
Medicare dropout						
Doctor Quality	-0.31	-0.17	-0.18	-2.66	-1.24	-1.61
	(0.23)	(0.15)	(0.16)	(0.67)	(0.54)	(0.49)
Foreign School	0.19	-0.30	-0.32	0.19	-0.30	0.15
	(0.62)	(0.36)	(0.37)	(0.62)	(0.35)	(1.45)
Prestigious School	1.07	0.24	0.02	1.07	0.24	-0.12
	(0.78)	(0.45)	(0.46)	(0.78)	(0.45)	(1.31)
HRR size (beneficiaries)	0.58	0.41	0.35	-0.32	0.56	-0.17
	(0.28)	(0.20)	(0.20)	(0.81)	(0.65)	(0.63)
HRR size (reimbursement)	-0.77	-0.51	-0.32	-2.36	-2.85	-2.31
	(0.31)	(0.22)	(0.22)	(0.84)	(0.68)	(0.66)
Cohort fixed effects		Yes	Yes		Yes	Yes
N	4228	4006	3765	2814	2814	2377
Pseudo R-squared	0.011	0.13	0.073	0.014	0.18	0.11

Table displays results from logit regressions of dropout indicators on quality measures and doctor characteristics. The top panel is for regressions using indicators for dropping out of PCI/CABG, and the bottom panel uses indicators for stopping billing Medicare. Columns (1), (2) and (3) display results for ICs; the remaining columns display results for CT surgeons. Columns (1), (2), (4) and (5) display results for the full samples; Columns (3) and (6) exclude specialists who are over 65 in 2005. 222 ICs in the 1986 graduation year cohort are dropped from fixed effect regressions (Columns (2) and (3)) in the bottom panel due to a lack of variation in the dependent variable. Effects are for a one standard deviation change in X on the probability of dropping out (in percentage points). For dummy variables the effect is for a discrete change from 0 to 1. Robust standard errors are in parentheses.

Table 4: Analysis of Moving Behavior

	ICs			CT Surgeons		
Change in zipcode	(1)	(2)	(3)	(4)	(5)	(6)
Doctor Quality	-1.63	-1.45	-1.60	-1.83	-2.30	-2.43
	(0.77)	(0.78)	(0.83)	(0.94)	(0.99)	(1.13)
Foreign School	0.38	1.82	2.65	-3.74	4.41	2.28
	(1.89)	(1.92)	(2.02)	(2.59)	(2.89)	(3.38)
Prestigious School	-2.46	-2.19	-1.87	-3.85	-2.19	-1.64
	(2.19)	(2.22)	(2.35)	(2.39)	(2.50)	(2.86)
HRR size (beneficiaries)	1.19	0.97	0.87	3.93	3.30	3.02
	(0.88)	(0.89)	(0.94)	(1.10)	(1.13)	(1.28)
HRR size (reimbursement)	-0.70	-0.75	-0.49	-2.43	-2.06	-2.07
	(0.89)	(0.91)	(0.95)	(1.10)	(1.14)	(1.30)
Cohort fixed effects		Yes	Yes		Yes	Yes
N	4228	4228	3722	2814	2814	2186
Pseudo R-squared	0.0013	0.014	0.0090	0.0052	0.048	0.045
Change in HRR						
Doctor Quality	-0.77	-0.55	-0.83	-2.57	-3.26	-2.95
	(0.52)	(0.51)	(0.53)	(0.81)	(0.81)	(0.92)
Foreign School	2.24	4.08	5.16	-7.43	1.75	2.76
	(1.46)	(1.45)	(1.53)	(2.03)	(2.58)	(3.06)
Prestigious School	1.68	2.38	1.98	-2.05	0.16	-0.74
	(1.70)	(1.69)	(1.75)	(1.99)	(2.06)	(2.33)
HRR size (beneficiaries)	-2.06	-2.25	-2.17	-1.33	-2.28	-1.93
	(0.66)	(0.63)	(0.65)	(0.95)	(0.95)	(1.08)
HRR size (reimbursement)	0.19	0.11	0.37	0.50	0.91	0.77
	(0.68)	(0.66)	(0.68)	(0.94)	(0.93)	(1.07)
Cohort fixed effects		Yes	Yes		Yes	Yes
N	4149	4149	3728	2778	2729	2145
Pseudo R-squared	0.0045	0.059	0.066	0.0083	0.090	0.087

Table displays results from logit regressions of moving indicators on quality measures and doctor characteristics. The top panel is for regressions using indicators for changing zipcodes, and the bottom panel uses indicators for changing HRRs. Columns (1), (2) and (3) display results for ICs; the remaining columns display results for CT surgeons. Columns (1), (2), (4) and (5) display results for the full sample; Columns (3) and (6) exclude dropouts. Regressions with change in HRR as the outcome variable exclude specialists with zipcodes that do not merge to HRRs. 49 observations are dropped in Columns (5) and (6) in the lower panel - these are CT surgeons in the 1952-1955 and 1957 cohorts, for whom there is no variation in the dependent variable. Effects are for a one standard deviation change in X on the probability of dropping out (in percentage points). For dummy variables the effect is for a discrete change from 0 to 1. Robust standard errors are in parentheses.

Table 5: Characteristics of First and Last HRR for Movers

ICs	First HRR	Last HRR	T-statistic	N
Beneficiaries	139,120	129,460	3.2	666
	(3,820)	(3,820)		
Reimbursement per beneficiary	6,481	6,335	2.4	666
	(55)	(53)		
Doctors/beneficiary	0.0034	0.0030	4.7	666
	(0.000055)	(0.000051)		
Malpractice cases/doctor	0.30	0.29	1.5	482
	(0.005)	(0.005)		
Malpractice premium	28,160	28,858	-0.69	482
	(681)	(685)		
CT Surgeons				
Beneficiaries	143,920	134,360	1.8	665
	(4,160)	(3,980)		
Reimbursement per beneficiary	6,567	6,457	1.8	665
	(56)	(54)		
Doctors/beneficiary	0.0024	0.0021	4.2	665
	(0.000049)	(0.000046)		
Malpractice cases/doctor	0.29	0.30	-0.37	451
	(0.0051)	(0.11)		
Malpractice premium	28,098	29,087	-1.0	451
	(661)	(701)		

Table displays mean characteristics of the first and last HRR of specialists who change HRR during the sample period. T-statistics are for the paired t-test that the mean difference between the first and last HRR is zero. Standard errors are in parentheses. Malpractice cases/doctor and malpractice premium vary at the state level, so these tests are for the sample of specialists changing states during the sample period.

