Does Experience Make Better Doctors?

Evidence from LASIK Eye Surgeries

Juan M. Contreras
Congressional Budget Office
juan.contreras@cbo.gov

Beomsoo Kim
University of North Carolina Greensboro
b_kim@uncg.edu

Ignez M. Tristao
Congressional Budget Office
ignez.tristao@cbo.gov

August 2007

ABSTRACT

We examine the "learning by doing" hypothesis in medicine using a longitudinal census of laser in situ keratomileusis (LASIK) eye surgeries collected directly from patient charts. LASIK surgery has precise measures of presurgical condition and postsurgical outcomes. Unlike any other surgery, the impact of unobservable conditions on outcomes is minimal. Learning by doing is identified through observations on surgical outcomes over time for each doctor. Our unique data set overcomes some of the major measurement problems in health outcomes, and enhances the possibility of identifying the impact of learning by doing separate from other effects. Our results do not support the hypothesis that doctors' individual learning improves outcomes, but we find strong evidence that experience accumulated by doctors as a group in a clinic significantly improves outcomes.

We are very grateful to the Maryland Population Research Center at the University of Maryland for providing financial support to fund this research. We are also thankful to the administration and doctors of the Clínica Oftalmológica de Antioquia (CLOFAN) for their help and collaboration, especially to Dr. Luis Fernando Pelaez, Dr. Mauricio Jaramillo, and Dr. Leonardo Orjuela for their support during our data collection in Colombia. We also would like to thank William Evans, Seth Sanders, John Rust, and Mark Duggan for helpful discussions and advice during the elaboration of this work. Chris Ruhm, David Ribar, Stephen Holland, Chris Swann, Doug Hamilton and Bob Dennis also provided many helpful comments. All errors are our own. Working papers in this series are preliminary and are circulated to stimulate discussion and critical comment. These papers are not subject to CBO's formal review and editing processes. The analysis and conclusions expressed in them are those of the authors and should not be interpreted as those of the Congressional Budget Office. References in publications should be cleared with the authors. Papers in this series can be obtained at http://www.cbo.gov/publications/.

1. INTRODUCTION

How much does experience enhance a physician's skills and improve outcomes for patients? Although economists have tried hard to empirically determine the existence of "learning by doing" in medicine, isolating its effects from other types of effects has proved elusive.

In this paper, we address this question using a new and unique data set of LASIK refractive eye surgeries, an operation with well-defined eligibility criteria, precise measures of previous conditions and postsurgery outcomes, and minimal postsurgical care. The data we use were taken directly from individual patients' charts and are part of a two-year longitudinal census of LASIK surgery patients from one of the largest ophthalmologic clinics in Colombia. Although our results do not support the hypothesis of doctors' individual learning, we find strong evidence that experience accumulated by doctors as a group improves outcomes.

In the last thirty years, many studies have analyzed the relationship between physician (or hospital) volumes of surgical procedures and patient outcomes. Halm et al. (2002) reviewed 135 papers studying this correlation in the health care industry between 1980 and 2000 and found that around 70 percent of the studies reported a statistically significant association between higher volume and better outcomes. Based on these studies, the *Washington Post* wrote that it is now a virtually sacrosanct belief in the medical world that patients fare better if operated on by a doctor who frequently performs the same operation (October 28, 2003). Researchers often have interpreted this correlation as evidence of a causal relationship that "practice makes perfect." Moreover, the view that "practice makes perfect" provides support for the expansion of regional or specialized medical centers to facilitate learning by doing.[1] But it is often debated whether

---

[1] In this context, regionalization of medical care entails the financing of larger medical centers instead of many small ones. See Luft et al. (1976) and Rathore et al. (2006).

society can get more benefit by receiving expensive care such as critical care or cardiovascular surgical services in specialized centers or by distributing those services among small clinics (Thompson et al. 1994; Menke and Wray 2001).

However, there are other possible explanations for the observed correlation between volume and outcome. One of them is selective referral. Primary care physicians may refer their surgery patients to more skilled surgeons, causing more skilled surgeons to show a higher volume and better outcomes, not because of their experience but because of their innate ability. If selective referral is indeed the cause of this high-volume/better-outcome relationship, then regionalization could increase the price of medical care by reducing competition without improving outcomes. Determining the existence and magnitude of a physician learning curve is at the heart of this issue.

The challenge of empirically measuring a learning curve in medical procedures is twofold: data limitations and hard to measure outcomes. Data limitations prevented previous studies from being able to observe the evolution of an individual doctor's experience over time. Instead, the studies could identify only the annual (or quarterly) volume for each surgeon or hospital. By contrast, we can observe the exact time of each surgery (with a precision of seconds), which enables us to measure a physician's experience directly and precisely from the cumulative number of procedures performed by the doctor at specific points in time.

The difficulty of defining and measuring health outcomes constitutes another challenge. Postoperative mortality is commonly used as an indicator of an adverse outcome because it is accurately recorded. But mortality is an extreme outcome, and relying on it alone is an inadequate means to capture both qualitative information and the full range of outcomes.[2] In addition, the effect of a surgeon's skill is hard to isolate from a patient's underlying conditions

---

[2] For example, morbidity or quality of life might also be important outcomes of a health procedure.

that may be closely related to outcomes.[3] Even in the most detailed data some patients' conditions are not available. And last, patients' outcomes also depend on many other factors such as the quality of the surgical team or of the postsurgical care provided by the hospital staff. Thus it is difficult for researchers to measure the impact of an individual physician's learning curve.

Focusing on LASIK surgery enables us to overcome these limitations. The outcome of this surgery is measured precisely on a detailed scale and is relatively unaffected by patients' unobserved underlying conditions (if patients are eligible for the surgery).[4] In addition, the procedure is performed by only one doctor, so we can measure the effect of each doctor's experience on outcomes. And patients require almost no postsurgical care, which can complicate the measurement of the effect of a physician's experience on outcomes.

This paper is divided into five sections. Section 2 provides a short review of the literature. Section 3 describes the data, the measures of outcome, and our empirical methodology. Section 4 presents and discusses our findings, and section 5 concludes.

2. LITERATURE REVIEW

There has been a great deal of research on the relationship between volume and outcomes in medical care, and most studies have found a positive correlation between the two. The first analysis of the correlation between volume and outcome in health care was done by Luft et al.

