NBER WORKING PAPER SERIES

DO BAD REPORT CARDS HAVE CONSEQUENCES? IMPACTS OF PUBLICLY REPORTED PROVIDER QUALITY INFORMATION ON THE CABG MARKET IN PENNSYLVANIA

Justin Wang
Jason Hockenberry
Shin-Yi Chou
Muzhe Yang

Working Paper 16225
http://www.nber.org/papers/w16225

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2010

We would like to thank the Martindale Center at Lehigh University for their generous support of this research. We would like to thank Mary Beth Deily, Thomas Hyclak and participants at the Lehigh University Economics seminar for helpful comments and suggestions. Disclaimer: The Pennsylvania Health Care Cost Containment Council (PHC4) is an independent state agency responsible for addressing the problem of escalating health costs, ensuring the quality of health care, and increasing access to health care for all citizens regardless of ability to pay. PHC4 has provided data to this entity in an effort to further PHC4's mission of educating the public and containing health care costs in Pennsylvania. PHC4, its agents, and staff, have made no representation, guarantee, or warranty, express or implied, that the data -- financial, patient, payor, and physician specific information -- provided to this entity, are error-free, or that the use of the data will avoid differences of opinion or interpretation. This analysis was not prepared by PHC4. This analysis was done by Justin Wang, Jason Hockenberry, Shin-Yi Chou and Muzhe Yang. PHC4, its agents and staff, bear no responsibility or liability for the results of the analysis, which are solely the opinion of the authors. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Do Bad Report Cards Have Consequences? Impacts of Publicly Reported Provider Quality Information on the CABG Market in Pennsylvania
Justin Wang, Jason Hockenberry, Shin-Yi Chou, and Muzhe Yang
NBER Working Paper No. 16225
July 2010
JEL No. I1,I11,I18

ABSTRACT

Since 1992, the Pennsylvania Health Care Cost Containment Council (PHC4) has published cardiac care report cards for coronary artery bypass graft (CABG) surgery providers. We examine the impact of CABG report cards on a provider's aggregate volume and volume by patient severity and then employ a mixed logit model to investigate the matching between patients and providers. We find a reduction in volume of poor performing and unrated surgeons' volume but no effect on more highly rated surgeons or hospitals of any rating. We also find that the probability that patients, regardless of severity of illness, receive CABG surgery from low-performing surgeons is significantly lower.

Justin Wang
School of Business
Worcester Polytechnic Institute
100 Institute Road, Worcester MA 01609
jwang@wpi.edu

Jason Hockenberry
Department of Health Management & Policy
College of Public Health
University of Iowa
200 Hawkins Drive, E206GH
Iowa City, IA 52242
and NBER
jason-hockenberry@uiowa.edu

Shin-Yi Chou
Department of Economics
College of Business and Economics
Lehigh University
621 Taylor Street
Bethlehem, PA 18015-3117
and NBER
syc2@lehigh.edu

Muzhe Yang
Department of Economics
Lehigh University
621 Taylor Street
Bethlehem, PA 18015
Tel: (610) 758-4962
Fax: (610) 758-4677
muzheyang@lehigh.edu

1 Introduction

Public disclosure of the performance of health care providers (e.g. hospitals and physicians) – often referred to as report cards – has been increasing since the early 1990's. In many states, these report cards rate providers on the performance of a particular procedure, most often reporting whether they had high, normal, or low risk adjusted mortality rates relative to the expected rates given the health characteristics of the patient population.[1] The agencies and regulators responsible for the provision of this public information often claim improved quality and efficiency in the provision of care as the main goal of providing this information.[2] Despite the economic undertone of this claim and the high level of attention paid by the medical and public health literature to this phenomenon,[3] only recently has the public disclosure of the performance of health care providers become the subject of published research among economists (Cutler, Huckman and Landrum 2004; Dranove, McClellan and Satterwaite 2003; Dranove and Sfekas 2008; Bundorf, Chun, Goda and Kessler 2009).[4]

Because of the highly specialized nature of health care, there exists an asymmetry of information related to the quality of the health care provider (Arrow 1963). By providing public information on the relative performance of providers, a regulator or public agency can increase the ability of patients, referring physicians, and health insurance plan providers to distinguish and select providers of high quality, thereby improving the health outcomes of those using the services provided in the health care market. This information may also ensure accountability of

---

[1] For example, Shearer and Cronin (2005) alone reviewed 51 online hospital performance reporting.
[2] As recommended by the Committee on Quality of Health Care in America, Institute of Medicine (2001), for example, "making the information describing the health care system's performance, evidence-based practice and patient satisfaction" transparent is an important step as we move to the 21st-century health care system.
[3] See detailed reviews in Epstein (2006) and Marshall et al. (2000).
[4] The impact of public disclosure of health plan performance has received as much, if not more, attention in recent economics literature (Beaulieu 2002; Dafny and Dranove 2008; Jin and Sorensen 2005; Chernew, Gowrisankaran and Scanlon 2008; Wedig and Tai-Seale 2002).

health care providers in the promotion of quality improvements and has the potential to increase efficiency and increase total welfare.

Welfare increases may not occur uniformly across different patient subpopulations as a result of public reporting. In fact, they may not even increase at all. In health care, particularly surgery which sometimes needs to be consumed with little advanced notice, the quality of and distance to the provider become the main choice variables. Those with more severe disease or in emergent condition may not have the time or ability to access the public information. Because those with less severe non-emergent disease can use the information to select the best providers, this could leave them unavailable to treat the patients who would arguably benefit more from the comparatively better performance of these surgeons. Thus report cards could increase market concentration for the particular procedure in a way that makes it more difficult for the neediest patients to access better providers.

Another effect public disclosure has on public welfare is related to the incentive this information gives providers to game the system. While relatively healthier patients may sort to better providers, this sorting may be further exacerbated by providers' efforts to improve or maintain quality ratings. Report cards might lead to patient selection or patient dumping, phenomena which have previously been documented (Ellis 1998; Dranove et al. 2003). In practice, providers often claim that the risk adjustment measures used in the calculations for the report cards fail to capture all the necessary information.[5] If providers are convinced that sicker patients raise the probability of having a poor rating and are convinced that poor ratings will lead to a decrease in demand for their service or potential sanctions by the regulator, they may select

[5] As a condition of reporting data to the state agency in Pennsylvania, hospitals are allowed to respond to the release of report cards in a public fashion with the results being posted on the website where the report cards are found. It appears as though hospitals ranked as poor performers will often point out that there is some facet of the system that fails to account for the poor health of their particular patients.

only the relatively healthy patients for the procedures which have publicly reported performance measures to reduce their likelihood of receiving poor rankings. This behavior would also constrain sicker patients' access to better providers, assuming the better providers are the ones able to select the risk profile of their patients.

Much of the previous literature also ignores the interaction between hospital and surgeon report cards and fails to account for hospital proximity to patient's residence.[6] Low performing hospitals may have surgeons who are rated proficient or high performing and vice versa. For those who are quite ill, hospital quality may not be as important as proximity to one's residence. If one expects to be hospitalized for a long period of time, proximity to family support networks may be important, and being farther away from home would raise the cost of this support.[7] In emergent cases this distance-to-hospital issue is even more salient, and report cards would be an even more minor factor in the hospital choice decision. However, in both cases the selection of surgeon might still be based on report cards conditional on arriving at a given hospital. In either case it is likely not the patient using the information first hand, rather a cardiologist (in the non-emergent cases) or the attending physician (in emergent cases) who use the report card information for referral purposes. Indeed, for hospitals receiving emergent cases, it is in their interest that these cases be steered toward better performing surgeons, as the hospital report card is composed of operations performed by each of the surgeries performed in that facility.

In this paper, we use more recent data (1998-2005) from Pennsylvania along with the publication of surgeon and hospital coronary artery bypass graft (CABG) performance data,

---

[6] Kessler (2005) proposed an alternative approach to rank hospitals on the basis of the travel distances of their Medicare patients. This hypothetical distance-based report card is more powerful than the outcome-based report card to distinguish high-mortality hospitals from the average hospital, but less powerful at distinguishing low-mortality hospitals from the average hospital.

[7] If one elects to get treated at a hospital further from their home then relatives and friends involved in supporting them during the procedure and recovery would need to travel further, potentially rent hotel rooms, take even more additional time off work than they otherwise would if the facility was local, etc.

collected by an independent state agency, the Pennsylvania Health Care Cost Containment Council (PHC4), to address two questions about the effects of quality reporting. First, we examine the effects report cards have on the patient volumes of hospitals and cardiac surgeons separately. We aggregate the patient-level data into hospital-quarter or surgeon-quarter, and use hospital or surgeon fixed effects to remove unobserved variations that are correlated with both ratings and volumes. The expectation is that patients', referring physicians', and/or health insurance plan providers' response to the information provided in the report cards will be observable through changes in patient volumes upon release of new information.

On the demand side, demand could decrease for poor performing surgeons, increase for high performing surgeons, or some combination thereof. On the supply side, poor-performing surgeons could change their patient selection decisions after learning their relative performance, perhaps voluntarily or as the result of administrative changes (Epstein 2006). The equilibrium volume will change in response to changes in demand-side factors, supply-side factors, or both. Moreover, we expect a surgeon's volume to be more responsive to the reported rating than hospital volume, because there are more surgeons in the market and the demand for surgeons' services will be more elastic than the demand for hospital services.[8] Our empirical results confirm this expectation.

Second, we examine the matching between patients and surgeons. If it is more beneficial for sicker patients to receive treatment from high-quality surgeons and if high-quality surgeons are less likely to shun sicker patients, then report cards will improve the social welfare by promoting better matching between patients and surgeons. We first follow Cutler et al. (2004)

[8] Again this is because conditional upon going to a particular facility the cost of choosing a particular surgeon that is at that hospital is only the cost of information, which is likely to be acquired and processed by the rest of the staff who can 'steer' patients toward better surgeons. There is no evidence of excessive wait times for CABG in PA during this period. There was an overall declining demand for CABG due to the diffusion of PTCA as a substitute revascularization procedure (Cutler and Huckman, 2003).

and examine surgical volumes at hospital or surgeon level by patient severity , so we can examine whether report cards have heterogeneous impacts on patients of different severity ratings.

However, there are two empirical problems which arise in this type of aggregate level analysis. The primary concern is that aggregate level analysis does not take patient heterogeneity into consideration. A secondary empirical concern, particularly for the surgeon-level analysis, is that the data constitute an unbalanced panel, which itself is probably due to non-random factors including exit of surgeons attributable to poor ratings. To address these concerns, we use a mixed logit model at the individual patient level to examine responses of patients with different severity ratings and emergent status to low-performing surgeons.

Overall, we find that surgical volume is negatively associated with surgeons' receiving a poor rating or being unrated, regardless of patient severity. We also find that the probability that patients, regardless of severity of illness, receive treatment from low-performing surgeons is significantly lower. The implication of these findings is that patients or referring physicians are aware of the report card publication and use the information to select high-quality providers, or alternatively stated, avoid low-performing or unrated providers. However, volumes of both types of patients respond similarly, which suggests report cards did not lead to an improvement in the matching process and excessively high demand for top surgeons, who it might be argued that should be treating the cases requiring the most skill (i.e. the most severely ill), could lead to a crowding issue.