Table 6: Analysis of Claim Volumes

	ICs			CT Surgeons		
Claims	(1)	(2)	(3)	(4)	(5)	(6)
Quality	-0.89	-0.32	-0.05	17.69	18.90	24.93
	(2.05)	(2.02)	(2.39)	(9.51)	(10.54)	(23.04)
Quality * Experience	0.08	0.04	-0.01	-1.02	-1.12	-1.90
	(0.13)	(0.13)	(0.15)	(0.68)	(0.77)	(1.65)
Experience	6.88	10.04	9.23	5.72	10.74	9.11
	(0.41)	(0.35)	(0.38)	(0.64)	(1.37)	(1.89)
Experience squared	-0.16	-0.16	-0.11	-0.15	-0.28	-0.19
	(0.01)	(0.01)	(0.01)	(0.02)	(0.04)	(0.06)
Cohort fixed effects		Yes	Yes		Yes	Yes
N	42,280	42,280	31,400	28,140	28,140	16,180
R-squared	0.065	0.090	0.10	0.015	0.027	0.050
PCI/CABG Claims						
Quality	2.38	3.09	3.91	14.71	14.73	8.06
	(2.69)	(2.68)	(3.11)	(2.29)	(2.27)	(3.46)
Quality * Experience	0.19	0.18	0.13	-0.21	-0.19	0.15
	(0.17)	(0.17)	(0.20)	(0.10)	(0.10)	(0.16)
Experience	4.41	8.96	8.10	2.06	0.39	-0.63
	(0.43)	(0.41)	(0.46)	(0.40)	(0.45)	(0.60)
Experience squared	-0.13	-0.21	-0.15	-0.09	-0.19	-0.13
	(0.012)	(0.011)	(0.012)	(0.009)	(0.011)	(0.016)
Cohort fixed effects		Yes	Yes		Yes	Yes
N	42,280	42,280	31,400	28,140	28,140	16,180
R-squared	0.018	0.033	0.034	0.079	0.14	0.099
Non-emergency PCI/CABG Claims						
Quality	6.92	7.89	7.15	17.28	17.51	8.37
	(4.81)	(4.79)	(5.89)	(3.31)	(3.32)	(5.06)
Quality * Experience	0.29	0.26	0.34	-0.25	-0.25	0.20
	(0.30)	(0.30)	(0.38)	(0.15)	(0.15)	(0.23)
Experience	6.26	12.26	11.90	2.89	3.02	1.92
	(0.78)	(0.75)	(0.92)	(0.56)	(0.67)	(0.86)
Experience squared	-0.17	-0.25	-0.19	-0.11	-0.22	-0.14
	(0.02)	(0.02)	(0.03)	(0.01)	(0.02)	(0.02)
Cohort fixed effects		Yes	Yes		Yes	Yes
N	38,052	38,052	28,260	25,326	25,326	14,562
R-squared	0.018	0.028	0.030	0.056	0.076	0.046

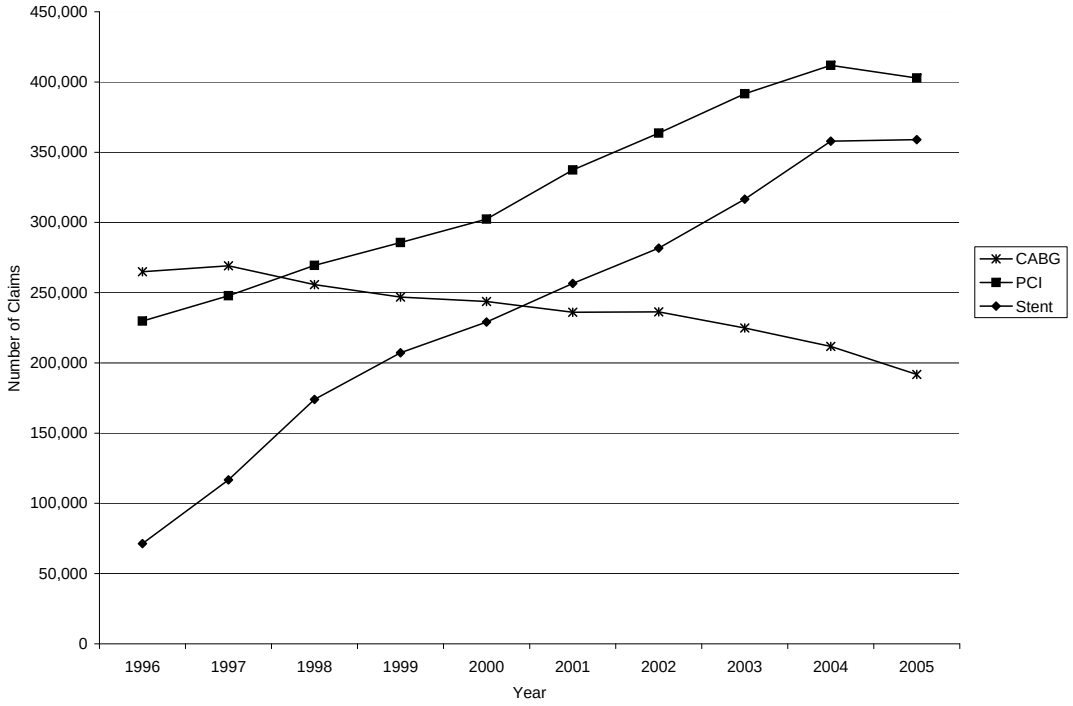
Table displays results from OLS regressions of claim volumes on quality, experience and a quality-experience interaction. All regressions include experience squared and the usual physician characteristics. In the top panel the outcome variable is Medicare claims, in the middle panel it is PCI/CABG claims, in the bottom panel it is non-emergency PCI/CABG claims. Non-emergency PCI/CABG regressions exclude observations from 1996, as these cannot be linked to inpatient records. Columns (1), (2) and (3) display results for ICs; the remaining columns display results for CT surgeons. Columns (1), (2), (4) and (5) display results for the full sample; Columns (3) and (6) exclude dropouts and movers. Quality effects are for a one standard deviation change in quality expressed as a percentage of the outcome variable. Experience effects are dy/dx as a percentage of the outcome variable. Standard errors, in parentheses, are clustered at the doctor level.