---

[3] For example, in coronary surgery some conditions like smoking behavior or history of previous heart disease can affect the outcome of the surgery.

[4] LASIK surgery is not recommended for everyone with poor visual acuity. The presence of subclinical keratoconus, corneal warpage syndrome, irregular astigmatism, or a thin cornea are generally contraindications for refractive surgery. LASIK also is not recommended for patients with an autoimmune disease (e.g., lupus, rheumatoid arthritis) or immunodeficiency (e.g., HIV). Some doctors also do not operate on patients younger than 18 years old or with diabetes. For details see Pallikaris and Siganos (1997) and FDA guidelines on laser surgeries http://www.fda.gov/cdrh/LASIK/when.htm).

(1979).[5] They compared outcomes between low- and high-volume hospitals for twelve surgical procedures and found that for certain operations mortality decreased with increases in hospital volume.

More recently, Halm et al. (2002) reviewed 135 papers studying this correlation in the health care industry between 1980 and 2000 and found that around 70 percent of the studies reported a statistically significant and positive association between higher volume and better outcomes. However, 90 of the studies reviewed by Halm et al. examined patient outcomes using the variation in cross-sectional hospital volume as Luft et al.[6] One shortcoming of this approach is that unobserved hospital characteristics can drive the results. For example, if high-volume hospitals have better technology than low-volume hospitals, then the difference in outcomes may be due to the difference in technology, which are often not well observed in the data. In addition, the volume-outcome relationship analyzed in those studies measured only correlation, not causality, because volume may be endogenous to outcomes—better outcomes may also lead to higher volume.

To overcome the problem of unobserved heterogeneity in cross-hospital comparisons, some researchers have tried different identification strategies. Ho (2002) examined variation across hospitals over time using the cumulative annual volume of percutaneous transluminal coronary angioplasty surgeries as a measure of experience. She found some evidence of learning by doing by observing outcome improvement over time using 13 years of annual volume of coronary surgery for hospitals in California. However, she could not control for the evolution of technology, which could have changed significantly over the 13-year span of the data. As a result, it is unclear from her results whether more experience or better technology improves

---

[5] Learning curves have been studied earlier in other areas. One example is Asher (1956), who studied the cost-quantity relationships in the airframe industry.

[6] This class of studies includes Hughes et al. (1988), Phillips et al. (1995), Jollis et al. (1994), and Kimmel et al. (1995).

outcomes over time. In addition, a hospital is a complex organization and postsurgical outcome depends on multiple factors such as the surgical team, operating surgeon's skill, technology, and postsurgical care. Because Ho could not measure the impact of each of these learning factors separately, the exact source of the improvement in outcomes is uncertain.

Several researchers have used comparisons among surgeons, based on the assumption that learning by doing occurs at the individual level. Birkmeyer et al. (2003) compared outcomes across high- and low-volume surgeons and found that outcomes had a stronger correlation with the volumes of individual surgeons than with volumes based on hospitalwide measures. However, the direction of causation with this correlation is unclear. Did experience improve outcomes (as implied by the learning by doing hypothesis)? Or did better outcomes lead to higher volumes (as implied by the competing hypothesis of selective referral)? The evidence presented in the study cannot distinguish between these two hypotheses.

Recently, a few papers have tried to determine the causal relationship between outcomes and volume using instrumental variables (IV) estimation. Gowrisankaran et al. (2006) focused on the outcome of three surgical procedures using quarterly hospital volume as an explanatory variable. They used the predicted quarterly hospital volume for each procedure as an instrumental variable and assumed that people would go to the closest hospital without selective referral. Their results indicate that for at least two of the three procedures they studied learning by doing plays an important role in explaining differences across hospitals in their risk-adjusted outcomes.

Ramanarayana (2006) tested learning by doing in a surgeon-level study using the annual volume of coronary artery bypass graft (CABG) surgeries in Florida. He used an exogenous increase in CABG volume caused by the exit of surgeons who practiced in the same area a year earlier.[7] His results support the existence of learning by doing in CABG surgery. By following doctors who perform surgeries in multiple hospitals in the same year, he also found that a surgeon's skill transfers across hospitals.

In comparison with previous studies, this paper has several advantages for the study of learning by doing in surgery. While other studies have used data that identify time of surgery only by year, forcing researchers to use annual volume as a proxy for experience,[8] our data identify the exact time of surgery as well as the day, month, and year. Therefore, we can measure experience directly from the first to the *n*th surgery for each surgeon. Moreover, we do not need to be concerned about changing technology, as all the surgeries were performed with the same machine. Using LASIK to analyze surgeons' learning by doing has other important advantages, which we discuss after giving some background on this surgery in the following section.

3. BACKGROUND ON LASIK SURGERY

LASIK is an elective refractive surgery that corrects visual acuity by reshaping the cornea using a special laser.[9] Figure 1 shows how the surgery is performed. First, a surgeon creates a thin flap on the cornea with a special tool called a microkeratome. The flap is folded back, and a laser is

---

[7] Volume increased for the remaining surgeons who took over the surgeries for their colleagues that left the area.

[8] In the most disaggregated unit of data that researchers have used, volume was given by quarters in each year.

[9] LASIK became very popular due to the fast vision recovery and minimal pain that accompany the procedure. Although there are other techniques for refractive surgery, LASIK is the most popular nowadays in most countries including the United States and Colombia, where we obtained our data set. Moreover, Colombia has been at the leading edge of developments in refractive surgeries (LASIK being one of multiple options) since Dr. José Barraquer laid down its theoretical and empirical bases in Bogotá in 1948. He provided doctors with knowledge about how much of the cornea had to be left unaltered to provide a stable long-term result, and created the procedure (called keratomileusis) and instrumentation (including the first microkeratoma) to cut and reshape the cornea. Later technical and procedural developments included the RK (radial keratomileusis) in the 1970s in Russia by Svyatoslav Fyodorov, the development of PRK (photorefractive keratomileusis) in the 1980s in Germany by Theo Seiler, and finally the introduction in the early 1990s by Italian doctor Luccio Burroto and Greek ophthalmologist Ioannis Pallikaris of laser techniques to reshape the cornea. It was only in 1999 that LASIK was approved by the US FDA.

used to remove a certain amount of corneal tissue. Before using the laser, some surgeons clean the cornea under the flap with sterilized drops, while others prefer to leave it as is to avoid injecting moisture into the corneal tissue. These surgeon-specific preferences and habits affect the final ablation[10] since dryness alters the absorption rate of the laser on the cornea. Therefore, each surgeon might use a different duration of the laser pulse for the same patient to get an identical ablation based on his or her own surgical habits. The surgeon pulses the laser on the targeted part of the cornea and then folds the flap down on the cornea.