2 Institutional Background and Previous Literature

2.1 Institutional Background

Health outcomes report cards are one mechanism by which health care provider quality information is disseminated to the public. Health outcomes report cards usually provide information related to adverse health outcomes, such as mortality rates and complications rates, at the provider- or plan-level, often for a specific procedure or treatment of a specific disease. Among the provider-level, procedure-specific category, CABG report cards are the most well-established, and New York and Pennsylvania were the first two states to make these publicly available (in 1990 and 1992, respectively). Although most states collect health care services data, to our knowledge only California, Massachusetts, New Jersey, and Virginia followed in the footsteps of New York and Pennsylvania to create similar reporting systems to publicly disclose the performance of CABG providers.

The Pennsylvania's Guide to Coronary Artery Bypass Graft Surgery (i.e., report card) is published by the Pennsylvania Health Care Cost Containment Council (PHC4). The report cards publicly disclose the aggregate health outcomes of those undergoing CABG at the hospital- and surgeon-level. Since 1992, PHC4 has published ten CABG report cards (for the years 1990, 1991, 1992, 1993, 1994/1995, 2000, 2002, 2003, 2004, and 2005).[9] Prior to 1998, the report cards were distributed to hospitals, surgeons, public libraries, business groups, legislature, the media, and any individual who requested them (Schneider and Epstein 1998). Beginning in 1998, PHC4 posted the 1994/95 and all subsequent report cards on the agency web site, making reports

[9] Reports cards are available at http://www.phc4.org. The report cards, with the exception of those for the years 1990, 1991, 1992 and 1993, are available at PHC4's website.

more accessible to health care consumers and their physicians. Report card data collection and publication dates are summarized in Table 1.

Providers are only given a report card if they meet the minimum volume threshold for the rating year, which is at least 30 CABG procedures performed. The report cards rate CABG providers (hospitals and surgeons) on four outcomes: in-hospital mortality, 30-day mortality, 7-day readmission and 30-day readmission. To arrive at the rating for each provider, PHC4 first constructs a 95% confidence interval for the expected risk-adjusted rate for each of the four outcomes, and compares the actual outcomes of each provider to the 95% confidence interval. Each hospital and surgeon receives one of three possible ratings: lower than expected, same-as-expected, or higher than expected in each of the four categories. The report cards also publish the average post-surgical length of stay for the patients of each hospital and each surgeon.

2.2 Literature Review

Over the last two decades, CABG mortality rates have been declining nationwide and research indicates public dissemination of provider report cards has accelerated this trend.[10] Epstein (2006) outlined three possible mechanisms through which CABG provider report cards would lead to reductions in mortality: changes in the population of patients, changes in the population of CABG providers, and better matching between patients and providers. However, empirical results are mixed on the mechanisms through which report cards improve patients' health outcomes.

New York has the longest history of publishing outcome information related to CABG, which is done through the New York Cardiac Surgeon Reporting System (CSRS). As a result, a large proportion of the empirical literature on the subject of report cards employs these data. The

[10] See Hannan et al. (1994) and Peterson et al. (1998).

CSRS is similar to the reporting system in Pennsylvania, disclosing health outcomes and provider ratings at the hospital- and surgeon-level. Using New York data from 1991 to 1999, Cutler et al. (2004) found that hospitals identified by a public report card as having high-mortality in the past 12 months experience a 10% decline in monthly CABG volume, but found no evidence that report cards had a significant increase on volume for low-mortality (high performing) hospitals. By examining patient characteristics, the authors determined that most of the decline in volume of hospitals receiving poor grades was due to the loss of relatively healthy (low-severity) patients. They attribute this to the healthier patients having lower search costs and having the ability to afford avoiding low-quality hospitals. The main limitation of this study is that it was focused on hospital-level analysis. As such, the model controlled only hospital fixed effects and year fixed effects and did not address surgeon-level effects.

Using New York Data from 1990 to 1993, Mukamel and Mushlin (1998) found that hospitals and surgeons with lower reported mortality rates experienced higher rates of annual growth in Medicare market shares. They also found that surgeons with lower reported mortality rates had higher rates of annual growth in prices charged submitted to Medicare for CABG surgery. Using New York Data from 1992 to 1995 Romano and Zhou (2004) found that hospitals with low mortality had a 22% increase in CABG volume in the first month after report cards were released, whereas hospitals with high mortality had a 16% decrease in CABG volume in the second month after report cards were released.

Other studies suggest that CABG report cards change the population of patients and improve the matching between elderly patients and providers. Schneider and Epstein (1996) randomly surveyed 50 percent of cardiologists and cardiac surgeons in Pennsylvania in 1995. They reported that 59 percent of cardiologists found it more difficult to place patients who were severely ill and required CABG, and that 63 percent of cardiac surgeons were less willing to perform CABG on severely ill patients. Using Medicare claim data from 1987 to 1994,[11] Dranove et al. (2003) found that the illness severity of elderly CABG patients in NY and PA declined compared to states which had not introduced public report cards before or during this period.[12] Their study provides evidence of selection behavior of CABG providers and suggests that, at least in the short term, the population of elderly CABG patients changed for Medicare beneficiaries. They also found that teaching hospitals in NY and PA experienced a greater share of severely ill patients, and that report cards led to delays of getting treatments for both healthy and sick patients. The authors conclude that report card publication led to better matching between patients and providers because teaching status is an indication of quality and the process of better matching is likely to take time.

However, the authors' argument that report cards led to better matching between patients and providers is based on inference rather than direct testing by incorporation of the information provided by the report cards, such as providers' ratings, into the model. In order to fully understand the impact of the report cards on the market equilibrium, one would need to examine the within-state impacts of the report cards of individual hospitals and surgeons. It also would be useful to incorporate non-Medicare and Medicaid patients, as they are usually younger, into the analysis. The reason for this is that if younger patients have a different likelihood of being aware

[11] During the period, only NY and PA have CABG report cards. NY's first CABG report card was published in 1991 and PA's first CABG report card was published in 1992.

[12] The illness severity measures used are the mean of patients' total hospital expenditures one year prior to admission and the mean of patients' total days in hospital one year prior to admission. While there is some correlation between recent spending and latent health status, there are a moderate number of very serious coronary artery disease cases that are managed and treated on an outpatient basis or not detected by health providers at all until sudden onset of AMI. Also, those who have had more recent hospital stays potentially could have a higher rate of hospital resource use and even improved their health status by using this resource prior to having CABG relative to resource non-users, further complicating the use of this as a measure of latent health status at the time of the decision of whether a patient will undergo CABG.

or using the report cards, or if the search and treatment costs are systematically different, then publicly provided information may have a differential impact on the younger population.

Overall, the literature on the impacts of report cards focuses generally on the effects of the information in the early 1990's and most often only deals with the effect they have immediately after introduction. The ability of patients, their primary providers, and their family to access, understand and effectively employ report cards in their decision making has changed drastically since the early 1990's, given the rapid diffusion of internet use during this period and that the reports in PA were not specifically published on the internet until 1998. Moreover, the diffusion of Percutaneous Coronary Intervention (PCI) as a substitute for CABG occurred quite rapidly in the mid 1990's, which may have changed the overall patient population receiving CABG (Cutler and Huckman 2003). Further investigation into the impacts of report cards in more recent periods is warranted.

3 Data and Sample

3.1 Data

We employed four different datasets in this study. The primary data are the Pennsylvania Inpatient Hospital Discharge Data collected by PHC4. In order to maintain consistency of this administrative dataset and to meet state requirements, Pennsylvania general acute care hospitals are required to use a uniform claims and billing form (UB92) to submit their data. This dataset contains very rich clinical and utilization information at the patient-level. Data elements include patients' race/ethnicity, gender, age, zip code of residence, severity of illness, insurance type, the type of admission, the quarter of admission, the principal diagnosis code and secondary diagnoses codes, the principal procedure code and secondary procedure codes, discharge status, a four-digit unique facility identifier, and the license number of the operating physician.

Pennsylvania hospitals use the computerized system MedisGroups to calculate the severity measure. This measure is calculated using clinical variables such as physician examinations, radiology findings, laboratory findings, and pathology findings.[13] Therefore, this measure is an independent proxy for patient severity upon admission. The patient severity is a score from 0 to 4. A higher score indicates a greater likelihood of in-hospital death (Iezzoni and Moskowitz 1988). The average severity of illness of patients in our sample is 1.5. Thus, we define low severity as a score of 0 or 1, and high severity as a score of 2, 3 or 4. Procedure codes, identified by International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) codes, enable us to define our CABG sample. The unique facility identifier enables us to identify at which hospital the patient underwent CABG surgery. The license number of the operating physician allows us to identify the surgeon who performed the operation on the patient.

The second dataset we use in our study is the Pennsylvania's Guide to Coronary Artery Bypass Graft Surgery, which we refer to as report cards. Because in-hospital mortality is a general indicator of surgical outcomes and PHC4 has consistently reported the in-hospital mortality ratings at hospital- and surgeon-level, we use provider's in-hospital mortality ratings as a quality measure. For each patient that underwent CABG surgery in our study period, we match the in-hospital mortality rating in the most recent report card of each patient's hospital and surgeon to the patient's individual-level data. The report card publication date and each patient's admission year and quarter allow us to identify the most recent rating for each patient's providers.

[13] See Iezzoni et al. (1996) for more detailed descriptions.

Our third data source is the web site of the Pennsylvania Department of State Bureau of Professional and Occupational Affairs. This web site allows us to extract the name and license issue date for CABG surgeons who have performed surgery in PA. The primary dataset contains only the surgeon's license number, and the report card lists only the surgeon's name, so this data set allows us to link a surgeon's license number with his/her name and thus combine the first two datasets. Additionally, we are able to construct the surgeon's experience given the surgeon's license issue date, and use the patient's admission year to control for surgeon heterogeneity in our models.[14] Finally, hospital characteristics were taken from our fourth data source, the American Hospital Association's Annual Survey of Hospitals.

3.2 Sample

Our study sample includes Pennsylvania residents (aged 30 and above) who were undergoing an isolated CABG procedure (CABG surgery with no other major heart surgery during the same admission)[15] in Pennsylvania hospitals and who were admitted between the third quarter of 1998 and the first quarter of 2006 (N=127,285). We drop patients whose admitting hospital performed fewer than 30 CABG surgeries in the reporting year and were thus not rated in the most recent report cards (N=12,099). Our final sample consists of 114,039 CABG patients without missing values, and 84,235 CABG patients if unrated surgeons are excluded. The location of hospitals providing CABG in our sample is shown in Figure 1.

There are two important reasons that we focus on the report cards that were published after 1998. First, report card information was made available online after the 1998 publication.

[14] There is some assumption that the license date corresponds with the beginning of the surgeon's career. This may not hold for all surgeons, particularly if they have moved from outside the mid-Atlantic region, as surgeons not practicing in states immediately adjacent to PA would not have reason to hold a PA license. Thus, our experience variables contain some measurement errors.

[15] For example, as discussed in Shahian et al. (2001), additional procedures are given at the same time as the CABG surgery, such as mitral valve repair, atrial septal defect repair or left ventricular aneurysmectomy, these cases are not classified as isolated CABG's and are not included in reporting.