Table 7: Analysis of Claim Volumes - Robustness

	ICs			CT Surgeons		
Claims	(1)	(2)	(3)	(4)	(5)	(6)
Quality	-1.30	0.56	3.88	23.42	5.07	51.31
	(3.04)	(2.75)	(3.00)	(12.62)	(4.91)	(35.37)
Quality * Experience	0.10	-0.10	-0.18	-1.17	0.28	-3.48
	(0.16)	(0.18)	(0.19)	(0.73)	(0.26)	(2.57)
Experience	10.35	10.18	9.22	13.35	7.29	10.61
	(0.51)	(0.44)	(0.46)	(2.24)	(0.68)	(3.68)
Experience squared	-0.15	-0.12	-0.11	-0.28	-0.13	-0.25
	(0.01)	(0.01)	(0.01)	(0.06)	(0.02)	(0.12)
HRR fixed effects			Yes			Yes
N	25,368	24,050	20,530	16,884	13,370	7,600
R-squared	0.070	0.11	0.24	0.027	0.038	0.20
PCI/CABG Claims						
Quality	4.47	2.18	5.75	14.29	8.56	6.81
	(3.97)	(3.47)	(3.67)	(2.71)	(3.79)	(4.32)
Quality * Experience	0.12	0.23	0.05	-0.20	0.25	0.04
	(0.21)	(0.22)	(0.23)	(0.11)	(0.17)	(0.20)
Experience	6.35	7.92	7.82	-3.98	-1.06	0.19
	(0.58)	(0.53)	(0.58)	(0.58)	(0.66)	(0.89)
Experience squared	-0.16	-0.15	-0.15	-0.07	-0.13	-0.15
	(0.015)	(0.014)	(0.016)	(0.013)	(0.017)	(0.024)
HRR fixed effects			Yes			Yes
N	25,368	24,050	20,530	16,884	13,370	7,600
R-squared	0.030	0.035	0.13	0.13	0.12	0.29
Non-emergency PCI/CABG Claims						
Quality	6.76	2.64	11.66	17.28	7.08	6.53
	(6.51)	(6.44)	(6.52)	(3.84)	(5.52)	(6.29)
Quality * Experience	0.33	0.49	0.04	-0.27	0.38	0.18
	(0.36)	(0.42)	(0.40)	(0.16)	(0.24)	(0.30)
Experience	9.45	12.12	11.62	-2.19	2.08	1.94
	(1.11)	(1.07)	(1.15)	(0.87)	(0.97)	(1.24)
Experience squared	-0.21	-0.20	-0.20	-0.12	-0.14	-0.15
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)
HRR fixed effects			Yes			Yes
N	25,368	21,645	18,477	16,884	12,033	6,840
R-squared	0.023	0.032	0.099	0.084	0.064	0.22

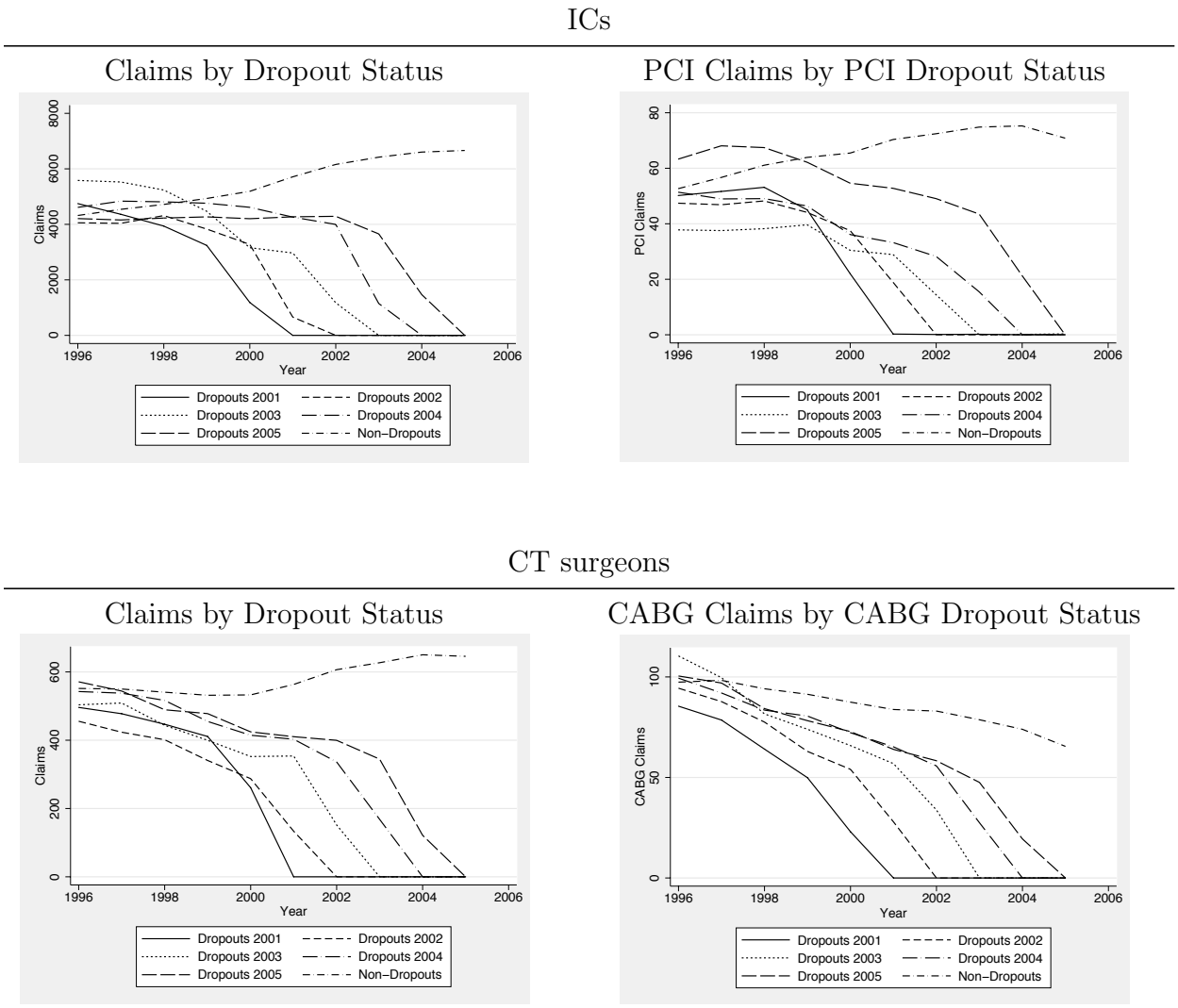
Table displays results from OLS regressions of claim volumes on quality, experience and a quality-experience interaction. All regressions include the usual physician characteristics, and cohort fixed effects. In the top panel the outcome variable is Medicare claims, in the middle panel it is PCI/CABG claims, in the bottom panel it is non-emergency PCI/CABG claims. Non-emergency PCI/CABG regressions exclude observations from 1996, as these cannot be linked to inpatient records. Columns (1), (2) and (3) display results for ICs; the remaining columns display results for CT surgeons. Columns (1) and (4) drop observations from 1996-1999; Columns (2) and (5) drop specialists in large group practices; in Columns (3) and (6) the sample is restricted to specialists in HRRs with at least 15 ICs. Quality effects are for a one standard deviation change in quality expressed as a percentage of the outcome variable. Experience effects are dy/dx as a percentage of the outcome variable. Standard errors, in parentheses, are clustered at the doctor level.