Outcomes of LASIK surgery depend mostly on two inputs: the machine used (i.e., the technology) and the skill of the ophthalmologist. There are two kinds of technology: hardware and software. In our paper, the hardware technology is fixed because we collected the data from a single machine.[11] However, the software, provided by the manufacturer, was upgraded once and this upgrade, which controls the delivery and fluence of the laser beams, could change the performance of the machine.[12]

There are several ophthalmologic skills involved in LASIK surgery. First, ophthalmologists use intraocular lens (IOL) formulas to calculate how much adjustment the eye needs before they start any surgical procedure. Second, they use a special tool called a microkeratome to make a flap that provides access to the corneal stroma;[13] the ability to make a smooth and clean flap is crucial for the outcome of the surgery as an irregular surface causes unclear vision. Third, doctors need to develop a surgical plan based on the patient's age, degree

---

[10] Laser ablation is defined as the process of removing material from a surface using a laser beam.

[11] LASIK surgical techniques have been evolving rapidly. However, applications of new techniques depend on the use of new machines. For example, eye tracking can improve the precision of the ablation but requires a new machine with eye tracking technology.

[12] Laser beam fluency is defined as the amount of energy per pulse that is distributed over a defined area (mJoules/cm2; Machat et al. 1999).

[13] The stroma is the thickest layer of the cornea and it is the part that is re-shaped in laser eye surgery.

of ametropia,[14] sex, and other factors (Machat et al. 1999). Although, in theory, IOL formulas tell a doctor how to reshape the cornea, in practice getting the exact ablation using a laser is complicated.

In particular, when doctors evaluate patients and make a plan for the surgery, they have to take into account the patient-specific factors listed above as well as their own surgical habits. To incorporate these elements into the surgical plan, doctors have developed an "adjustment rule" called a nomogram.[15] As described by Machat et al. (1999, p. 67), "The process of developing a LASIK nomogram requires four steps: obtaining patient data, formulating an initial nomogram, entering data into the laser's computer and evaluating data and outcomes, and making adjustments based on this information." The manufacturer of the laser machine recommends an initial nomogram based on their test data. However, it is recommended that doctors develop a personalized nomogram based on their own surgical habits. The disadvantage of this approach is that it takes time to accumulate the data needed to formulate the nomogram.

In the clinic where we collected data all the doctors use a single nomogram. They review their surgical data and outcomes regularly and, based on their conclusions, the nomogram is updated accordingly. Updates were done twice during our data period: in December 2003 and May 2004. Because all the doctors use and update the nomogram, it reflects learning as an accumulated experience for the whole clinic.[16]

---

[14] In other words, the degree of an eye abnormality, such as nearsightedness, farsightedness, or astigmatism, resulting from faulty refractive ability of the eye.

[15] Nomograms are not unique to LASIK. The National Cancer Institute defines a nomogram as a mathematical device or model that shows relationships between things. For example, a nomogram of height and weight measurements can be used to find the surface area of a person, without doing the math, in order to determine the right dose of chemotherapy (http://www.cancer.gov/Templates/db_alpha.aspx?CdrID=439410).

[16] Their update is mainly based on their experience as surgeons and not on econometric analysis of the data. Because we know only the month but not the day of the nomogram update, we assume it to be on the 15th. We checked the sensitivity of our results to this assumption by redoing our analysis using the first and the last date of the month. Our results are qualitatively robust.

3.1 Why LASIK?

LASIK surgery offers several advantages for measuring learning by doing. First, as mentioned above, the patient charts contain all the information relevant to the outcome of the surgery as impacts on outcome due to a patient's unobserved underlying conditions are minimal in LASIK surgery compared to other types of surgeries. For example, outcomes of coronary surgeries, the most extensively studied surgical procedure, can be affected by weight, previous myocardial infarction, previous cardiac surgery, peripheral vascular disease, diabetes, renal function, hypertension, angina, dyspnoea (breathlessness), and smoking (National Adult Cardiac Surgical Database Report 1999-2000 (UK); Roques et. al 1999). In most studies, researchers have not been able to observe underlying conditions such as smoking behavior or a history of heart disease.

Second, LASIK is performed in most cases by one operating doctor, unlike many other surgeries. In our data, the operating surgeon is assisted by a nurse and an optometrist, who is responsible for the maintenance of the LASIK machine. Because there is only one optometrist in the clinic and the role of the assisting nurse is minimal, LASIK is an ideal surgery for capturing the surgeon's learning curve. In addition, LASIK does not require hospitalization, meaning that postsurgical care has a relatively limited impact on the final outcome compared to other surgeries that require a patient to stay in the hospital for several days.[17]

Third, LASIK offers a unique possibility of measuring the outcome of a surgery precisely because we can accurately measure pre- and postsurgery eyesight. We explain this in more detail in section 4.2.

---

[17] The mean length of hospital stay after CABG was 8.2 days (Lazar et al. 1995).

4. EMPIRICAL METHODOLOGY

4.1 Data

The data set used in this study is the population of patients who underwent LASIK surgery in a surgical center by the ophthalmologists of the Clínica Oftalmológica de Antioquia (CLOFAN) in Medellín, Colombia. This surgical center opened in July 2003 with a brand new Schwind ESIRIS laser machine and now has the biggest market share in Medellín.[18] Before July 2003, most CLOFAN doctors performed refractive surgery in two other surgical centers in Medellín with older laser technologies. However, because all CLOFAN doctors are shareholders of this new surgical center they have great incentive to use their own laser machine.[19] In addition, at the time of the collection of the data, CLOFAN's Schwind ESIRIS LASIK machine was the best available technology in the city.[20]

CLOFAN provides not only LASIK surgery but also other eye-related treatments. Due to the large number of patients' charts that CLOFAN maintains, we could not collect them all. Patient charts are sorted by patient name and do not have any flag for LASIK surgery. Fortunately, CLOFAN's new surgical center has a log file for each surgery performed using the Schwind ESIRIS laser machine. Using this log, we identified all the patients that underwent LASIK surgery and collected the data from their individual patient charts at the clinic. We did not have access to the files for the other two surgical centers that CLOFAN doctors used before opening the new surgical center.[21]

---

[18] The clinic's market share is estimated at about 57 percent of all refractive surgery procedures done in Medellín.