Most previous studies found that earlier report card publications in Pennsylvania had no impact on hospital volumes. We expect that the availability of the report card information online may increase their impact on patients, referring physicians, health plan providers, hospitals, and cardiac surgeons. Second, the Certificate of Need (CON) regulation for cardiac care was terminated in 1996. In the three years prior to the termination of CON in Pennsylvania, 1994-1996, the number of open-heart surgery programs was stable (from 43 to 44). In contrast, the number of programs climbed from 44 to 55 in the three years following CON termination (Robinson, Nash, Moxey and O'Connor 2001). Although the CON had no significant impact on inpatient mortality rates for CABG (Robinson et al. 2001), it did decrease the average procedure volume for CABG, and the average cost per CABG patient (Ho 2006). These trends became relatively more stable after 1998. Thus, we concentrate on the four most recent report cards to eliminate the confounding factors due to movement to online provision and the repeal of CON regulation.

In Table 2, we list the number of hospitals and surgeons by ratings across four report card episodes, as well as the sample years that are attached to each report card. Overall, we have more CABG hospitals and fewer CABG surgeons over time. The number of hospitals with high mortality flags fluctuates while the number of surgeons with high mortality flags decreases between 1994 and 2003. The reduction of surgeons is largely due to the reduction of unrated surgeons. Between reports of 1994 and 2003, the number of CABG surgeons decreased from 518 to 221, and the reduction of unrated surgeons accounts for 97% of the reduction.

One might argue that the reduction of CABG surgeons is not purely due to the publication of report cards. Another compelling reason to explain this trend is technological substitution: more patients undergoing percutaneous coronary intervention (PCI) rather than

CABG. Indeed, Cutler and Huckman (2003) show that the substitution of PCI for CABG occurs over time, with much of the substitution effect occurring in the 1990's. Understanding the causal impact of report card on the patterns in the decreasing number of surgeons, particularly the unrated surgeons, is important, but is left for future research. To fully address the question we are addressing here, we keep those unrated surgeons in our sample in order to have a complete sample of rated hospitals, because a hospital may be rated even if some or all of its' affiliated surgeons are not.

Table 3 contains the summary statistics of in-patient discharge data for CABG patients. In both samples, about 2% of patients died in-hospital. 7% of patients were admitted to hospitals with high mortality risk flags, while 10% of patients had hospitals with low mortality risk flags. Four to six percents of patients received CABG from surgeons with high mortality flags, while 2-3% of patients received their surgeries from surgeons with low mortality flags. 28% of patients received CABG surgeries from unrated surgeons.[16]

4 Conceptual Framework

4.1 Supply

On the supply side, the health economics literature typically assumes that physicians maximize their expected profit by choosing quantity and price (summarized in McGuire (2000)). Physicians, as well as hospitals may have different responses to the potential effects of report cards on the future demand for their services and future income.

[16] Our patient composition is as follows. Respectively, 86.5%, 0.3%, 3.3%, and 9.9% of the patients are white, Asian, black, and other races. 70% are male patients. 2.7% of the patients are 30 to 44 years old, 38.4% are 45 to 64 years old, and 58.9% are at least 65 years old. The average severity of illness of all patients is 1.5. 33.1% are emergency cases. Uninsured, private insurance, Medicare, and Medicaid patients make up 0.8%, 38.3%, 53.6%, and 3.9% respectively.

One potential response is to engage in up-coding. Strictly defined, up-coding occurs when a provider increases a patient's recorded severity so that the provider can get reimbursed at a higher rate from the insurer. Analogous to up-coding for increased reimbursement, up-coding for a better 'grade' may be an issue, as the increase of a patients' severity changes the risk-adjusted mortality calculations in a way that would be potentially favorable for a surgeon's (or hospital's) rating. For example, when the CABG report cards were first introduced in New York, Green and Wintfield (1995) found a significant increase in the prevalence of five comorbidities, which accounted for 41 percent of the decrease in risk-adjusted mortality. While the empirical evidence is still unclear about whether or not public reporting induces true up-coding behavior, or simply induces more accurate coding, the former is nonetheless a possible gaming response by providers. We argue, however, that any trend toward up-coding would have occurred during the initial transitions to the report card market environment and led to a new equilibrium level of coding intensity.[17] Therefore, our examination of the impact of the latter years' report cards provides some insight into the effects they have net of this up-coding behavior.

Second, providers may engage in patient selection. While formal patient dumping at the hospital-level is difficult given the regulatory environment, surgeons have more freedom whether to accept a patient. Patient selection in this setting will occur when surgeons avoid operating sicker patients who they perceive as having a "true" probability of dying that exceeds the expected risk-adjusted mortality probability. The reason the providers would avoid these patients is the belief that these patients could disproportionately hurt their reputation via a poor rating on a report card.

17 There is a trade-off in risk between the additional payment and improvement of report cards gained and the risk of additional scrutiny up-coding may draw from payers. While we do not deal with it explicitly here, profit-maximization theory would suggest transition to a new equilibrium level of coding intensity would happen quickly.

Third, report card publication may change a surgeon's willingness to offer CABG at all. Poor performing surgeons practicing in states with report cards systems in place may switch to other specialties or exit the market either by retiring early or moving to states without a publicly reported performance system. Chassin (2002) and Hannan et al. (1995) found that 27 low-volume, high-mortality surgeons exited the New York market or switched to performing other surgery. In summary, these actions by surgeons are all theoretical possibilities, while actually determining whether the report card publication has any impact on the market remains an empirical question.

As for hospitals, given that more than 60 percent of hospitals nationally are not-for-profit, economists typically assume that hospitals maximize an objective function with different maximands, such as quality and quantity, rather than using a traditional profit maximization condition. This objective function is subject to a break-even budget constraint which assumes that the equity capital has been obtained through philanthropic donations, debt, and retained earnings (Newhouse 1970; Frank and Salkever 1991). In recent literature[18], empirical studies have suggested that not-for-profit hospitals are similar to for-profit hospitals in terms of quality of care, prices, uncompensated care provisions, technology adoption, etc.

Regardless of their objectives, hospital administrators may respond to report cards similar to the way surgeons respond. Hospital managers may encourage up-coding, may restrict sicker patients from being considered eligible for the rated procedure, may allocate fewer patients to poor-performing surgeons, or may revoke the admitting privileges of poor-rated surgeons. Differentiating between these potential explanations, while interesting and important, requires more detailed data than we had at our disposal and is therefore left to future research; here we are focused on whether there is an observable impact of report cards on procedure volume.

[18] See a comprehensive review in Sloan (2000).

4.2 Demand

On the demand side, when a patient becomes ill, the patient and his/her cardiologist may jointly determine where and from whom to have the CABG surgery. [19] Patients' utility function depends on both observable and unobservable consumer and providers' characteristics. Suppose the indirect utility that patient *i* (=1, 2, ..., N) derives from receiving CABG surgery from surgeon *k* at hospital *j* is U_{jik}, which is a function of patient's characteristics, report card ratings and exogenous variables such as the distance between a patient's residence and the chosen hospital. The patient chooses the hospital and the surgeon pair with the greatest indirect utility.

If demanders of surgery respond to the information provided in the report cards, we can expect patients will choose high quality providers and avoid low quality providers. Sicker patients may be more likely to choose better surgeons because they have more to gain by doing so. On the other hand, relatively healthier patients may also be more likely to choose better surgeons because they could have more search time and be better informed about surgeon quality.[20]

4.3 Equilibrium

We address in this paper two potential effects of public reporting of provider quality information: the impact of report cards on the patient volume of providers and the impact report cards have on patient matching. In our empirical specifications described in the next section, we first perform the analysis at hospital- and surgeon-levels, and then focus on the patient-level hospital-surgeon choice. It is important to point out that our analyses do not distinguish whether

[19] In cases where a patient is admitted in an emergency situation, the choice is determined by the patient and possibly family and friends. This may be done after brief consultation with emergency responders, but that is inconsequential to the analysis at hand.

[20] This choice may not be made by the patient alone. It could be done in consultation with their primary care provider, referring cardiologist and family, but it is a choice nonetheless, and the exact people involved in helping in this choice does not matter given the primary aim of this model.

the volume change or hospital-surgeon choice is due to demand- or supply-side factors. The equilibrium volumes or patient's final choice may change, due to the report card publication, in response to the change of demand-side factors, supply-side factors, or both.

In the patient-level analysis, supply-side factors will enter the choice model through the changing choice set of the patients. The choice set will change as a result of new information. Poor performing surgeons may exit the market by retiring earlier, relocating to other states without public reporting, or switching to types of surgery in which outcomes are not publicly reported. Alternatively, they may be forcibly selected out of the market by health plans that drop them or by hospitals revoking or constraining admitting privileges. Thus, the supply side responds to the information, and the choice set for each patient and referring cardiologist will change. The potential patient selection of health care providers will also affect patient's final choice of hospital/surgeon.

5 Empirical Specification

5.1 Hospital-Level and Surgeon-Level Volume Analysis

To identify the effects of report cards on hospital-level and surgeon-level volumes, we take advantage of the panel structure of the data. Unobserved providers' quality may affect both ratings and volumes, which confounds the causal impacts of report cards. For example, providers with better unobserved quality may attract higher volumes because of word-of-mouth or formal referral, which could result in better health outcomes either because they are better quality providers or through the volume-outcome relationship. Alternatively, providers with poor unobserved quality may manipulate both ratings and volumes by up-coding or cherry picking. As a result, cross-sectional analyses that rely on rating variations across the providers will yield biased results. Our identification strategy is to employ the hospital or surgeon fixed effects that exploit the variation within the hospital or surgeon. By removing the time-invariant unobserved heterogeneity with fixed effects, we will be able to identify the causal impact of ratings on volumes.

To test the effects of report cards on hospital-level volume, we use the number of CABG procedures performed in a hospital in a quarter as the dependent variable.[21] We estimate the following regression

$$HospVolume_{jqt} = \alpha_0 + \alpha_1 High_{jq} + \alpha_2 Low_{jq} + \alpha_3 H_{ij} + \delta_t + \gamma_j, \qquad (1)$$

where j indexes hospital, q indexes quarter of hospital admission (q = 1, … , 31; from the third quarter of 1998 to the first quarter of 2006), and t indexes year of hospital admission.
The first independent variable of interest is a dummy variable indicating whether the hospital received a high in-hospital mortality flag ($High_{jq}$) in the most recent report card prior to performing the CABG surgery.[22] The coefficient α_1 is expected to be negative, i.e. hospitals flagged with high mortality perform fewer procedures relative to their counterparts.

The second independent variable of interest is a dummy variable indicating whether the hospital received a low in-hospital mortality flag (Low_{jq}) in the most recent report card prior to performing the CABG surgery. We expect α_2 to be positive, i.e. hospitals flagged with low mortality perform more procedures relative to their counterparts. Both of the dummies measure the impact relatively to the excluded group, which consists of hospitals whose in-hospital mortality ratings are same-as-expected, i.e. their actual mortality rates inside the 95% confidence interval. Hospital characteristics H_{ij} include ownership types, bed size, and teaching status, δ_t is a vector of time fixed effects, and γ_j is a vector of hospital fixed effects.

[21] To simplify our notations, we drop the subscript of severity s hereafter. However, we will perform the empirical analyses by patients' severity of illness.
[22] Cutler et al. (2004) did not find that the presence of any residual effect on volume of being poorly rated on the older report cards: it was only the most recent report card that mattered.