Figure 1: CABG, PCI and Stent Claims Over Time



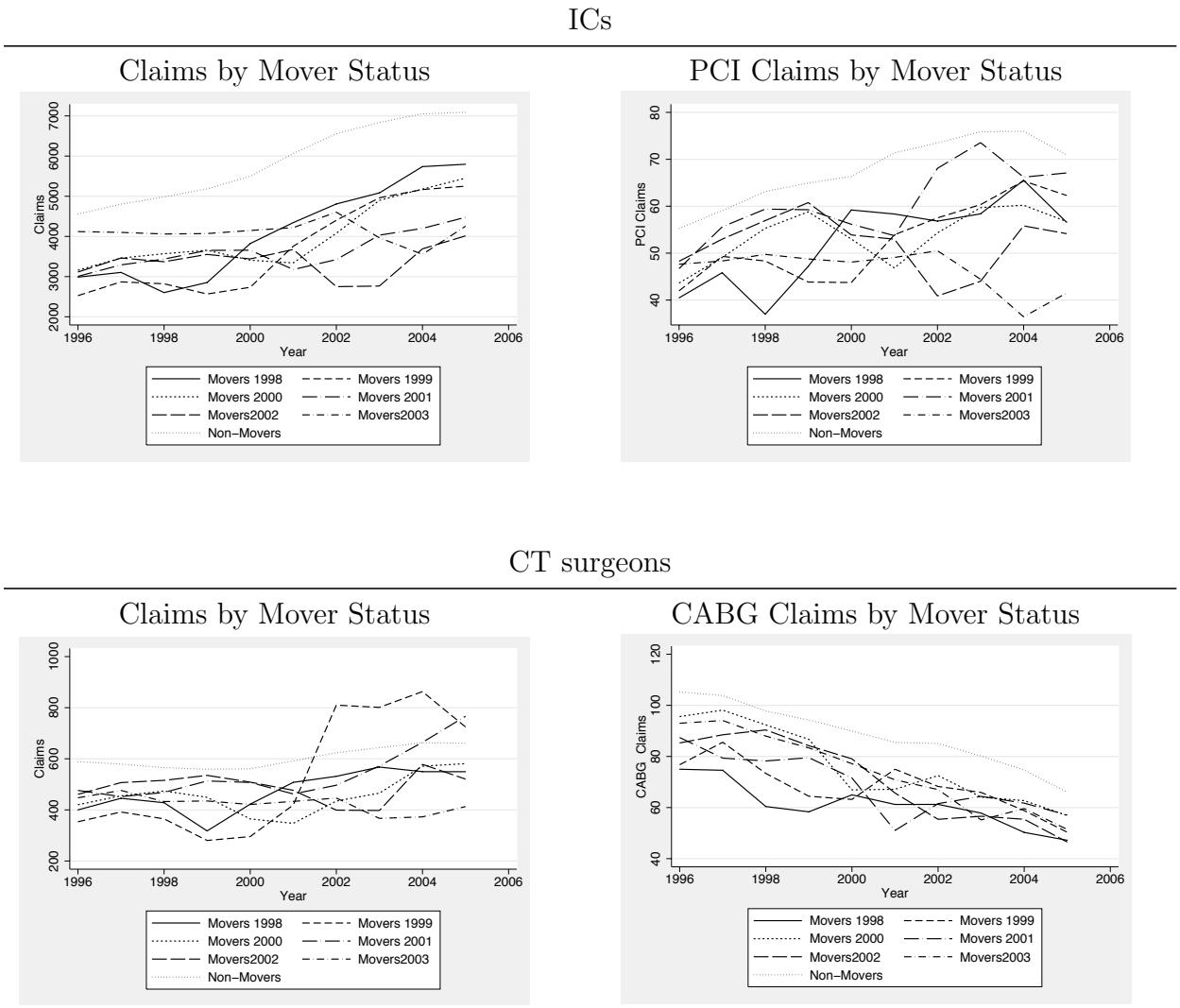
Plot points are the total number of Medicare Part B claims involving each procedure (PCI (of any kind), PCI with stent placement, and CABG.) Duplicate and denied claims have been excluded.

Figure 2: Time Trends in Claim Volumes by Dropout Status



Figures display referral volumes over time for dropouts and non-dropouts. For dropouts, volumes are plotted separately for each dropout cohort, with the cohort defined by the year they drop out. The top panel displays results for ICs and the bottom panel for CT surgeons. The figures on the left plot claims for doctors who drop out of billing Medicare, and the figures on the right plot PCI/CABG claims for doctors who drop out of performing PCI or CABG.

Figure 3: Time Trends in Claim Volumes by Mover Status



Figures display referral volumes over time for movers and non-movers. For movers, volumes are plotted separately for each mover cohort, with the cohort defined by the year of the move. The top panel displays results for ICs and the bottom panel for CT surgeons. The figures on the left plot claims, and the figures on the right plot PCI/CABG claims.

Appendix - Not for Publication

A Data Extract Details and Supplementary Tables

The primary file used in the analysis is an extract created by Centers for Medicare and Medicaid Services (CMS) from the 100 percent CMS Carrier file, which contains all final action claims submitted to Medicare by non-institutional providers (i.e, physicians, non-md practitioners such as physician assistants and nurse practitioners, ambulance providers and free-standing surgery centers). This file includes icd-9 diagnosis codes and HCPCS procedure codes, dates of service, reimbursement amount, and CMS identification (UPIN) numbers for the performing providers. The extract was produced by CMS contractors by selecting claims from physicians on a submitted list.[61] All claims with the UPIN as the referring or performing physician were selected.

The submitted list was designed to include all practicing ICs and CT surgeons. The UPINs were identified using the 100 percent sample of the Medicare inpatient claims file for 1998-2005, and a 20 percent random sample of the Carrier file for 1998 and 2005. UPIN numbers appearing on more than five CABG claims in any two years in the inpatient files were included on the list.[62] In addition to these UPINs, any physician listed as the performing doctor on greater than five Carrier file claims for CABG in either 1998 or 2005 were included on the list.