[19] In fact, CLOFAN doctors as a group need to perform a certain number of surgeries per month to make their equipment pay off its cost and generate a profit.

[20] Surgeons from other clinics can rent the laser machine for their surgeries.

[21] The exceptions are patients that had LASIK surgery with an older machine before having it redone with the new Schwind ESIRIS machine. In those cases, we observe the whole history since it is recorded in the same patient chart.

We have 2 years of data, from July 2003 to August 2005, with a total of 2,042 patients and 3,892 surgery cases (unit of observation is an eye). From the patients' charts, we collected presurgery eyesight measures, the name of the operating surgeon, and all postsurgery follow-up evaluations ("follow-ups"). We also recorded patients' basic demographic characteristics such as gender, age, marital status, place of birth, occupation, neighborhood and city of residence. The patient charts also include detailed information about the surgery: the date, start time (to the precision of seconds), operating surgeon, laser control software version, and whether or not there was any complication during the procedure.

We dropped the 5.98 percent of patients living abroad[22] from our analysis since they can be different from the rest of patients. In particular, most of them did not stay in Medellín long enough for the follow-up evaluations.[23] In addition, we dropped two other categories of patient observations: those that did not have a presurgery refraction measure,[24] and those who failed to return after the surgery for any follow-ups (around 10 percent of LASIK surgery patients). We posit two possible reasons for patients not coming back for follow-ups: either complete satisfaction or strong dissatisfaction. The second case is unlikely as there are no charges for follow-ups or for resurgeries if they are needed; thus, we believe there is only a very small chance of losing observations with adverse outcomes. In addition, we checked whether follow-up observations are correlated with time or with the cumulative number of surgeries for each surgeon. We found that they are not.

The ideal approach would be to observe each surgeon's entire history of LASIK surgeries. However, our observations include only the census of LASIK surgeries performed by doctors

---

[22] The majority of patients living abroad are from the United States. Our conversations with the clinic staff confirmed that some people get their surgery while visiting family in Colombia.

[23] Based on conversations with the doctors, we learned that most of those patients had already booked their airline ticket and could not stay for the clinic follow-ups.

[24] The refraction measure gauges visual acuity using sphere, cylinder, and axis expression. More details are given in section 4.3.

using the new Schwind ESIRIS laser machine. As a result, each surgeon had a different level of experience at the time they performed their first surgery with the new laser machine. To control for these time-invariant surgeon-specific differences we used a surgeon-specific fixed-effect model. A surgeon-specific fixed-effect model can accommodate time-invariant differences such as ability, education, and so forth.

Our data is censored at the time of collection. A patient operated in July 2003 had two years for possible follow ups. However, a patient operated on a day before the data collection had only one day for possible follow ups. We use 200 days as a window to follow the patient throughout the data. We dropped the surgeries that happened less than 200 days from the end of our data period. We choose 200 days (just over 6 months) because 70 percent of repeat surgeries occurred within 200 days of the first surgery. Figure 2 presents the cumulative density function of the duration between the original surgery and resurgery dates. By choosing a shorter window, we would have had more observations but suffered from incomplete outcome measures. On the other hand, a longer window would probably have yielded better value for the final outcome but would have resulted in fewer observations.

The key feature of these data is that we can observe each physician's LASIK surgeries over time and thus test for the existence of learning by doing in this medical procedure. If indeed practice makes perfect, then we should observe a learning curve—that is, we should observe an improvement in the physicians' outcomes as they accumulate more experience.

4.2 Measures of Eyesight

The availability of precise measures of eyesight to define previous patient conditions and surgical outcomes is one of the main advantages of using LASIK surgeries to examine learning by doing in surgery. In this section we describe how ophthalmologists measure eyesight and explain how we use these measures in our analysis.

Two different methods are commonly used to examine visual acuity: the Snellen and the refraction measures. For the first, the patient reads letters of different sizes from a distance of 20 feet. Snellen measures visual acuity on a scale from 20/10 to 20/800, depending on the letter sizes that the patient can read.[25] Although the Snellen measure is informative, ophthalmologists need to perform a refraction assessment in order to determine the refractive error and prescribe a corrective lens.[26] The results are expressed as sphere, cylinder, and axis.[27] The sphere and the cylinder are measured in units called "diopters" and determine the lens prescription; the axis is measured in degrees and signifies the direction of an astigmatism.[28] A negative sphere indicates myopia (near-sightedness) and a positive sphere, hyperopia (far-sightedness). The higher the absolute value of the sphere, the worse the visual acuity. The cylinder reflects the degree of astigmatism.[29] A value of zero indicates a perfect sphere or cylinder, meaning that the patient does not have myopia/hyperopia or astigmatism.[30]

Ophthalmologists use a standard metric called the defocus equivalent to obtain a composite measure based on the refraction measure of the eye.[31] The defocus equivalent is obtained by the following formula:

---

[25] The Snellen chart cannot measure visual acuity that is worse than 20/800.

[26] Refraction refers to how light waves are bent as they pass through the cornea and lens.

[27] The usual expression is sphere = cylinder * axis. The "=" and "*" do not mean the mathematical "equals" or multiplication but are a convention used in the field to express the refraction measure.

[28] The value of the axis is not necessary in order to calculate visual acuity.

[29] Ophthalmologists and optometrists use negative and positive signs as a custom. All the measures in our data set use the negative sign norm.

[30] In order to measure the cornea and obtain values for the sphere, the cylinder, and the axis, the doctor has several options; one is to measure the cornea directly with an automated refractometry; another is to try several combinations of lenses to correct the vision. These exams can be conducted with the eye muscles relaxed using eye drops ("dilated" measures) or without the use of drops.