We estimate a similar regression using the number of CABG procedures performed by a surgeon in a quarter as the dependent variable to test the effects of report cards on surgeon-level volume as follows:

$$SurgVolume_{kjqt} = \beta_0 + \beta_1 High_{kjq} + \beta_2 Low_{kjq} + \beta_3 NotRated_{kjq} + S_{kjqt} + \delta_t + \gamma_j + \eta_k \quad (2)$$

where k indexes surgeons, q indexes quarter of hospital admission (q = 1, … , 31; from the third quarter of 1998 to the first quarter of 2006), t indexes year of hospital admission, and j indexes hospital. Surgeons who have multiple admitting privileges will be treated as different observations. η_k is a vector of surgeon fixed effect. S_{kjqt} is a vector of a surgeon k at hospital j's observable characteristics, including experience and experience squared. $High_{kjq}$ and Low_{kjq} indicate whether the surgeon received a high or low in-hospital mortality flag in the most recent report card prior to performing the CABG surgery. The excluded group consists of surgeons whose actual in-hospital mortality rates are within the 95% confidence interval. We expect β_1 to be negative and β_2 to be positive.

The dummy $NotRated_{kjq}$ indicates whether the surgeon appears in the most recent report card. Due to minimum volume thresholds for reporting outcomes set by PHC4, surgeons will be unrated if they performed less than 30 CABG surgeries in PA in the years when the report card information was collected. Because surgeons who had a very low volume of CABG or newly entered the market will not be rated, thus, by definition, β_3 should be negative. Our main interest is not β_3; rather, we are more interested in α_1, α_2, β_1, and β_2. However, hospitals that have many non-rated surgeons may be rated. We will run regressions separately with and without including patients whose surgeries were performed by non-rated surgeons.

Finally, to test the persistency of report card effects and capture the "news" content, we interact *High* and *Low* indictors with indicators for the number of years since the most recent report (e.g., 1 year, 2 years, 3 years and more than 3 years).

5.2 Patients' Choice Analysis

Unlike the volume analyses described above, the patient's choice model allows us to account for patients' heterogeneity, and therefore an alternative and more detailed insight into the matching between patient and surgeon. We use a discrete choice model to formulate a patient's selection of a hospital-surgeon pair. Surgeons who have multiple admitting privileges will be treated as different alternatives. Because the decision of where and from whom to have a CABG surgery is likely to be jointly determined by the patient and his or her cardiologist, individual preference or knowledge about the CABG surgery can induce correlations in choosing those alternatives. We model such correlations using a mixed logit model, also known as the random parameter (or coefficient) logit model (Cameron and Trivedi, 2005; Hole, 2007; Train, 2009).[23]

Using a random utility (U) model consistent with discrete choice, we specify its representative part (V) and idiosyncratic part (ε) for individual i (i = 1,2, ⋯, N) as follows:

$$U_{ijk}^{s} = V_{ijk}^{s} + \varepsilon_{ijk}^{s} \quad (s \in \{\text{low, high}\}; j = 1, 2, \ldots, J; k = 1, 2, \ldots, K), \quad (3)$$

where s denotes the severity of cardiac illness, j denotes a hospital, and k denotes a surgeon. Following the literature on discrete choice models, we use a linear specification for the representative utility V_{ijk}^{s} (Train, 2009), which is additively separable in hospital and surgeon

[23] We use the user-written "mixlogit" command in Stata (Hole, 2007) and estimate the mixed logit model using the maximum simulated likelihood method, which uses Halton draws to simulate the likelihood function. "The superior coverage and the negative correlation over observations that are obtained with Halton draws combine to make Halton draws far more effective than random draws for simulation...... Bhat (2001) found that 100 Halton draws provided more precise results for his mixed logit than 1000 random draws." (Train, 2009, p.228). In the example given by Train (2009, p.229), 100 Halton draws gave very similar results to 1,000 Halton draws. We use 100 Halton draws in all our mixed logit estimations to reduce computation burden.

characteristics and parameterized by α and β:

$$V_{ijk}^{s} = \alpha' H_{ij}^{s} + \beta' S_{ijk}^{s}, \quad (4)$$

where H_{ij}^{s} is a vector of hospital *j*'s observable characteristics, S_{ijk}^{s} is a vector of surgeon *k*'s observable characteristics.

To take into account correlations in choosing a surgeon (an alternative) but leaving the correlation structure in ε_{ijk}^{s} in equation (3) unspecified, we use the mixed logit model (Train, 2009). This model assumes that the heterogeneity in a decision-maker's preference or evaluation for alternatives induces the correlation in choosing alternatives. Specifically, the marginal utility (β) in the representative utility (V_{ijk}^{s}) in equation (4) is modeled as a random variable as opposed to a fixed parameter:

$$V_{ijk}^{s} = \alpha' H_{ij}^{s} + \beta_i' S_{ijk}^{s}. \quad (5)$$

The idiosyncratic part of the random utility (ε_{ijk}^{s}) in equation (3) is assumed to have the type I extreme value distribution, which gives the following choice probability *conditional* on β_i:

$$\Pr(y_i^{s} = k \mid H_{ij}^{s}, S_{ijk}^{s}, \beta_i) = \frac{\exp(\beta_i' S_{ijk}^{s} + \alpha' H_{ij}^{s})}{\sum_{g=1}^{J} \sum_{l \in B_g} \exp(\beta_i' S_{ijl}^{s} + \alpha' H_{ig}^{s})}. \quad (6)$$

Note that the set of alternatives $\{1, 2, \cdots, K\}$ is grouped by hospital into *J* subsets ($B_1, B_2, \ldots, B_J$) such that $\{1, 2, \cdots, K\} = \cup_{j=1}^{J} B_j$. We herein follow the literature on the mixed logit model (Train, 2009), assuming the difference between β_i and its mean $\overline{\beta}$ to be normally distributed. Specifically, we have

$$\beta_i = \overline{\beta} + \mathbf{u}_i \text{ and } \mathbf{u}_i \sim N(\mathbf{0}, \Sigma_{\beta}). \quad (7)$$

Substituting equations (6) and (7) into equation (3), we have

$$U_{ijk}^{s} = \alpha' H_{ij}^{s} + \overline{\beta}' S_{ijk}^{s} + v_{ijk}^{s}, \text{ where } v_{ijk} \equiv \mathbf{u}_i' S_{ijk}^{s} + \varepsilon_{ijk}^{s}. \quad (8)$$

Thus, the mixed logit model takes into account any pairwise correlation between U_{ijk}^{s} and U_{ijl}^{s} ($k, l = 1, 2, \cdots, K$) conditional on the hospital and surgeon level characteristics (H_{ij}^{s} and S_{ijk}^{s}) through the following:[24]

$$Cov(v_{ijk}^{s}, v_{ijl}^{s}) = S_{ijk}^{s\prime} \Sigma_{\beta} S_{ijk}^{s}, \text{ for } k \neq l. \quad (9)$$

In this mixed logit model, despite that ε_{ijk}^{s} is independent across individuals ($i = 1, 2, \cdots, N$) and alternatives ($k = 1, 2, \cdots, K$), introducing the random parameters (β_i 's) allows for correlation between any pair of U_{ijk}^{s} and U_{ijl}^{s} ($k \neq l$), in which the two surgeons are not necessarily affiliated with the same hospital.

The *unconditional* choice probability can be obtained by integrating the conditional choice probability in equation (4) over a probability density function of β (with subscript *i* suppressed) parameterized by θ:

$$\Pr(y^{s} = k \mid H_{j}^{s}, S_{jk}^{s}, \theta) = \int \frac{\exp(\beta' S_{jk}^{s} + \alpha' H_{j}^{s})}{\sum_{g=1}^{J} \sum_{l \in B_g} \exp(\beta' S_{jl}^{s} + \alpha' H_{g}^{s})} f(\beta \mid \theta) d\beta. \quad (10)$$

For the normal probability distribution of β, θ refers to $\overline{\beta}$ and Σ_{β} in equation (7). The mixed logit model obtains θ estimates using the maximum simulated likelihood method (Hole, 2007; Train, 2009).

Note that the unconditional choice probability specified in equation (10) is a weighted average of the choice-probability conditional on β, in which the weight is $f(\beta \mid \theta)$ — the probability density of β in the entire population (of patients). Using the Bayes' rule, as shown in

[24] If the pairwise correlation coefficient is a nonzero constant within each hospital, the mixed logit model is equivalent to a nested logit model.

Train (2009), we can obtain a distribution of β at the level of each patient whose surgeon-choice has been observed in our sample,

$$g(\beta \mid H_j^s, S_{jk}^s, \theta, y^s = k) = \frac{\Pr(y^s = k \mid H_j^s, S_{jk}^s, \beta) f(\beta \mid \theta)}{\Pr(y^s = k \mid H_j^s, S_{jk}^s, \theta)}. \quad (11)$$

Equation (11) is computable because $\Pr(y^s = k \mid H_j^s, S_{jk}^s, \theta)$ can be estimated by maximum simulated likelihood, $\Pr(y^s = k \mid H_j^s, S_{jk}^s, \beta)$ is specified by equation (6), and the unconditional density of β, $f(\beta \mid \theta)$, is assumed to be normal. Based on equation (11), we can obtain the average conditional distribution of β at each patient level as follows (Train, 2009):

$$\begin{aligned} \overline{\beta}_i &= \int \beta g(\beta \mid H_{ij}^s, S_{ijk}^s, \theta, y_i^s = k) d\beta \\ &= \int \beta \left(\frac{\Pr(y_i^s = k \mid H_{ij}^s, S_{ijk}^s, \beta) f(\beta \mid \theta)}{\int \Pr(y_i^s = k \mid H_{ij}^s, S_{ijk}^s, \beta) f(\beta \mid \theta) d\beta} \right) d\beta. \end{aligned} \quad (12)$$

The estimate of $\overline{\beta}_i$ represents the individual level (each patient's) preference towards characteristics of alternatives (surgeons) in the choice set, whereas θ characterizes (the moments of) the distribution of the preference (β) for all patients. The former reveals information about a particular patient's preference in choosing surgeons conditional on his or her decision already made. The latter gives the distribution of the preference in the entire population (of patients).

It is important to note that $\overline{\beta}_i$, the average preference in the conditional distribution ($g(\beta \mid H_{ij}^s, S_{ijk}^s, \theta, y_i^s = k)$), is conceptually distinct and numerically different from $\overline{\beta}$, the average preference in the unconditional distribution ($f(\beta \mid \theta)$). However, the similarity between the average of all estimated $\overline{\beta}_i$'s across all patients in the estimation sample and the estimated $\overline{\beta}$ suggests that the mixed logit model in equation (10) is correctly specified and accurately estimated (Train, 2009, p.270), which serves as a specification check in our following empirical analyses.