ICs were identified similarly. UPIN numbers appearing on more than five PCI claims in any two years in the inpatient files were included on the list.[63] In addition to these UPINs, any physician listed as the performing doctor on greater than five Carrier file claims for PCI in either 1998 or 2005 were included on the list.

[61] The logical alternative approach would have been to have CMS contractors identify ICs and CT surgeons using the 100 percent Carrier file and procedure codes. Then after identifying ICs and CT surgeons, in the second step they could pull all claims associated with those UPINs. However, this two-step approach would have doubled the cost of the data.

[62] When one or more physician on a claim listed their specialty as cardiothoracic surgery, the cardiothoracic surgeons were the UPINs identified with the claim, whether they were listed as the attending, operating or other physician. When no cardiothoracic surgeon was on the claim, the operating UPIN was identified with the claim. If this was blank, then the attending, followed by the other physician were used.

[63] When one or more physician on the claims listed their specialty as cardiology, the cardiologists' UPINs were the UPINs identified with the claim, whether they were listed as the attending, operating or other physician. When no cardiologist was on the claim, the operating UPIN was identified with the claim. If this was blank, then the attending, followed by the other physician were used.

Table A.1: Patient Level Summary Statistics

	PCI Sample		CABG Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Outcome Measure (%)				
Death in hospital	1.4	12.0	3.0	17.0
Risk Characteristics (%)				
AMI	20.0	40.0	5.0	21.9
Heart failure	5.4	23.0	2.6	15.8
Peripheral vasc disease	1.7	13.0	1.6	12.5
Cerebrovascular disease	0.59	7.7	1.4	11.9
Chronic pulmonary disease	0.82	9.0	0.78	8.8
Rheumatoid arthritis	0.016	1.3	0.017	1.3
Diabetes w/o complications	1.5	12.0	1.4	11.6
Renal dsease	0.29	5.3	0.29	5.4
Demographics				
Disability (%)	9.4	29.0	5.1	22.0
End stage renal disease (%)	1.0	10.0	0.65	8.0
Age	72.0	8.2	72.7	7.2
Female (%)	43.0	49.0	36.4	48.1
Black (%)	5.3	22.0	4.1	19.9
Asian (%)	0.57	7.5	0.72	8.4
Hispanic (%)	1.4	12.0	1.6	12.6
Zipcode income	21,163	9,031	21,174	9,062
Procedure Information (%)				
Angioplasty only	41.4	49.3		
Stent	54.5	49.8		
Count of vessels	121	47.8		
Left circumflex artery	12.7	33.3		
Left anterior descending	17.8	38.2		
Right coronary artery	16.7	37.3		
Diagnostic cath	43.7	49.6		
Reoperation			10.0	30.0
Artery			68.5	46.5
Vein			28.6	45.2
Artery & vein			62.9	48.3
Number vessels			2.65	1.08
Number arteries			1.13	0.45
N	1,067,018		1,019,770	

Sample includes all claims involving PCI or CABG in 1996-1999, as determined from procedure codes. Excludes duplicate and denied claims.

Table A.2: Coefficients from Correlated Random Effects Logits

	PCI Sample		CABG Sample	
	Regressors	Mean Regressors	Regressors	Mean Regressors
Risk Characteristics				
AMI	1.57	-1.703	0.678	-1.547
	(0.025)	(0.6)	(0.033)	(0.47)
Heart Failure	0.695	-0.0314	0.98	-1.588
	(0.04)	(0.31)	(0.037)	(0.79)
Periph Vasc Disease	0.0589	0.942	1.137	-1.154
	(0.069)	(0.64)	(0.032)	(0.29)
Chronic Pulmonary Disease	0.13	0.787	-0.0942	0.583
	(0.085)	(0.67)	(0.074)	(0.67)
Diabetes w/o Complications	-0.478	-1.388	-0.0469	-1.405
	(0.087)	(0.47)	(0.057)	(0.46)
Renal Disease	0.775	2.095	0.614	4.424
	(0.1)	(1.72)	(0.077)	(1.75)
Demographics				
Disability	0.406	-0.0293	0.376	-1.166
	(0.053)	(0.62)	(0.035)	(0.76)
End stage renal disease	1.297	1.166	1.165	2.512
	(0.079)	(1.11)	(0.054)	(1.46)
Age	-0.0238	-0.513	-0.00518	-0.103
	(0.013)	(0.16)	(0.01)	(0.22)
Age squared	0.000605	0.00372	0.000305	-0.000029
	(0.000088)	(0.0011)	(0.000068)	(0.0015)
Female	0.809	1.013	0.79	-6.534
	(0.16)	(2.37)	(0.12)	(3.03)
Black	0.0225	4.037	0.857	6.381
	(0.3)	(3.36)	(0.22)	(4.76)
Asian	0.0592	-0.699	0.332	-0.614
	(0.11)	(0.55)	(0.064)	(0.45)
Hispanic	0.106	0.0892	0.0582	1.082
	(0.072)	(0.32)	(0.048)	(0.27)
Age*Female	-0.00651	-0.0108	-0.00525	0.0884
	(0.0021)	(0.032)	(0.0016)	(0.041)
Black*Female	-0.0497	-1.814	-0.471	-2.632
	(0.081)	(1.1)	(0.057)	(1.35)
Age*Black	-0.000574	-0.038	-0.00716	-0.0695
	(0.0041)	(0.049)	(0.0031)	(0.067)
AMI*Black	-0.195	-0.261	-0.21	3.239
	(0.079)	(1.03)	(0.11)	(0.89)
Heart Failure*Black	-0.113	-1.725	-0.152	-2.519
	(0.12)	(1.49)	(0.12)	(2.91)
AMI*Female	-0.0402	-0.335	0.153	0.0991
	(0.033)	(0.67)	(0.045)	(1.27)
Heart Failure*Female	-0.205	1.548	-0.192	1.822
	(0.054)	(1.19)	(0.053)	(1.77)
Zipcode Income	-0.00000239	-0.0000194	-0.00000127	-0.00000179
	(0.0000011)	(0.0000032)	(0.00000081)	(0.0000033)
Procedure Information				
Angioplasty Only	0.534	-0.409		
	(0.054)	(0.22)		
Stent	-0.0181	-0.0475		
	(0.054)	(0.23)		
Count of Vessels	-0.108	0.201		
	(0.02)	(0.09)		
Left Circumflex Artery	-0.0558	-0.248		
	(0.03)	(0.25)		
Left Anterior Descending	0.155	0.444		
	(0.025)	(0.22)		
Right Coronary Artery	-0.0901	0.00419		
	(0.027)	(0.24)		
Diagnostic Cath	0.117	0.193		
	(0.019)	(0.055)		
Reoperation			0.932	-0.126
			(0.017)	(0.27)
Artery			0.206	2.026
			(0.072)	(0.84)
Vein			0.194	-1.05
			(0.24)	(4.57)
Artery & Vein			-0.192	-0.556
			(0.24)	(-4.59)
$Sigma_u$	0.428		0.491	
	(0.0137)		(0.010)	
Number of vessel dummies	Yes		Yes	
Year Dummies	Yes		Yes	
Pseudo R-squared	0.088		0.057	
N	1,067,018		1,019,770	