[31] Spherical equivalence (SE) is used more widely and is calculated as $|\text{cylinder}/2| + |\text{sphere}|$. However, SE can be misleading as it does not fully consider the amount of astigmatism. In addition, it can have a negative sign not

$$\text{Defocus equivalent} = |\text{Cylinder}/2 + \text{Sphere}| + |\text{Cylinder}/2|.$$

If the refraction evaluation is –2.5 of sphere, –3.5 of cylinder, and 180° of axis (which means a myopia of 2.5 diopters and an astigmatism of 3.5 diopters, measured in the 180° axis), the defocus equivalent will be $|(-3.5/2) + (-2.5)| + |(-3.5)/2| = |-4.25| + |-1.75| = 6$. The defocus equivalent for a perfect eye is zero, but any measure between 0 and 0.5 diopters is considered close to perfect eyesight. In our data set we use the defocus equivalent to measure visual acuity because the Snellen test cannot measure high-degree visual acuity problems.

Visual acuity is recorded several times for each patient. Before the surgery, 86 percent of our LASIK surgery patients had a refraction measure of their visual acuity.[32] Because observations with no presurgery refraction measure are not correlated (i.e., associated) with any specific doctor, number of cumulative surgeries, or month, it seemed safe to drop these observations.

After the surgery, doctors typically evaluate a patient's eyesight with the Snellen measure; for patients that demonstrate good eyesight, doctors often do not perform a refraction evaluation. Based on this criterion, 24 percent of cases were evaluated using only the Snellen measure. We used observations for which both Snellen and refraction measures were reported to predict a measure of the defocus equivalent for the observations that have only Snellen measure.[33]

---

associated with any specific meaning. Holladay et al. (1991) proposed the defocus equivalent to overcome these shortcomings of SE.

[32] There are dilated and nondilated measures. For a dilated measure doctors use eye drops to stimulate or prevent the action of the iris muscle. Because only 32 percent of LASIK cases have dilated measures and postsurgery eyesight measures are not dilated, in our data set we use only nondilated measures both for calculations of the defocus equivalent and for pre- and postsurgery evaluations.

[33] A linear regression was used in order to fill out the missing data.The Snellen and refraction measures use different standards for evaluation. Snellen measures how well the patient can see, and refraction measures near- or far-sightedness as well as astigmatism. In addition, Snellen can only measure up to 20/800. In our data set, 60 percent of the observations with a postsurgery Snellen measure but no refraction measure have 20/20 or better eyesight. We ran a regression with age, sex, and Snellen measure as independent variables.

4.3 Measures of Outcomes

We use four outcome measures. The first outcome is the postsurgical defocus equivalent, measured in the last follow-up observed, at least 1 day after the surgery.[34] The second measure is a dummy for the success or failure of the procedure. A defocus equivalent of less than or equal to 0.5 diopters is considered a success, while values higher than 0.5 diopters are considered a failure in the ophthalmology literature.[35] The third outcome measure is an indicator of whether the patient needed at least one repeat surgery.[36] The fourth outcome measure is the number of follow-up visits after the surgery, used as a proxy of how often a patient had to return to the office in order to ensure a satisfactory outcome. Higher numbers of follow-ups are regarded as adverse outcomes as they may indicate complications after the surgery.

Table 1 presents descriptive statistics of the data. The average number of surgeries per doctor through January 19, 2005, with the CLOFAN laser machine is 109. We present patient characteristics in the lower block. The mean age of patients in the sample is 38 years, and 64 percent are female. At the bottom of Table 1, we present the mean and standard deviation of our four measures of outcomes. Mean postsurgery defocus equivalent is 1.07 diopters; 74 percent of surgeries had a defocus equivalent higher than 0.5; 8 percent of patients required resurgery; and on average, patients had three follow-ups.

---

[34] It is worth noting that the attempted outcome is not always a defocus equivalent of zero. For instance, based on the patient's lifestyle (occupation, recreational activities, etc.), age, eyeglass prescription, and accommodation of the eye muscles, the physician may consider that a full correction is not attainable or advisable. Moreover, the physician may decide to specialize one eye for near-sightedness and the other for far-sightedness.

[35] See Waring (2000) for standard measures for reporting refractive surgery outcomes.

[36] It is worth noting that not all surgeries outside ±0.5 diopters required a repeat surgery.

4.4 Econometric Model

The main question of this paper is whether there is learning by doing in LASIK eye surgery. We define learning by doing as an improvement in surgery outcomes due to the ophthalmologist's accumulated experience in performing LASIK procedures. Our empirical strategy aims to identify this effect.

To test for learning by doing effects, we estimate the following equation:

$$Outcome_{ijk} = X_j \beta + L\,\delta + \upsilon_k + \omega_i + \mu_{ijk} \quad (1)$$

Where $Outcome_{ijk}$ is the outcome for the surgery on the eye (i=[*left*,*right*]) of patient j operated on by doctor k. For the defocus equivalent outcome, the best possible outcome has a value of zero; a higher number indicates postoperative myopia/hyperopia and/or astigmatism. For this outcome we consider only the first surgeries in our sample, as a repeat surgery may differ in many ways,[37] although we accumulate all surgeries (first and repeat surgeries) when calculating doctors' experience. X_j is a vector of patient characteristics such as age, sex, and presurgery defocus equivalent. To control for the nonlinearity of presurgery eyesight we include the square of the presurgery defocus equivalent.[38]

L = [log(cumulative surgeries$_k$), dummies for nomogram updates] is a vector of learning effects containing individual learning (represented by the cumulative number of LASIK surgeries for each surgeon k) and group learning (represented by the nomogram updates). We use the logarithm of the cumulative number of LASIK surgeries for each surgeon as the functional form that captures the learning curve and a dummy variable for each nomogram

---

[37] For example, the goal of repeat surgery may differ from that of the initial surgery as it corrects the remaining uncorrected visual acuity. Doctors might therefore choose a different approach for the follow-up surgery. In addition, postsurgery outcome from the initial surgery, used as a presurgery visual acuity for the follow-up surgery, is no longer exogenous.