Mixed logit model offers at least two advantages over other discrete choice models such as a nested logit model. First, mixed logit model is arguably the most flexible way to deal with the problem of independence from irrelevant alternatives (IIA) pervasive in discrete choice models (Train, 2009, p.134). Second, we can compute the individual-level average preference estimate specified by equation (12) to conduct a specification check for the mixed logit model. [25]

We estimate the mixed logit model separately for each report card period after 2000.[26] In equation (4), the vector H_{ij}^{s} includes the most recent hospital report card rating (i.e. high mortality flag), hospital ownership type, number of beds, teaching status, the Euclidean distance measured between patient *i*'s residence zip code to hospital *j*'s zip code, and the distance-squared term. In the most inclusive specification, we include the interactions of hospital high mortality flag with hospital characteristics. The vector S_{ijk}^{s} includes a surgeon's most recent report card rating, years of experience, and its squared term. In our estimation sample, all hospital-surgeon pairs within a 50-mile radius of a patient's residence constitute his or her choice set.[27]

To examine the heterogeneity in the effects of report cards on a patient's surgeon choice, we also estimate the mixed logit model by the patient's severity of illness and admission type (emergency versus non-emergency) at the time of the surgery. In our patient-level surgeon-choice analyses, no alternative-invariant variables, such as patient characteristics, are directly included into the estimation model. Conceivably, a patient's choosing whom to perform a surgery can be largely explained by a surgeon's characteristics (which are alternative-varying) as

---

[25]We use the user-written "mixlbeta" command in Stata (Hole, 2007) with 100 Halton draws.
[26]We are not able to perform the mixed logit estimation on earlier periods because of too many nonrated surgeons, which make the choice set too large to implement the mixed logit estimator.
[27]In our sample, the average distance measured between a patient's residence zip code and the zip code of a hospital included in his or her choice set is 17.7 miles, and the standard deviation is 30 miles.

opposed to the marginal utility the patient may derive from each alternative according to his or her own (alternative-invariant) characteristics. Empirically, estimating the return to each alternative-invariant patient characteristic requires interacting it with each alternative in the choice set. Thus, two challenges arise if we include alternative-invariant variables into our non-binary discrete choice models. First, the sign and the magnitude of any coefficient (and its marginal effect) estimate for alternative-invariant variables depend on which alternative is set to be the base category. The interpretation of those estimates relative to the base category will add to empirical ambiguity because there is no natural base category (a lone surgeon) in our empirical setting. Second, such an estimation procedure adds formidable challenges and computation burden to our high-dimensional nonlinear numerical optimization.

Our mixed logit analyses follow the two-part procedure proposed by Train (2009, pp.280-281). First, as specified in equation (7), we obtain the estimates of the means and standard deviations of the random parameters for the population. Second, we compute the means and standard deviations of the random parameters at each patient level conditional on our sample in which each patient's actual choice (revealed preference) has been observed. As explained by Train (2009, p.281), there are two reasons why obtaining the patient-level conditional distributions is preferred to including patient's characteristics directly into the discrete choice model. First, adding patient's characteristics to the estimation equation requires that the effect to be additive and homogeneous across patients, which is unnecessarily restrictive. In mixed logit models, a patient's evaluation or preference for each alternative, either based on alternative-invariant or alternative-varying characteristics, is modeled as a separate random variable multiplied with the characteristics of each alternative. This leads to the second point in favor of our two-part procedure: the conditional distribution at each patient-level can reveal patterns that cannot be related to observed patient characteristics (Train, 2009, p.281), especially in the presence of unobserved heterogeneities among patients.

6 Empirical Results

6.1 Hospital-Level and Surgeon-Level Volume Analysis

Table 4 presents the results of our hospital-level volume analysis. Our sample consists of 1,469 hospital-quarters in Pennsylvania between the third quarter of 1998 to the first quarter of 2006. The first panel of Table 4 shows the results using hospital quarterly volume on all CABG cases as the dependent variable. We first regress quarterly volume on report card ratings while controlling for year fixed effects and hospital characteristics. The results suggest that being identified as a high-mortality hospital in the most recent report card is associated with a decline of 9 CABG surgeries per quarter. This decline is not statistically significant. However, being identified as a low-mortality hospital in the most recent report card is associated with an increase of 33.41 CABG surgeries per quarter, and this increase is statistically significant at the ten percent level.

In our second specifications, we include hospital fixed effects to control for unobserved heterogeneity associated with hospital quality. The inclusion of hospital fixed effects significantly reduced the point estimates of all rating coefficients. In our final specification with the inclusion of both hospital characteristics and hospital fixed effects, hospitals with poor ratings are associated with a decrease of 5.60 CABG surgeries and ones with good ratings are associated with an increase of 5.13 CABG surgeries. Though coefficients are not precisely estimated, it is interesting to note that the overall volume effects are of similar magnitudes but in opposing directions, that is, patients not treated at hospitals with poor ratings will undergo surgery at hospitals with higher ratings. It implies that hospital report cards do not change the population of patients who received the CABG surgeries.

We repeat our regressions for low-severity ((4)-(6)) and high-severity CABG cases ((7)-(9)) in Table 4. The results show a similar pattern to the analysis of the whole volume. Controlling hospital fixed effects significantly reduces the magnitude of the coefficient, suggesting that analyses based on the cross-section data yield biased results. In the most inclusive specification (column (6)), being identified as a high-mortality hospital in the most recent report card is associated with a decline of 4.47 low-severity CABG surgeries per quarter, though this decline is not statistically significant. 79% of this decrease is almost entirely picked up by the hospitals with good ratings. We find similar results for volume on high-severity CABG cases; the only difference is that the coefficients are smaller (column (9)). Hospitals with poor ratings are associated with a decrease of 1.20 CABG surgeries per quarter, and this decrease is picked up by hospitals with good ratings.

In the second model of Table 4, we interact *High* and *Low* indictors with indicators for the number of years since the most recent report (e.g., 1 year, 2 years, 3 years and more than 3 years). In the most inclusive specification, hospital volumes significantly decrease two years after receiving a high mortality flag (column (3)), particularly the volumes on low-severity patients (column (6)). There are several possible explanations for this finding. First, it takes time for information to diffuse and volume to decline at bad hospitals. In particular, report cards were published every year after 2002, thus, the variations of 2-3+ years interactions only come from earlier report cards. It may take longer for information to diffuse in earlier periods. Second, it takes time for insurance companies to update their networks, and/or the fact that employers only really contract once a year with different insurers, and they may go toward insurers that exclude poor hospitals. Third, surgeons who feel they are better than average will leave bad hospitals or steer patients toward admission at better hospitals if they have privileges, but this cannot happen instantaneously.

Overall, there are four important things to consider in light of the results of the estimation reported in Table 4. First, hospital report cards appear to have no significant impact on surgical volume at the hospital-level. Second, hospital report cards do not appear to change the population of patients who received CABG surgeries. Third, if we consider only the magnitude of the coefficients, hospital report cards have a larger impact on the distribution of healthier patients across hospitals, consistent with the idea that these patients have more time to gather information on patient quality. Fourth, bad ratings take about a year to have a negative effect (conditional on there being an effect), however, the effect is not persistent.

Table 5 presents the results on surgeon-level volume analysis. We run samples with (columns (1)-(3)) and without (columns (4)-(6)) non-rated surgeons separately. Results from these two samples are very similar. Thus, we only discuss one set of the results in more detail below. Again, we add hospital characteristics, hospital fixed effects and surgeon fixed effects incrementally. The iterative introduction of each set of fixed effects reduces the magnitude of the coefficients relative to the previous specification. We focus on and report only the most inclusive specifications.

For all CABG cases (column (1)), being identified as a high-mortality surgeon in the most recent report card is associated with a decline of 4.76 CABG surgeries per quarter and the coefficient is significant at the 1% level. Being identified as a low-mortality surgeon is associated with an increase of 4.63 CABG surgeries per quarter, though this coefficient is not precisely estimated. Surgeons with no report card, either a low-volume surgeon or a new surgeon in the market, perform 8.04 fewer surgeries after the release of the report cards and the coefficient here is also statistically significant at the 1 % level.

We repeat our regressions for low-severity (column (2)) and high-severity CABG cases (column (3)). Overall, our results suggest a similar pattern to the whole volume analysis: high mortality (low-performing) and non-rated surgeons experience a subsequent decrease in volume. Surgeons with poor ratings in the most recent report card will have a CABG volume reduction of 3.15 and 1.53 per quarter on low and high severity patients, respectively. Unrated surgeons in the most recent report card will have a CABG volume reduction of 5.17 and 3.69 per quarter on low and high severity patients, respectively.

Again, the fact that the magnitude of the reduction is higher for low severity patients is consistent with the idea that healthier patients have more time to gather information before making a choice of surgeons. The healthier patients also can afford to wait for the better rated surgeon to be available. However, these results have potentially negative implications for patient-surgeon matching. One would arguably want the best surgeons operating on sicker patients. Instead, conditional upon staying in the market, a low-performing surgeon's volume of relatively healthier patients is decreasing more than the volume of relatively sicker patients, suggesting that the patient population they face after a poor rating or a non-rating is sicker than before the recent report cards. Of the total reduction of volume associated with low-performing or unrated surgeons, 4.07 of the procedures on low severity patients appears to be captured by high-performing surgeons, but there is little evidence of reallocation of high severity patients to high-performing surgeons.

This strengthens the assertion that report cards affect the distribution of patients across surgeons and therefore may not result in the improved patient-surgeon matching that is often cited as a potential benefit of these report cards. It also raises concerns that high severity patients may have a more difficult time in getting a CABG, consistent with the findings of Schneider and Epstein (1996) in which cardiologists reported having a tougher time placing sicker patients after the advent of public reporting in Pennsylvania.

Interestingly, the surgeons with a high-mortality flag experienced persistent declines in volumes (Table 5, lower panel). The estimated coefficients are statistically significant and tend to be more negative as the time since the report increases. Similar to previous findings, volume reductions are larger on low-severity patients. The report cards significantly increased volume for those surgeons receiving low-mortality flags in the first two years following a report. The increases are statistically significant for low-severity patients in particular. The coefficient estimate is negative in subsequent years, but not statistically precise. This finding is consistent with a short-term 'bump' in volume that occurs as a result of a good report card and eventually subsides due to the trend toward overall decreasing CABG volume throughout this period.

To sum up, in our provider-volume analysis we found that CABG report cards have an impact on surgeon-level volume but very limited on hospital-level volume. The implication is that the market responds more to the surgeon rating than the hospital rating. Related to surgeon-volume analysis, we find that surgeons with poor ratings or who are non-rated will have lower surgical volumes after the publication of report cards, regardless of patient severity. The impacts are persistent over time for poor-performing surgeons. However, 49% of the reducing volumes from healthier patients will be shifted to better surgeons but only 18% occurs among sicker patients. Overall this analysis suggests hospitals may be able to respond to poor ratings by re-allocating the patients across surgeons within a hospital, and thus patients and referring doctors are less apt to avoid specific hospitals, rather they are more apt to avoid specific surgeons who are rated as poor or not rated at all.

6.2 Patients' Choice Analysis

Having found that the market is more responsive to the surgeons' report card, we further study the demand-side responses to surgeons' ratings. We estimate a mixed logit model separately for each report card period. We also estimate models separately with and without non-rated surgeons. In the aggregate analysis, the causality between non-rated and volumes is bi-directional (i.e. non-rated will lead to lower volume, but lower volume will make surgeons non-rated). However, from the individual patient level, choosing a surgeon may or may not make the surgeon nonrated.[28] Thus, the causal relationship between non-rated and patients' choice is likely to be one way.