PCI Sample (Columns (1) and (2)) includes any claim involving PCI in 1996-1999, as identified from procedure codes, excluding denied and duplicate claims. CABG Sample (Columns (3) and (4)) includes any claim involving CABG in 1996-1999, as identified from procedure codes, excluding denied and duplicate claims. Results are from logit regressions of patient mortality in-hospital on the listed explanatory variables, with doctor level random effects. Coefficients are presented. Standard errors are in parentheses.

Table A.3: Effect of Controlling for Patient Sorting on Quality Measures

	ICs		CT Surgeons	
Difference in:	Rank	Measure	Rank	Measure
High Risk Patients	42.2	0.06	50.9	0.09
	(10.8)	(0.01)	(13.9)	(0.02)
Low Risk Patients	-85.5	-0.08	-63.4	-0.10
	(6.8)	(0.01)	(11.9)	(0.02)
T-statistic (difference in means)	10.0	10.2	6.2	6.9

Standard errors in parentheses. Table compares my preferred quality measures with measures constructed without controlling for patient sorting for doctors with patients in the top and bottom risk quartiles. Columns (1) and (3) give the difference in the doctor's ordinal rank (with higher quality doctors ranked lower). The difference is the ranking controlling for sorting minus the measure not controlling for sorting. Columns (2) and (4) give the difference in the measures, expressed as a fraction of one standard deviation of each measure.

Table A.4: Analysis of Dropout Behavior by Experience

	ICs			CT Surgeons		
	Young	Middle	Old	Young	Middle	Old
PCI/CABG dropout	(1)	(2)	(3)	(4)	(5)	(6)
Doctor Quality	-0.41	-0.26	-2.14	-1.92	0.04	-6.96
	(0.55)	(0.69)	(0.93)	(0.90)	(1.24)	(1.72)
Foreign School	-2.07	-4.96	-4.18	0.86	3.95	-5.01
	(1.34)	(1.32)	(2.07)	(4.33)	(4.26)	(3.90)
Prestigious School	-2.76	-2.15	-0.84	1.29	-0.55	0.51
	(1.39)	(1.60)	(2.70)	(2.54)	(3.38)	(4.32)
HRR size (beneficiaries)	-1.07	-0.38	1.63	-0.20	-1.38	0.31
	(0.67)	(0.81)	(1.12)	(1.20)	(1.47)	(2.03)
HRR size (reimbursement)	1.94	0.63	-3.56	-1.34	-0.53	-5.86
	(0.61)	(0.70)	(1.29)	(1.23)	(1.64)	(1.96)
N	1435	1395	1157	884	797	696
Dependent variable mean (%)	6.2	7.2	13.2	9.6	15.8	26.7
Pseudo R-squared	0.018	0.016	0.023	0.011	0.0038	0.039
Medicare dropout						
Doctor Quality	-0.07	-0.10	-0.71	-0.38	0.58	-7.25
	(0.25)	(0.33)	(0.50)	(0.73)	(0.98)	(1.43)
Foreign School	0.81		-0.38	1.74	1.40	-0.94
	(0.82)		(1.30)	(3.78)	(3.14)	(3.57)
Prestigious School	-0.13	0.15	0.28	2.36	-1.21	-1.74
	(0.69)	(0.92)	(1.61)	(2.09)	(2.45)	(3.71)
HRR size (beneficiaries)	0.43	0.29	0.59	-0.09	-1.23	0.37
	(0.25)	(0.38)	(0.65)	(0.87)	(1.20)	(1.84)
HRR size (reimbursement)	-0.04	0.01	-1.52	-1.09	-1.36	-5.37
	(0.27)	(0.38)	(0.70)	(0.89)	(1.20)	(1.88)
N	1435	1090	1157	884	797	696
Dependent variable mean (%)	1.1	1.1	4.0	5.2	8.7	20.5
Pseudo R-squared	0.026	0.0057	0.019	0.011	0.011	0.051

Table displays results from logit regressions of dropout indicators on quality measures and doctor characteristics by experience cohort. Cohorts were defined to equalize samples across groups: For ICs young cohorts graduated medical school from 1983-1989, middle cohorts from 1977-1982, and older cohorts from 1967-1976 (ICs near retirement (65 or older in 2005) are excluded). For CT surgeons, young cohorts graduated medical school from 1982-1989, middle cohorts from 1975-1981, and older cohorts from 1967-1974 (CT surgeons 65 years or older in 2005 are excluded). The top panel is for regressions using indicators for dropping out of PCI/CABG, and the bottom panel uses indicators for stopping billing Medicare altogether. Effects are for a one standard deviation change in X on the probability of dropping out (in percentage points). For dummy variables the effect is for a discrete change of from 0 to 1. Robust standard errors are in parentheses. Foreign school is dropped from Medicare dropout regressions for middle cohorts due to perfectly predicting failure.