[38] The results are robust even if we include higher-order terms.

update to identify the clinicwide learning.[39] $\delta = [\ \delta_{\mathrm{INDIV}}\ ,\ \delta_{\mathrm{GROUP}}\ ]$ is a vector that measures the slope of the learning curve. If there is individual learning by doing, doctors should get better outcomes (i.e., defocus equivalent measures closer to zero) in the n^{th} surgery than in the $n^{\text{th}-1}$ surgery. If the experience accumulated by doctors as a group in the surgical center matters, doctors should get better outcomes after the nomogram is updated. We should expect a negative sign in δ when there is individual or group learning since our measures of outcome are defined as adverse.

We include a surgeon-specific fixed effect (υ_k) to eliminate permanent differences across surgeons. Standard errors are clustered by surgeon to take into account any nonlinear surgeon-specific residuals not captured by this fixed effect.

We also include type of surgery fixed effects (ω). The surgical methods are different depending on the refractive error. For example, myopia requires laser ablation to flatten the central cornea, whereas hyperopia requires laser ablation to make the cornea steeper (Machat et al. 1999, p. 17). Our data set comprises seven different surgical methods to treat myopia, myopic astigmatism, hyperopia, hyperopic astigmatism, and astigmatism as well as custom LASIK and a "multizone" technique.[40] For higher-order aberrations, surgeons measure the complete shape of the surface of the cornea to prepare an individually tailored surgery plan.[41]

---

[39] We also tried square root and a quadratic in the vector of learning as a functional form. Results are robust to all these functional forms.

[40] "Multizone" is a technique in which the surgeon divides the cornea into small zones and makes a separate ablation for each zone. It has the advantage of a more accurate ablation. The choice of multizone might not be exogenous. However, we do not observe a time trend for the patients that chose multizone.

[41] This more personalized procedure is called custom LASIK. Higher-order aberrations cannot be corrected by optical devices such as glasses or by conventional LASIK surgery (http://www.seewithlasik.com/docs/custom-lasik.html).

Last, we include a microkeratome fixed effect.[42] To determine whether the type of microkeratome makes any difference in the outcome, we included four different versions of this tool (hansatome, hansatome excelsius, carriazo-barraquer, and pendular) in our data.

Additionally, we controlled for the software version in our regressions. The software for the Schwind ESIRIS LASIK machine and its upgrade were provided by the manufacturer. We include a dummy variable to capture the upgrade installed in the clinic's machine on June 24, 2004.[43]

For the outcome that is measured as a dummy variable (= 1, if the postsurgery defocus equivalent is higher than 0.5 diopters—i.e., a failure), we apply a linear probability model using only the first surgery case for the same reason described above for the defocus equivalent outcome. In the case of the last two outcome measures, repeat surgery and number of follow-ups after surgery, we use the whole sample.

4.5 Random Assignment

Selection across doctors and over time is a possible source of bias.[44] Selection across doctors can occur if some doctors treat more difficult cases compared to others. One possible scenario is that more experienced (or better performing) doctors would be assigned to more serious cases. In this case, outcomes for doctors treating severe cases would be underestimated. Selection over time may occur if a doctor treats more (or fewer) severe cases over time.

---

[42] As mentioned above, a microkeratome is a tool used to make a flap in the cornea that provides the surgeon access to the corneal stroma in order to reshape it. A smooth flap is crucial for the outcome of the surgery because an irregular surface causes unclear vision.

[43] Machat et al. (1999) list operating temperature and humidity as other factors that can affect surgery outcomes. Although we have those variables in our data, they are poorly recorded and therefore not regarded as an important part of the patient charts; for example, records show large variations within a day, which is impossible as the operating room is a temperature- and humidity-controlled space. Therefore, we decided not to use those variables in our analysis.

[44] In our case, two selections are possible: by the patient or by the doctor. We do not differentiate them here.

Selection across doctors and over time depends on how many patients looked for a specific doctor, and how the clinic assigns patients among doctors. Our conversations with receptionists at CLOFAN revealed that most patients visit the clinic without a specific doctor in mind. When a patient arrives at the clinic for a LASIK surgery consultation without requesting a specific doctor, the receptionist assigns the patient randomly based on doctors' availability. Because each doctor is an equal shareholder of the clinic, receptionists assign patients so that each doctor has the same workload.

Nevertheless, LASIK surgeries were performed very disproportionately across doctors. The reason is that, although LASIK surgery is regarded as a general procedure that can be handled by any doctor in the clinic, some doctors specialize in performing specific procedures such as cataract surgery, and as a result have less available time to take LASIK cases. Unfortunately, we cannot observe this effect directly from our data because they include only LASIK patients.

In order to check for selection in our data, we test whether patients' observable characteristics vary over time or across doctors. As mentioned above, one of the big advantages of using LASIK surgery in our study is that unobservable conditions are minimal. To check selection across doctors, we ran a regression with the presurgery defocus equivalent as a dependent variable and the total number of LASIK surgeries for the operating surgeon as an independent variable. If there is no selection across doctors, we will not get a statistically significant coefficient. In fact, the coefficient is –0.0009 with a standard error of 0.0017, which is small and statistically insignificant.

To check for selection over time, we analyzed the changes in presurgery defocus equivalent of all LASIK patients in the clinic (shown in Figure 3 by month of surgery) and did not observe any pattern. We also tested whether presurgery eyesight changes over time or over experience by using a separate regression. Monthly time trend was used as an independent variable and log cumulative number of surgeries was used for experience. We included a surgeon-specific fixed effect for both estimations. The result is statistically insignificant at the 5 percent level, meaning that patients' severity did not change over time or with the individual doctor's level of experience.

5. RESULTS

If there is a learning curve in LASIK surgeries, outcomes will improve either with individual doctors' experience (measured as the logarithm of the cumulative number of surgeries) or with their group experience (measured using the nomogram update). For all our measures of outcome, this translates into a negative slope for increases in the log of cumulative number of surgeries if individual experience matters, or for the nomogram updates if the group experience is relevant, given that better outcomes are associated with smaller numerical values.