Due to space constraint, we only report results for the 2003 report card episode.[29] The results (in Table 6) suggest that CABG patients are less likely to choose a high-mortality surgeon or an unrated surgeon, these results are robust regardless of whether interactions of high mortality flag and other hospital variables are included (columns (3) and (4)). The coefficients of high mortality flag remain significantly negative after excluding non-rated surgeons (columns (2) and (4)).

These results are consistent with our findings in Table 5 and suggest that someone, whether patients, referring physicians, health plan providers, or even hospital administrators, is responsive to surgeons' ratings. Though beyond the scope of our data, further research into whether it is a patient in conjunction with their cardiologist, or hospital management using report

[28] The only way an individual patient could willfully change a surgeon from non-rated to rated is if they knew they would be the surgeon's 30th patient in the rating year (the threshold for getting rated) which is a highly unlikely scenario.
[29] Mixed logit results for other report card periods are largely similar. Those results are available upon request.

cards to steer patients to better providers, is necessary and important to improve our understanding of the mechanisms through which report cards impact medical care markets.

Our two-part estimation procedure described in Section 5.2 provides a diagnostic check for the mixed logit model specification. If the mixed logit model is correctly specified, then the means of the estimated random parameters for the unconditional distribution (in Table 6) should be similar to the means of the estimated random parameters at each patient level for the conditional distribution (in Table 7) (Train, 2009, p.270). Our mixed logit estimates from the population distribution (our study population including all patients; Table 6) and the conditional distribution (our sample in which each patient's choice is observed; Table 7) are almost identical, suggesting the validity of our mixed logit model specification.

Hospital characteristics also play significant roles in determining patients' selection. We find that patients are more likely to choose teaching, nonprofit, larger and closer hospitals. Our result shows that patients have higher probability to go to high mortality flag hospitals, which is counterintuitive and inconsistent with our findings in Table 5. When we account for patients' heterogeneity in the mixed logit model, an important explanation of this counterintuitive result emerges, though.

Related to the issues discussed in Section 5.2, to further account for patients' heterogeneity related to health status at the time of the procedure, we repeat our mixed logit estimations, stratifying the full sample by patient severity and admission type (emergency versus non-emergency) for each report card period. The results for the 2003 report card episode with interactions of high mortality flag and other hospital characteristics are summarized in Tables 8-9.[30] Overall, all patients, regardless of severity of illness, are less likely to choose high-mortality surgeons or unrated surgeons, but that hospital ratings do not play a significant role in patients'

[30] Mixed logit results for other report card periods are largely similar. Those results are available upon request.

choices (Table 8). In Table 9, both non-emergency and emergency patients are less likely to choose high-mortality surgeons, but the former are more responsive to the poor rating of surgeons. Hospital's high mortality flag is not statistically significant for the non-emergency patients. However, these coefficients are significantly positive for emergency patients. These results suggest that the positive coefficients of HMF that we see in Table 6 are largely driven by patients with emergent care needs. These patients might select (or be routed to) "bad" hospitals, because they are located much closer than any other hospital, due to the nature of the optimal treatments for the underlying disease for which one receives CABG. However, conditional upon being at a bad hospital they are still routed away from "bad" surgeons.[31]

So far we only focus on the signs of coefficients, but the question now becomes how to interpret the magnitude of these coefficients. In our mixed logit model, for the variables with standard deviation estimates of the associated coefficients statistically significant from zero, we can obtain the information about the distribution of patient preferences or evaluations for surgeon characteristics. Take the results without nonrated surgeons as an example (Table 6, (4)). Our results suggest that 94%[32] of patients choose experienced surgeons and 93%[33] patients would avoid surgeons with high mortality flag.[34] For nonemergency patients (Table 9, (1)), 92% choose experienced surgeons, 90% avoid low-rated surgeons and 80%[35] avoid non-rated surgeons. Taken all results together, 90%-94% patients avoid having a surgery performed by a

---

[31] If a patient is suffering an AMI, which is often the event that leads to CABG, time to reperfusion is an important predictor of the outcome of treatment, thus there would be a tension between getting to a 'better' hospital and minimizing the time to treatment.

[32] This figure is given by $\Phi(b_k / s_k)$, where Φ is the cumulative standard normal distribution, b_k and s_k are the mean and the standard deviation, respectively, k denotes the years of experience.

[33] This figure is given by $\Phi(-b_k / s_k)$, where Φ is the cumulative standard normal distribution, b_k and s_k are the mean and the standard deviation, respectively, k denotes a surgeon with a high-mortality flag.

[34] We calculate the figures for all variables that have statistically significant standard deviations. Results are available upon request.

[35] This figure is given by $\Phi(-b_k / s_k)$, where Φ is the cumulative standard normal distribution, b_k and s_k are the mean and the standard deviation, respectively, k denotes a nonrated surgeon.

poorly rated surgeon.

Another way to explain our coefficients is to use an equivalence-type interpretation of our mixed logit estimates (Train, 2009, p.272). For example, we find that a surgeon having received a high mortality flag is perceived by a patient as having approximately 17 years less experience (Table 6, (1)). Furthermore, our results suggest that this figure in a comparable setting (i.e. the most inclusive specification) ranges from 15 years (high severity patients) to 34 years (nonemergency patients).

7 Conclusion

Report cards publicly disclose information on the performance of healthcare providers, as measured by patient health outcomes. The intent of mandating public reporting is to improve the quality and efficiency of medical care, thus improving social welfare. In this paper we investigate what impacts public reporting of provider performance have had within the market for CABG's in Pennsylvania, and whether there is evidence of improved social welfare.

Our analyses of the impacts of public information on provider volume are done at both the hospital level and at the surgeon level, as public information may have different impacts for these two groups. Indeed we find that public reporting led to a decrease in volume for unrated and poor performing surgeons, but interestingly, the volume of the high performing surgeons does not increase by an offsetting amount. In addition, we do not find a statistically significant effect on hospital volume once we control for unobserved heterogeneity, which is in contrast to findings of Cutler et al. (2004). These findings persist when we analyze the impacts by patient severity.

Subsequent the volume impact, we investigate the matching between patients and providers. Results of our patient choice modeling suggest that public reporting leads to poor performing or unrated surgeon avoidance. This model demonstrates that distance to the hospital is likely a more relevant factor in the choice of hospital than the rating it received, but conditional upon being admitted to the hospital the patients sort to better rated surgeons. The mechanism of exactly whether this low performing surgeon avoidance behavior is patient preference driven, referring doctor driven, or somehow due to hospital management decisions is left to future research.

The main question related to the impact of publicly provided healthcare provider report cards is whether they improve the market. If report cards accurately reflect the quality of healthcare provider, welfare improvement can be achieved through enhanced information symmetry between patients and healthcare providers and reduced search costs for patients who value high quality. However, report cards raise the stakes for a physician performing a surgery on a high risk patient. Receiving a high-mortality flag could be perceived by a patient as the surgeon having 10-20 fewer years of experience, affecting the demand for their services. In this sense, report cards could potentially lead to crowding of higher quality surgeons (or other healthcare providers) for higher risk patients, either because of surgeons' unwillingness to operate on the patient or because the healthier patients are using the report card information to select better providers. Given the last point, the net welfare gain can be ambiguous. Further research is needed to assess the degree to which report cards affect total welfare, as well as examining the mechanisms by which the report cards lead to the sorting and avoidance behavior we have noted here.

Reference

Arrow, K. J., 1963. "Uncertainty and Welfare Economics of Medical-Care," *American Economic Review*, 53 (5), 224-239.

Beaulieu, N. D., 2002. "Quality Information and Consumer Health Plan Choices," *Journal of Health Economics*, 21 (1), 43-63.

Bundorf, M. K., N. Chun, G. S. Goda, and D. Kessler, 2009. "Do Markets Respond to Quality Information? The Case of Fertility Clinics," *Journal of Health Economics,* 28 (3), 718-727.

Cameron, A. C. and P. K. Trivedi, 2005. *Microeconometrics: Methods and Applications*, Cambridge: Cambridge University Press.

Chassin, M. R., 2002. "Achieving and Sustaining Improved Quality: Lessons from New York State and Cardiac Surgery," *Health Affairs*, 21 (4), 40-51.

Chernew, M., G. Gowrisankaran, and D. P. Scanlon, 2008. "Learning and the Value of Information: Evidence from Health Plan Report Cards," *Journal of Econometrics,* 144 (1), 156-174.

Cutler, D. M. and R. S. Huckman, 2003. "Technological Development and Medical Productivity: The Diffusion of Angioplasty in New York State," *Journal of Health Economics*, 22 (2), 187-217.

---, ---, and M. B. Landrum, 2004. "The Role of Information in Medical Markets: An Analysis of Publicly Reported Outcomes in Cardiac Surgery," *American Economics Review*, 94, 342-346.

Dafny, L. S. and D. Dranove, 2008. "Do Report Cards Tell Consumers Anything They Don't Already Know? The Case of Medicare HMOs?," *Rand Journal of Economics,* 29 (3), 790-821.

Dranove, D. and A. Sfekas, 2008. "Start Spreading the News: A Structural Estimate of the Effects of New York Hospital Report Cards," *Journal of Health Economics*, 27 (5), 1201-1207.

---, M. McClellan, and M. Satterwaite, 2003. "Is More Information Better? The Effects of 'Report Cards' on Health Care Providers," *Journal of Political Economics*, 111 (3), 555-588.

Ellis, R. P., 1998. "Creaming, Skimping and Dumping: Provider Competition on the Intensive and Extensive Margins," *Journal of Health Economics*, 17 (5), 537-555.

Epstein, A. J., 2006. “Do Cardiac Care Surgery Report Cards Reduce Mortality? Assessing the Evidence,” *Medical Care Research and Review*, 63 (4), 403-426.

Frank, R. G. and D. S. Salkever, 1991. “The Supply of Charity Services by Nonprofit Hospitals: Motives and Market Structure,” *The Rand Journal of Economics*, 22 (3), 430-445.

Green, J. and N. Wintfield, 1995. “Report Cards on Cardiac Surgeons-Assessing New York State's Approach,” *New England Journal of Medicine*, 332 (18), 1229-1232.

Hannan, E. L., A. L. Siu, D. Kumar, H. Kilburn Jr., and M. R. Chassin, 1995. “The Decline in Coronary Artery Bypass Graft Surgery Mortality in New York State: The Role of Surgeon Volume,” *Journal of the American Medical Association*, 273 (3), 209-213.

---, H. Kilburn Jr, M. Racz, E. Shields, and M. R. Chassin, 1994. “Improving the Outcomes of Coronary Artery Bypass Surgery in New York State,” *Journal of American Medical Assocation*, 271 (10), 761-766.

Ho, V., 2006. “Does Certificate of Need Affect Cardiac Outcomes?,” *International Journal of Health Care Finance and Economics*, 6 (4), 300-324.

Hole, A. R., 2007. “Fitting Mixed Logit Models by Using Maximum Simulated Likelihood.” *Stata Journal* 7(3): 388-401.

Iezzoni, L. I., A. S. Ash, M. Shwartz, J. Daley, J. S. Hughes, and Y. D. Mackierman, 1996. “Judging Hospitals by Severity-Adjusted Mortality Rates: The Influence of the Severity-Adjusted Method,” *American Journal of Public Health*, 86 (10), 1379-1387.

--- and M. A. Moskowitz, 1988. “A Clinical Assessment of MedisGroups,” *Journal of American Medical Association*, 260 (21), 3159-3163.