Table A.5: Analysis of Moving Behavior by Experience

	ICs			CT Surgeons		
	Young	Middle	Old	Young	Middle	Old
Change in zipcode	(1)	(2)	(3)	(4)	(5)	(6)
Doctor Quality	-0.63	-2.51	-0.70	-5.50	-1.18	-1.32
	(1.35)	(1.33)	(1.47)	(1.75)	(1.75)	(1.76)
Foreign School	9.08	2.08	-8.66	-4.86	1.21	9.42
	(3.37)	(3.33)	(3.35)	(7.33)	(5.49)	(4.54)
Prestigious School	1.35	-4.52	-3.90	-6.05	-1.62	3.27
	(3.69)	(3.87)	(4.17)	(4.24)	(4.60)	(4.86)
HRR size (beneficiaries)	0.10	0.57	1.76	3.87	0.94	4.39
	(1.51)	(1.54)	(1.67)	(1.93)	(2.04)	(2.18)
HRR size (reimbursement)	-0.11	-0.10	-0.31	-3.26	1.53	-2.09
	(1.58)	(1.47)	(1.74)	(1.92)	(2.04)	(2.20)
N	1435	1395	1157	884	797	696
Dependent variable mean (%)	53.4	43.2	39.8	59.4	43.7	36.2
pseudo R-squared	0.0037	0.0030	0.0054	0.014	0.0020	0.010
Change in HRR						
Doctor Quality	-0.15	-1.07	-0.95	-6.31	-2.80	-1.18
	(1.09)	(0.86)	(0.78)	(1.75)	(1.42)	(1.23)
Foreign School	11.80	3.19	-3.73	0.57	-0.39	3.30
	(3.22)	(2.50)	(1.78)	(7.01)	(4.53)	(3.34)
Prestigious School	4.49	-0.20	-1.00	-0.58	2.84	-1.69
	(3.47)	(2.82)	(2.22)	(4.27)	(3.95)	(3.30)
HRR size (beneficiaries)	-3.75	-1.33	-1.16	-2.21	-3.76	1.13
	(1.33)	(1.11)	(0.94)	(2.01)	(1.71)	(1.58)
HRR size (reimbursement)	-0.06	-1.18	1.96	0.89	3.74	-2.34
	(1.40)	(1.03)	(1.03)	(1.92)	(1.62)	(1.70)
N	1415	1370	1126	871	788	684
Dependent variable mean (%)	24.7	14.4	9.0	40.4	21.6	13.3
pseudo R-squared	0.017	0.0066	0.013	0.013	0.013	0.0084

Table displays results from logit regressions of moving indicators on quality measures and doctor characteristics by experience cohort. Cohorts were defined to equalize samples across groups: for ICs young cohorts graduated medical school from 1983-1989, middle cohorts from 1977-1982, and older cohorts from 1967-1976 (ICs near retirement (65 or older in 2005) are excluded). For CT surgeons, young cohorts graduated medical school from 1982-1989, middle cohorts from 1975-1981, and older cohorts from 1967-1974 (CT surgeons 65 years or older in 2005 are excluded). The top panel is for regressions using indicators for changing zip code, and the bottom panel uses indicators for changing HRR. Effects are for a one standard deviation change in X on the probability of dropping out (in percentage points). For dummy variables the effect is for a discrete change of from 0 to 1. Robust standard errors are in parentheses.

Table A.6: Analysis of Dropout Behavior within HRRs

PCI/CABG dropout	ICs (1)	CT Surgeons (2)
Doctor quality	-1.14	-3.12
	(0.57)	(1.18)
Foreign school	-3.28	5.30
	(1.28)	(3.26)
Prestigious school	0.15	4.38
	(1.55)	(3.21)
HRR fixed effects	Yes	Yes
N	2490	1319
Pseudo R-squared	0.06	0.06
Medicare dropout		
Doctor quality	-0.72	-3.01
	(0.43)	(1.02)
Foreign school	-0.42	5.55
	(1.13)	(2.85)
Prestigious school	1.96	3.21
	(1.52)	(2.78)
HRR fixed effects	Yes	Yes
N	1757	1319
Pseudo R-squared	0.05	0.06

Table displays results from logit regressions of dropout indicators on quality measures and doctor characteristics with HRR fixed effects. The top panel is for regressions using indicators for dropping out of PCI or CABG, and the bottom panel uses indicators for stopping billing Medicare altogether. Column (1) displays results for ICs; column (2) for CT surgeons. Effects are for a one standard deviation change in X on the probability of dropping out (in percentage points). For dummy variables the effect is for a discrete change of from 0 to 1. Robust standard errors are in parentheses.

Table A.7: Analysis of Moving Behavior within HRRs

Change in zipcode	ICs (1)	CT surgeons (2)
Doctor Quality	-3.78	-1.52
	(1.11)	(1.55)
Foreign School	-1.74	-7.13
	(2.66)	(4.09)
Prestigious School	0.101	-5.14
	(3.09)	(3.83)
HRR fixed effects	Yes	Yes
N	2,589	1,336
Pseudo R-squared	0.079	0.087
Change in HRR		
Doctor Quality	-0.49	-3.39
	(0.69)	(1.22)
Foreign School	0.233	-10.4
	(1.85)	(2.53)
Prestigious School	4.41	-3.88
	(2.38)	(2.83)
HRR fixed effects	Yes	Yes
N	2,419	1,317
Pseudo R-squared	0.056	0.072

Table displays results from logit regressions of dropout indicators on quality measures and doctor characteristics. The top panel is for regressions using indicators for changing zipcode, and the bottom panel uses indicators for changing HRR. Column (1) displays results for ICs; column (2) for CT surgeons. Effects are for a one standard deviation change in X on the probability of dropping out (in percentage points). For dummy variables the effect is for a discrete change of from 0 to 1. Robust standard errors are in parentheses.

Table A.8: Analysis of Claim Volumes - Additional Measures

	ICs			CT Surgeons		
Patients	(1)	(2)	(3)	(4)	(5)	(6)
Quality	-1.14	-0.66	-1.02	6.05	6.40	5.79
	(1.85)	(1.85)	(2.17)	(3.97)	(4.07)	(7.75)
Quality * Experience	0.06	0.03	0.00	-0.31	-0.33	-0.45
	(0.12)	(0.12)	(0.14)	(0.23)	(0.25)	(0.49)
Experience	5.52	8.06	7.28	5.38	6.47	5.44
	(0.37)	(0.34)	(0.35)	(0.56)	(0.55)	(0.83)
Experience squared	-0.13	-0.15	-0.10	-0.14	-0.22	-0.14
	(0.011)	(0.011)	(0.011)	(0.01)	(0.01)	(0.02)
Cohort fixed effects		Yes	Yes		Yes	Yes
N	42,280	42,280	31,400	28,140	28,140	16,180
R-squared	0.059	0.076	0.087	0.029	0.040	0.044
Allowed charges						
Quality	-0.76	0.03	0.51	12.73	12.31	8.51
	(2.10)	(2.09)	(2.43)	(2.45)	(2.45)	(4.04)
Quality * Experience	0.15	0.11	0.05	-0.33	-0.30	-0.21
	(0.13)	(0.13)	(0.16)	(0.12)	(0.12)	(0.22)
Experience	6.94	10.66	9.92	4.83	3.22	2.02
	(0.36)	(0.39)	(0.46)	(0.38)	(0.43)	(0.55)
Experience squared	-0.17	-0.16	-0.11	-0.15	-0.23	-0.17
	(0.010)	(0.011)	(0.013)	(0.010)	(0.011)	(0.013)
Cohort fixed effects		Yes	Yes		Yes	Yes
N	42,280	42,280	31,400	28,140	28,140	16,180
R-squared	0.059	0.095	0.10	0.081	0.12	0.081
Patient risk index ($x\hat{\beta}$)						
Quality	2.9	3.5	1.9	-1.5	-1.2	1.2
	(2.14)	(2.10)	(1.11)	(2.5)	(2.4)	(2.5)
Quality * Experience	-0.28	-0.33	-0.12	-0.025	-0.056	-0.048
	(0.15)	(0.15)	(0.061)	(0.14)	(0.14)	(0.11)
Experience	-7.0	-10.6	-5.0	3.02	12.66	5.99
	(0.62)	(0.80)	(0.47)	(0.15)	(0.74)	(0.22)
Experience squared	0.23	0.42	0.10	-0.0015	0.0017	0.0013
	(0.020)	(0.026)	(0.013)	(0.00010)	(0.018)	(0.0016)
Cohort fixed effects		Yes	Yes		Yes	Yes
N	42,280	42,280	31,400	28,140	28,140	16,180
R-squared	0.035	0.052	0.038	0.069	0.19	0.13