In order to illustrate graphically how outcomes evolve over time, Figure 4 presents the average postsurgery defocus equivalent by month. In reading this figure, it is useful to bear in mind that the nomogram was updated in December 2003 and May 2004, and that the bigger the average postsurgery defocus equivalent, the worse the outcome. During the LASIK machine's first month of operation, we observed a high average postsurgery defocus equivalent—that is, more adverse outcomes; thereafter, outcomes remained stable until November 2003, except for a spike of bad outcomes in October 2003. From December 2003 to April 2004 outcomes varied little except for a relatively high number of adverse outcomes observed in March 2004, which occurred because of four surgeries with particularly bad outcomes.[45] Although there is more variation in outcomes after May 2004, the overall levels are lower than those observed before then.

In Table 2, we show the results under various specifications using the defocus equivalent as the outcome variable. In the first column, we include only observations with postsurgery visual acuity measured by refraction. The second column includes the estimated defocus equivalent imputed from the Snellen measure, which causes the number of observations to increase by 50 percent. This increase in turn causes the standard error for the log of the cumulative number of surgeries in the second column to decrease substantially with respect to the first column; at the same time, the point estimate decreases by 0.003 without changing the qualitative results. The sign is negative as we expected, meaning that adverse outcomes decrease with experience, but there is no statistically significant improvement as an individual doctor performs more surgeries using the same technique.

In the third column we include other learning factors—specifically, the two nomogram updates. The coefficient for log cumulative surgeries in the third column is –0.0542 and is imprecisely measured. When a doctor performs 100 additional surgeries, outcomes improve 0.05 percent, which constitutes a tiny improvement.

The coefficients corresponding to the two nomogram updates are presented in the second and third rows respectively. The first nomogram update does not have the expected negative sign and is estimated imprecisely. The coefficient that corresponds to the second nomogram update is negative and significant at the 5 percent level. The second update lowers the postsurgery defocus equivalent by 0.1 diopter, which is equivalent to 10 percent of the average postsurgery defocus equivalent and represents a substantial improvement. Age shows consistent

---

[45] We concluded that these adverse outcome are random after close examination of the data.

significance, meaning that older patients have significantly worse outcomes. If a patient's age increases by 1 year, the posturgery defocus equivalent is higher by 0.01 diopter. If the patient is a woman, the postsurgery defocus equivalent is again 0.1 diopter higher, and this variable shows consistent statistical significance except in the first column.

To control for the nonlinearity of the presurgery eyesight measure, we included a quadratic term in our model. Both terms in this measure are jointly significant, meaning that when patients have particularly poor presurgery eyesight, the postsurgery outcome is likely to be worse. The software upgrade was imprecisely estimated in the third and fourth columns. The choice of microkeratome is not statistically significant in the regression.[46] A microkeratome fixed effect, included in the last column, did not change any of the estimates qualitatively. Throughout the regression we included a surgeon-specific fixed effect, and a type of surgery fixed effect to control for time-invariant factors across surgeons and types of surgeries. Standard errors are clustered by surgeon to capture nonlinear common factors associated with a particular surgeon.

In Table 3, we present the results for all four outcome variables. In the first column, the dependent variable is the postsurgery defocus equivalent, the same variable displayed in the last column of Table 2. In the second column, the outcome variable is whether the postsurgery defocus equivalent is greater than 0.5 diopter (i.e., a failure). When we use postsurgery defocus equivalent as an outcome, a single outlier could have a big effect on the estimates. Using a dummy for success or failure provides a different way of measuring learning that avoids this problem. Individual learning, measured by the logarithm of the cumulative number of surgeries, shows the expected sign but it is statistically insignificant and tiny in magnitude. Group learning, reflected in the nomogram updates, shows qualitatively the same results compared to the first

---

[46] We do not report those estimates in the table in order to save space. They are available upon request.

column. The sign of the first nomogram update is positive but statistically insignificant. However, the second nomogram update lowers the adverse outcome rate by 0.07, which represents a 10 percent improvement and is statistically significant at the 5 percent level.

In the third column we use repeat surgery as the adverse outcome. There are more observations here because we include all the resurgery cases that are excluded in the outcome measures shown in the first and second columns. The repeat surgery rate is 8.2 percent. In this case, the coefficient of individual experience is not precisely measured. Consistent with our previous results, the second nomogram update lowers the repeat surgery rate by 8 percentage points and is statistically significant. Repeat surgery outcome shows the most dramatic change in terms of magnitude after the second nomogram update.

In the last column, we use the number of follow-ups as a measure of outcome. The log of the cumulative number of surgeries and the two nomogram updates shows a negative sign but we cannot reject the null hypothesis of no learning effects given the lack of statistical significance. The rest of the independent variables show stable estimates across different outcome variables.

In Table 4 we sort the doctors based on their volume of surgeries. Because we observe from Tables 2 and 3 that group learning substantially improves the outcome, in Table 4 we want to examine possible heterogeneous effects by surgery volume on various outcomes and also determine whether specific doctors are driving the results. We divide our sample in two groups: doctors who performed more than 100 LASIK cases during our data period and those who performed fewer than 100. As mentioned above, the nomogram is updated in a meeting attended by all the ophthalmologists in the clinic. They review the surgeries performed using the existing nomogram and discuss possible changes to improve outcomes. In this process, it is possible that updates happened based on the opinion of doctors who performed more surgeries.

The upper block of Table 4 repeats the results reported in Table 3 for all doctors, the middle block shows the results for doctors with more than 100 surgeries, and the bottom block shows the results for doctors with fewer than 100 surgeries. The number of observations in the middle and bottom blocks is similar.

In the first column of Table 4, when individual learning is measured by the log of the cumulative number of surgeries, high-volume doctors show bigger point estimates in absolute terms than low-volume doctors, meaning that high-volume doctors learn faster (although none of the estimates are precise). Results in the case of the second nomogram are striking. The magnitude of the coefficient associated with the second nomogram update for high-volume doctors is almost twice as large as that for all doctors in absolute value, and 16 times larger than the coefficient for low-volume doctors. This coefficient is estimated precisely for high-volume doctors but not for low-volume doctors. This means that the second nomogram improves the surgery outcome for high-volume doctors much more than for low-volume doctors, who did not get nearly as much benefit from the nomogram update. This result is not surprising, as high-volume doctors probably play a larger role in defining the nomogram adjustments.