Institute of Medicine, 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*, Washington, D.C.: National Academy Press.

Jin, G. Z. and A. Sorensen, 2005. “Information and Consumer Choice: The Value of Publicized Health Plan Ratings,” *Journal of Health Economics*, 25 (2), 248-275.

Kessler, D. P., 2005. “Can Ranking Hospitals on the Basis of Patients' Travel Distances Improve Quality of Care?,” *NBER Working Paper #11226*.

McGuire, T. G., 2000. “Physcian Agency,” in A. J. Culyer and J. P. Newhouse, eds., *Handbook of Health Economics*, Elsevier Science B. V., chapter 9.

Mukamel, D. B. and A. I. Mushlin, 1998. “Quality of Care Information Makes a Difference: An Analysis of Market Share and Price Changes after Publication of New York State Cardiac Care Surgery Mortality Reports,” *Medical Care*, 36 (7), 945-954.Newhouse, J. P., 1970.

"Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital," *American Economic Review*, 60 (1), 64-74.

Peterson, E. D., E. R. Delong, J. G. Jollis, L. H. Muhlbaier, D. B. mark, G. T. O'Connor, and K. A. Eagle, 1998. "The Effects of New York's Bypass Surgery Provider Profiling on Access to Care and Pateint Outcomes in the Elderly," *Journal of the American College of Cardiology*, 32 (4), 993-999.

Robinson, J. L., D. B. Nash, E. Moxey, and J. P. O'Connor, 2001. "Certificate of Need and the Quality of Cardiac Surgery," *American Journal of Medical Quality*, 16 (5), 155-260.

Romano, P. and H. Zhou, 2004. "Do Well-Publicized Risk-Adjusted Outcome Reports Affect Hospital Volume?," *Medical Care*, 42 (4), 367-377.

Schneider, E. C. and A. M. Epstein, 1996. "Influence of Cardiac-Surgery Performance Reports on Referral Practices and Access to Care," *New England Journal of Medicine*, 335 (4), 251-256.

--- and ---, 1998. "Use of Public Performance Reports: A Survey of Patients Undergoing Cardiac Surgery," *Journal of American Medical Association*, 279 (20), 1638-1642.

Shahian, D. M., S.-L. Normand, D. F. Torchiana, S. M. Lewis, J. O. Pastore, R. E. Kuntz, and P. I. Dreyer, 2001. "Cardiac Surgery Report Cards: Comprehensive Review and Statistical Critique," *Annals of Thoracic Surgery*, 72 (6), 2155-2168.

Shearer, A. and C. Cronin, 2005. *The State-of-the-Art of Online Hospital Public Reporting: A Review of Fifty-One Websites*, Easton, MD: Delmarva Foundation.

Sloan, F. A., 2000. "Not-For-Profit Ownership and Hospital Behavior," in A. J. Culyer and J. P. Newhouse, eds., *Handbook of Health Economics*, Elsevier Science B. V., chapter 9.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Second edition. Cambridge and New York, Cambridge University Press.

Wedig, G. J. and M. Tai-Seale, 2002. "The Effect of Report Cards on Consumer Choice in the Health Insurance Market," *Journal of Health Economics*, 21 (1), 1031-1048.

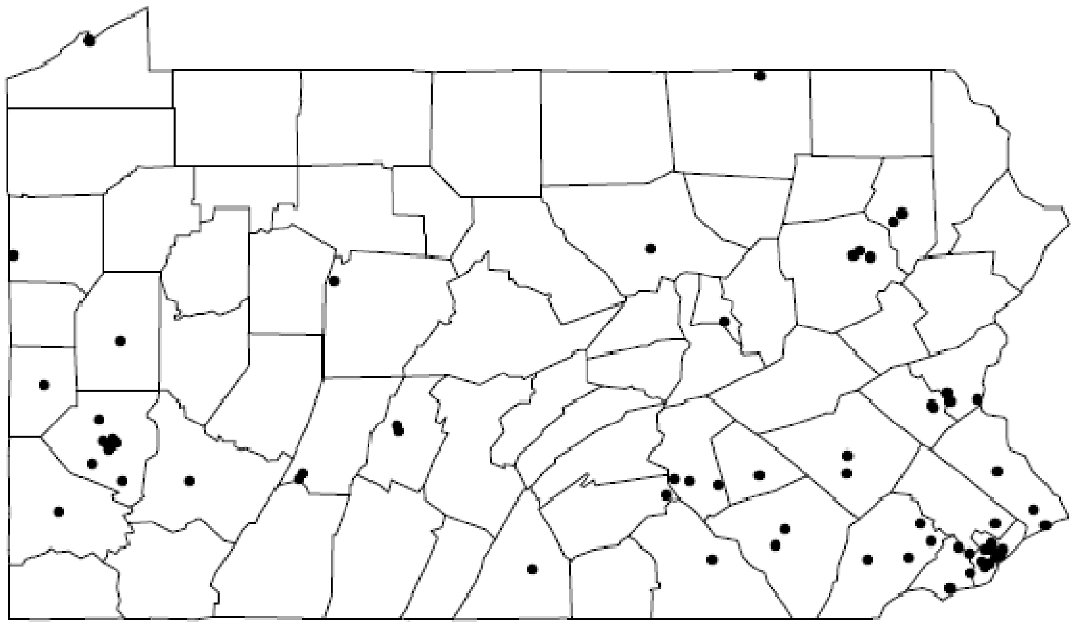

Figure 1: Distribution of CABG Hospitals in Pennsylvania

Table 1: Cardiac Surgery Report Card Data Collection and Publication Dates in Pennsylvania

Publication		Data Collection Period	Online Publication
Year	Quarter		
1992	4th	1990	No
1993	4th	1991	No
1994	4th	1992	No
1995	2nd	1993	No
1998	2nd	1994-1995	Yes
2002	2nd	2000	Yes
2004	1st	2002	Yes
2005	1st	2003	Yes
2006	1st	2004	Yes
2007	2nd	2005	Yes

Table 2: Number of Hospitals and Surgeons by Ratings

Year in which data was collected	1994/1995	2000	2002	2003
Number of Hospitals by Hospital Ratings				
High-mortality flag	4	4	7	3
Low-mortality flag	3	3	2	1
Same-as-expected mortality flag	35	48	50	55
Total number of hospitals	42	55	59	59
Number of Surgeons by Surgeon Ratings				
High-mortality flag	13	7	9	5
Low-mortality flag	5	2	0	0
Same-as-expected mortality flag	112	120	125	117
Not-rated surgeon	388	173	104	99
Total number of surgeons	518	302	238	221
Release Year/Quarter	1998/2	2002/2	2004/1	2005/1
Time period in which the report card is matched to	1998/3-2002/2	2002/3-2004/1	2004/2-2005/1	2005/2-2006/1

Table 3: Sample Statistics

Variable	All Sample Mean	All Sample Std. Dev.	Excluding Unrated Surgeons Mean	Excluding Unrated Surgeons Std. Dev.
Death	0.024	0.153	0.021	0.144
Hospital Ratings				
High mortality flag	0.072	0.258	0.069	0.253
Low mortality flag	0.097	0.296	0.097	0.296
Same-as-expected mortality flag	0.831	0.375	0.834	0.372
Surgeon Ratings				
High mortality flag	0.037	0.190	0.055	0.227
Low mortality flag	0.020	0.139	0.032	0.177
Same-as-expected mortality flag	0.664	0.472	0.913	0.282
Not rated	0.279	0.449		
Sample Size	114039		84235	

Table 4: Regression of Quarterly CABG Volume on Publically Reported Mortality Flag at Hospital Level

Dep. Variable =	All CABG Cases			Low-Severity CABG Cases			High-Severity CABG Cases		
Hospital Quarterly Volume	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mean [Std. Dev] of Dep. Var		76.5 [51.3]			45.5 [30.3]			30.3 [23.4]	
Model 1									
High Mortality Flag (HMF)	-9.112	-5.098	-5.600	-6.162	-4.120	-4.477	-2.694	-0.471	-1.195
	(14.508)	(5.275)	(5.25)	(7.936)	(3.596)	(3.542)	(6.785)	(0.835)	(2.077)
Low Mortality Flag (LMF)	33.413*	6.341	5.125	18.494*	4.669	3.55	15.068	1.436	1.578
	(19.594)	(9.547)	(9.286)	(10.053)	(6.714)	(6.168)	(10.027)	(0.734)	(4.394)
Model 2									
1 year since HMF	-10.073	-5.001	-5.607	-6.019	-3.488	-3.884	-3.958	-1.301	-1.845
	(10.547)	(4.829)	(4.583)	(5.826)	(3.817)	(3.624)	(4.857)	(1.471)	(1.363)
2 year since HMF	-16.864	-14.892***	-15.510**	-11.459	-10.900***	-11.171**	-4.662	-3.149	-4.175
	(14.421)	(5.588)	(6.021)	(6.951)	(3.817)	(4.345)	(7.810)	(3.335)	(2.953)
3 year since HMF	-0.436	6.629	6.209	-0.035	4.182	3.928	0.205	2.668	2.084
	(33.133)	(8.342)	(7.942)	(19.433)	(5.253)	(5.033)	(14.118)	(3.352)	(3.024)
3+ year since HMF	4.109	6.588	5.984	-2.362	-0.704	-1.452	6.351	6.78	7.319
	(41.672)	(12.649)	(13.171)	(24.539)	(7.557)	(7.974)	(17.767)	(5.528)	(5.868)
1 year since LMF	33.289*	-3.848	-5.176	17.587**	-1.35	-2.634	15.976*	-2.665	-2.489
	(17.307)	(7.683)	(8.091)	(8.306)	(5.310)	(5.327)	(9.506)	(4.607)	(4.711)
2 year since LMF	37.749*	19.734	18.695	19.489*	11.128	10.178	18.353*	8.543	8.517
	(20.647)	(16.115)	(14.827)	(10.949)	(10.685)	(9.129)	(10.214)	(6.604)	(6.659)
3 year since LMF	37.674	18.351*	17.891*	26.982	16.611**	16.169*	11.161	2.04	2.096
	(30.095)	(9.995)	(10.415)	(17.476)	(8.302)	(8.623)	(13.368)	(2.795)	(2.897)
3+ year since LMF	22.246	2.56	0.826	11.068	0.642	-0.935	10.746	1.318	1.235
	(25.086)	(7.062)	(7.660)	(13.545)	(4.444)	(4.828)	(12.187)	(3.547)	(3.650)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hospital Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Hospital Chracteristics	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes

Notes: Sample size is 1469. Standard deviations are in brackets. Standard errors are in parentheses. All standard errors are clustered by hospital. Hospital characteristics include a dummy indicating not-for-profit status, bed size and a dummy indicating teaching status. *, **, *** Significant at the 10%, 5% and 1% level for a two-tail test.