Table displays results from OLS regressions of claim volumes on quality, experience and a quality-experience interaction. All regressions include the usual physician characteristics. In the top panel the outcome variable is unique patients, in the middle panel it is charges allowed by CMS, in the bottom panel it is the patient risk index, which is equal to the doctor-level mean of predicted patient mortality in-hospital, where the prediction uses coefficients from Appendix Table A.2s. Columns (1), (2) and (3) display results for ICs; the remaining columns display results for CT surgeons. Columns (1), (2), (4) and (5) display results for the full sample; Columns (3) and (6) exclude dropouts and movers. Quality effects are for a one standard deviation change in quality. Coefficients are expressed as a percentage of the outcome variable in the top two panels, in the bottom panel coefficients are expressed as a percentage of one standard deviation in the patient risk index. Standard errors, in parentheses, are clustered at the doctor level.

Table A.9: Analysis of Claim Volumes by Experience

	ICs			CT Surgeons		
	Young	Middle	Old	Young	Middle	Old
Claims	(1)	(2)	(3)	(4)	(5)	(6)
Quality	1.93	0.92	7.06	-3.43	-6.71	-4.51
	(2.92)	(5.07)	(11.6)	(5.71)	(8.88)	(16.10)
Quality * Experience	-0.22	-0.15	-0.28	0.16	0.50	-0.10
	(0.37)	(0.36)	(0.44)	(0.63)	(0.54)	(0.67)
Experience	11.71	7.09	9.59	12.43	-2.86	6.11
	(1.09)	(1.38)	(1.10)	(1.74)	(2.58)	(7.51)
Experience squared	-0.21	-0.039	-0.12	-0.38	0.15	-0.10
	(0.056)	(0.047)	(0.023)	(0.086)	(0.079)	(0.15)
N	10,040	10,950	10,410	4,790	5,280	4,340
Dependent variable mean	4,956	5,650	5,971	521	566	539
R-squared	0.13	0.088	0.082	0.086	0.027	0.022
PCI/CABG Claims						
Quality	-0.79	3.59	16.3	3.99	10.6	28.5
	(3.72)	(6.68)	(13.5)	(5.40)	(7.77)	(-12.3)
Quality * Experience	0.43	0.286	-0.40	0.27	0.038	-0.66
	(0.44)	(0.52)	(0.51)	(0.49)	(0.42)	(0.47)
Experience	12.5	8.76	10.8	9.8	-4.96	-4.1
	(1.19)	(1.71)	(1.60)	(1.56)	(1.94)	(3.20)
Experience squared	-0.41	-0.18	-0.20	-0.67	-0.013	-0.044
	(0.062)	(0.057)	(0.033)	(0.076)	(0.057)	(0.066)
N	10,040	10,950	10,410	4,790	5,280	4,340
Dependent variable mean	64.7	63.1	62.0	80.7	81.6	77.0
R-squared	0.057	0.033	0.029	0.060	0.100	0.12
Non-emergency PCI/CABG Claims						
Quality	-3.93	13.7	43.4	8.77	9.1	12.9
	(7.88)	(18.7)	-29.0	(7.76)	(14.5)	(23.5)
Quality * Experience	1.15	0.062	-1.13	-0.09	0.19	0.05
	(0.89)	(1.40)	(1.06)	(0.69)	(0.78)	(0.90)
Experience	16.7	10.9	18.2	9.7	4.9	9.9
	(2.41)	(4.27)	(3.45)	(2.20)	(3.93)	(5.87)
Experience squared	-0.47	-0.16	-0.31	-0.54	-0.24	-0.29
	(0.12)	(0.14)	(0.068)	(0.10)	(0.11)	(0.12)
N	9,036	9,855	9,369	4,311	4,752	3,906
Dependent variable mean	19.0	18.9	19.5	30.4	30.4	28.1
R-squared	0.036	0.032	0.034	0.038	0.040	0.050

Table displays results from OLS regressions of claim volumes on quality, experience and a quality-experience interaction. All regressions include the usual physician characteristics and cohort fixed effects. In the top panel the outcome variable is Medicare claims, in the middle panel it is PCI/CABG claims, in the bottom panel it is non-emergency PCI/CABG claims. Non-emergency PCI/CABG regressions exclude observations from 1996, as these cannot be linked to inpatient records. Columns (1), (2) and (3) display results for ICs; the remaining columns display results for CT surgeons. Columns (1) and (4) display results for young ICs/CTs; Columns (2) and (4) for middle cohorts; and Columns (3) and (6) for older cohorts. Cohorts were defined to equalize samples across groups: young ICs graduated medical school from 1983-1989, middle cohorts from 1977-1982, and older cohorts from 1967-1976. ICs near retirement (65 or older in 2005) as well as dropouts and movers are excluded. For CT surgeons, young cohorts graduated medical school from 1982-1989, middle cohorts from 1975-1981, and older cohorts from 1967-1974 (CT surgeons 65 years or older in 2005 as well as dropouts and movers are excluded). Quality effects are for a one standard deviation change in quality expressed as a percentage of the outcome variable. Experience effects are dy/dx as a percentage of the outcome variable. Standard errors, in parentheses, are clustered at the doctor level.