Table 5: Regression of Quarterly CABG Volume on Publically Reported Mortality Flag at Surgeon Level

Dep. Variable =	Sample with Non-rated Surgeons			Sample without Non-rated Surgeons		
Surgeon Quarterly Volume	All	Low-Severity	High-Severity	All	Low-Severity	High-Severity
	(1)	(2)	(3)	(4)	(5)	(6)
	n=6586	n=6586	n=6586	n=4338	n=4338	n=4338
Mean of the Dep. Var	21.9	13.0	8.7	25.1	14.8	10.1
[Std. Dev. Of the Dep. Var]	[14.9]	[9.3]	[6.8]	[13.83]	[8.58]	[6.75]
Model 1						
High Mortality Flag	-4.762***	-3.147***	-1.527**	-7.911***	-4.946***	-2.872***
	(1.407)	(0.845)	(0.679)	(2.013)	(1.311)	(0.779)
Low Mortality Flag	4.634	4.076**	0.921	3.288	2.835**	0.578
	(3.79)	(1.71)	(1.865)	(2.747)	(1.375)	(1.417)
Not Rated	-8.042***	-5.168***	-3.695***			
	(1.163)	(0.675)	(0.524)			
Model 2						
1 year since High	-3.371***	-2.165***	-1.095**	-4.184***	-2.508***	-1.466***
	(0.892)	(0.589)	(0.450)	(0.888)	(0.595)	(0.474)
2 year since High	-5.987***	-4.547***	-1.280**	-5.323***	-4.051***	-1.052*
	(1.237)	(0.819)	(0.627)	(1.174)	(0.790)	(0.630)
3 year since High	-8.789***	-5.253***	-3.504***	-7.327***	-4.234***	-2.885***
	(1.586)	(1.054)	(0.810)	(1.492)	(1.007)	(0.804)
3+ year since High	-6.573***	-3.867***	-2.759***	-5.957***	-3.503***	-2.317**
	(1.865)	(1.243)	(0.957)	(1.716)	(1.161)	(0.928)
1 year since Low	8.167***	7.169***	1.509	4.708**	5.460***	-0.377
	(2.074)	(1.355)	(1.026)	(1.935)	(1.295)	(1.030)
2 year since Low	5.341**	3.171**	2.440**	2.39	1.695	0.881
	(2.119)	(1.399)	(1.068)	(1.951)	(1.314)	(1.048)
3 year since Low	-0.307	2.541	-2.510*	-1.557	1.9	-3.103**
	(2.697)	(1.781)	(1.361)	(2.470)	(1.664)	(1.328)
3+ year since Low	-2.32	-1.044	-1.018	-3.628	-1.827	-1.546
	(2.789)	(1.845)	(1.412)	(2.549)	(1.719)	(1.372)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Hospital Chracteristics	Yes	Yes	Yes	Yes	Yes	Yes
Hospital Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Surgeon Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Sample size is 1469. Standard deviations are in brackets. Standard errors are in parentheses. All standard errors are clustered by surgeon. Hospital characteristics include a dummy indicating not-for-profit status, bed size and a dummy indicating teaching status. *, **, *** Significant at the 10%, 5% and 1% level for a two-tail test.

Table 6: Estimates of Mixed Logit Model

Year in which data were collected	2003			
Period the report is matched to	2005/2 - 2006/1			
	(1)	(2)	(3)	(4)
Surgeon characteristics				
High mortality flag				
Mean	-1.404***	-2.043***	-1.442***	-2.068***
	(0.321)	(0.508)	(0.436)	(0.503)
Standard deviation	0.315	1.337**	0.481	1.381**
	(1.269)	(0.572)	(1.168)	(0.553)
Experience				
Mean	0.085***	0.234***	0.087***	0.236***
	(0.013)	(0.029)	(0.014)	(0.029)
Standard deviation	0.064***	0.155***	0.065***	0.156***
	(0.014)	(0.017)	(0.014)	(0.017)
Experience squared				
Mean	-0.003***	-0.007***	-0.003***	-0.007***
	(0.000)	(0.001)	(0.000)	(0.001)
Standard deviation	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Not rated				
Mean	-0.845***		-0.857***	
	(0.124)		(0.138)	
Standard deviation	0.896***		0.935***	
	(0.243)		(0.260)	
Hospital characteristics				
High mortality flag (HMF)	0.438***	0.522***	1.047**	1.101**
	(0.069)	(0.076)	(0.461)	(0.506)
Teaching	0.257***	0.433***	0.288***	0.453***
	(0.035)	(0.038)	(0.038)	(0.040)
Nonprofit	0.576***	0.674***	0.587***	0.681***
	(0.084)	(0.090)	(0.085)	(0.090)
Number of beds	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Distance	-0.374***	-0.378***	-0.377***	-0.381***
	(0.006)	(0.007)	(0.006)	(0.007)
Distance squared	0.004***	0.004***	0.004***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
HMF x Teaching			-0.548***	-0.450**
			(0.188)	(0.203)
HMF x Number of beds			-0.001	-0.001
			(0.001)	(0.001)
HMF x Distance			0.014	0.022
			(0.019)	(0.021)
HMF x Distance squared			-0.000	-0.001
			(0.000)	(0.000)
Number of observations	527,460	369,401	527,460	369,401

Notes: Estimation is based on patient-level data. Samples include nonrated surgeons under odd-numbered columns and exclude nonrated surgeons under even-numbered columns. The number of observations is the number of patient-alternative pairs. Standard errors are in parentheses. *, **, *** Significant at the 10%, 5%, and 1% level for a two-tailed test.

Table 7: Averages of Patient-Level Random Parameter Estimates

Year in which data were collected	2003			
Period the report is matched to	2005/2 - 2006/1			
	(1)	(2)	(3)	(4)
Surgeon characteristics				
High mortality flag				
Mean	-1.404	-2.042	-1.442	-2.068
Standard deviation	0.015	0.182	0.030	0.192
Experience				
Mean	0.085	0.234	0.087	0.236
Standard deviation	0.025	0.094	0.025	0.095
Experience squared				
Mean	-0.003	-0.007	-0.003	-0.007
Standard deviation	0.001	0.001	0.001	0.001
Not rated				
Mean	-0.844		-0.857	
Standard deviation	0.227		0.245	
Number of patients	9,476	8,245	9,476	8,245

Notes: Estimates for the conditional distribution are calculated based on the associated mixed logit estimates for the population distribution. Samples include nonrated surgeons under odd-numbered columns and exclude nonrated surgeons under even-numbered columns. The last (first) two columns include (exclude) the interaction terms of hospital-level characteristics.

Table 8: Estimates of Mixed Logit Model by Patient Severity

Year in which data were collected	2003			
Period the report is matched to	2005/2 - 2006/1			
Patient severity	Low Severity		High Severity	
	(1)	(2)	(3)	(4)
Surgeon characteristics				
High mortality flag				
Mean	-2.043***	-2.315***	-1.359***	-2.662***
	(0.576)	(0.645)	(0.148)	(0.656)
Standard deviation	1.381**	1.658***	0.058	1.908***
	(0.585)	(0.587)	(1.083)	(0.553)
Experience				
Mean	0.084***	0.218***	0.091***	0.240***
	(0.019)	(0.042)	(0.019)	(0.037)
Standard deviation	0.061***	0.140***	0.069***	0.160***
	(0.019)	(0.025)	(0.019)	(0.022)
Experience squared				
Mean	-0.003***	-0.007***	-0.003***	-0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Standard deviation	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Not rated				
Mean	-1.214***		-0.676***	
	(0.224)		(0.136)	
Standard deviation	1.465***		0.538	
	(0.309)		(0.419)	
Hospital characteristics				
High mortality flag (HMF)	1.065*	1.037	0.994	1.235
	(0.623)	(0.677)	(0.693)	(0.771)
Teaching	0.190***	0.311***	0.430***	0.650***
	(0.050)	(0.053)	(0.057)	(0.062)
Nonprofit	0.321***	0.394***	1.013***	1.140***
	(0.105)	(0.110)	(0.147)	(0.158)
Number of beds	0.000***	0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Distance	-0.359***	-0.362***	-0.405***	-0.408***
	(0.008)	(0.009)	(0.010)	(0.011)
Distance squared	0.004***	0.004***	0.004***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
HMF x Teaching	-0.647**	-0.499*	-0.501*	-0.468
	(0.265)	(0.282)	(0.273)	(0.297)
HMF x Number of beds	-0.001	-0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
HMF x Distance	0.029	0.049*	0.008	-0.004
	(0.025)	(0.027)	(0.029)	(0.033)
HMF x Distance squared	-0.001	-0.001**	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Number of observations	290,067	205,302	233,945	161,852

Notes: Estimation is based on patient-level data. Samples include nonrated surgeons under odd-numbered columns and exclude nonrated surgeons under even-numbered columns. The number of observations is the number of patient-alternative pairs. Standard errors are in parentheses. *, **, *** Significant at the 10%, 5%, and 1% level for a two-tailed test.

Table 9: Estimates of Mixed Logit Model by Emergency Admission Status

Year in which data were collected	2003			
Period the report is matched to	2005/2 - 2006/1			
Patient severity	Nonemergency		Emergency	
	(1)	(2)	(3)	(4)
Surgeon characteristics				
High mortality flag				
Mean	-2.945***	-3.262***	-1.416***	-1.558***
	(0.739)	(0.752)	(0.198)	(0.229)
Standard deviation	2.328***	2.645***	0.063	0.001
	(0.543)	(0.541)	(0.595)	(0.635)
Experience				
Mean	0.086***	0.232***	0.075***	0.168***
	(0.018)	(0.033)	(0.025)	(0.045)
Standard deviation	0.062***	0.148***	0.055**	0.141***
	(0.018)	(0.019)	(0.027)	(0.029)
Experience squared				
Mean	-0.003***	-0.007***	-0.003***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
Standard deviation	0.001***	0.002***	0.002***	0.002**
	(0.000)	(0.000)	(0.000)	(0.001)
Not rated				
Mean	-1.294***		-0.794***	
	(0.236)		(0.180)	
Standard deviation	1.567***		0.920***	
	(0.313)		(0.357)	
Hospital characteristics				
High mortality flag (HMF)	0.453	0.459	7.203***	7.675***
	(0.563)	(0.612)	(2.428)	(2.632)
Teaching	0.288***	0.424***	0.243***	0.447***
	(0.051)	(0.055)	(0.079)	(0.085)
Nonprofit	0.454***	0.541***	0.844***	0.880***
	(0.114)	(0.121)	(0.184)	(0.194)
Number of beds	0.000***	0.000***	0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Distance	-0.350***	-0.353***	-0.433***	-0.435***
	(0.008)	(0.009)	(0.014)	(0.015)
Distance squared	0.004***	0.003***	0.005***	0.005***
	(0.000)	(0.000)	(0.000)	(0.000)
HMF x Teaching	-0.508**	-0.404	-2.518***	-2.625**
	(0.231)	(0.248)	(0.970)	(1.045)
HMF x Number of beds	-0.001	-0.001	-0.004	-0.005
	(0.001)	(0.001)	(0.003)	(0.004)
HMF x Distance	0.044*	0.055**	-0.099	-0.099
	(0.023)	(0.025)	(0.102)	(0.108)
HMF x Distance squared	-0.001**	-0.001**	-0.001	-0.001
	(0.000)	(0.000)	(0.002)	(0.003)
Number of observations	259,990	183,716	129,470	89,468

Notes: Estimation is based on patient-level data. Samples include nonrated surgeons under odd-numbered columns and exclude nonrated surgeons under even-numbered columns. The number of observations is the number of patient-alternative pairs. Standard errors are in parentheses. *, **, *** Significant at the 10%, 5%, and 1% level for a two-tailed test.