In the second column, where we use failure (= 1 if the postsurgery defocus equivalent is greater than 0.5 diopter) as an outcome measure, we observe the same qualitative results as in the first column. High-volume doctors improve their outcomes after the second nomogram update, the magnitude of the improvement is bigger than that of all doctors, and the results are statistically significant. We cannot reject the null hypothesis of no improvement after the second nomogram update for low-volume doctors. The point estimate for low-volume doctors is one-sixth the estimate for high-volume doctors in absolute value.

When we consider repeat surgery as an outcome in the third column, none of the learning variables are estimated precisely, except for the second nomogram update in the case of low-volume doctors. However, the magnitude of the estimates is almost the same for the high- and low-volume doctors. In the results presented in the last column, where the outcome measure is the number of follow-ups, none of the variables from the learning vector are precisely estimated.

We also want to examine which patients achieve better outcomes due to learning and whether the severity of their presurgery condition affects outcomes. We analyze those cases in Table 5. Outcomes might improve with more experience on the margin, and easy cases are probably not the marginal cases. Therefore, we divide the patients based on the severity of their previous condition, measured by the presurgery defocus equivalent.

The first column in Table 5 is analogous to the first column in Table 3, which contains the whole population. The second column includes patients with a presurgery defocus equivalent lower than the 50th percentile, and the third column shows those above the 50th percentile.[47] In this table, we use postsurgery defocus equivalent as an outcome. If we look at the first row, where the coefficient associated with individual learning is presented, none of the columns show statistically significant estimates. The coefficient associated with the first nomogram update is also estimated imprecisely. In the third row, we observe that the second nomogram update was disproportionately more beneficial to patients with serious initial conditions. As reported in the third column, the coefficient of the second nomogram update is statistically significant for serious cases: The second nomogram update improved the postsurgery defocus equivalent for these cases by 19 percent, which is 7 percentage points higher than that of the whole sample. On the other hand, the easier cases, shown in the second column, did not obtain any statistically significant improvement on the postsurgery defocus equivalent after the second nomogram update. The improvement for the more severe cases, shown in the coefficient in column 3, is almost 4 times larger in absolute value than the improvement caused by the second nomogram

---

[47] We also tried 25th percentile and 75th percentile as cutoffs, but, we could not obtain precise estimates due to the decreased observation numbers.

update to the easier cases shown in the second column. Severe cases constitute 58 percent of the sample.

Table 6 reproduces the analysis presented in Table 5 for three other outcomes. In the case of the outcome measure that takes the value of 1 when the postsurgery defocus equivalent is higher than 0.5 diopter, we do not observe any significant results for individual or group learning. In the case of the repeat surgery outcome, shown in the middle block, individual learning does not reveal any consistent pattern; however, the second nomogram upgrade makes a big difference between the easy and serious cases. The latter improve more than the easy cases and the coefficient is statistically significant at the10 percent level as a result of group learning. In contrast, easy cases did not show any statistically significant improvement after the second nomogram update. In the bottom block the number of follow-ups shows no significant changes in outcomes for either easy or serious cases.

Finally, we explore the possibility that there is some learning from sources other than a doctor's individual experience or from the clinic's group experience. For example, learning could happen over time by reading professional journals. In order to capture general learning we add a monthly time trend to our model and present the results in Table 7. In the first column, the coefficient associated with the second nomogram update increases slightly and is significant at the 10 percent level. When we use failure as an outcome, the second nomogram update is no longer statistically significant after controlling for monthly time trend. However, the coefficient for repeat surgery is still statistically significant and has the same magnitude. For the number of follow-ups, monthly time trend is significant but lowers the number of follow-ups by only 0.04, a 1.5 percent decrease. Using the outcome measure that defines a postsurgical defocus equivalent of 0.5 diopter or higher as a failure, we could not determine if doctors are learning from the second nomogram update or because there is general learning over time. However, results obtained using the other outcome measures were consistent with the previous results and qualitatively robust.

6. CONCLUSIONS

In this paper, we examine the existence of learning by doing in LASIK eye surgeries. We chose LASIK because this procedure has a well-defined outcome measure on which a patient's unobservable conditions have little bearing. In addition, we use a remarkable data set that allows us to observe the evolution of these well-defined outcomes for a group of doctors since they began performing laser surgeries in June 2003.

The distinguishing feature of this paper, in comparison with previous studies, is the use of a longitudinal data set with good measures of doctors' experience and precisely defined medical outcomes. Past studies have used data that identify only annual surgical volumes and often confound measures of outcomes with many unobserved patient conditions; in those studies, it was difficult to isolate the effect of learning by doing from other effects such as selective referral.

The main question addressed in this paper is whether patients' outcomes improve with physicians' experience. We use two different measures of learning in LASIK procedures. First, we measure individual learning by the cumulative number of surgeries for each doctor. Second, we measure group learning by using updates of a nomogram, an adjustment algorithm used in the LASIK surgeries. We did not find evidence that as doctors increase the number of surgeries performed they obtain better outcomes. However, we did find evidence of group learning, as outcomes significantly improved after the second nomogram update.

We used four different measures of outcomes: postsurgery defocus equivalent, a dummy variable for success or failure, repeat surgery, and the number of follow-ups after the surgery. The first two measures of outcome show a consistent 10 percent improvement after the second nomogram update, with a statistical significance of 10 and 5 percent respectively. The other two measures of outcome show a similar improvement at the 5 percent level of statistical significance. Patients with serious cases benefited more as a result of the second nomogram update and doctors with higher volume showed a greater improvement. This may be because the high-volume doctors participate most actively in the updates of the nomogram. We found that group learning plays a substantial role in LASIK surgery performance whereas individual learning does not.

References

Alchian A. (1963). Reliability of progress curves in airframe production. *Econometrica*, 31(4):679–694.

Argote L, Beckman SL, Epple D. (1990). Learning curves in manufacturing. *Science*, 247:920–924.

Asher H. (1956). Cost-Quantity Relationships in the Airframe Industry. RAND Corporation Study No. R-291.

Benkard CL. (2000). Learning and forgetting: The dynamics of aircraft production. *American Economic Review*, 90(4):1034–1054.

Birkmeyer JD, Siewers AE, Finlayson E. (2002). Hospital volume and surgical mortality in the United States. *Annals of Surgery*, 346:1128–1137.

Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, Lucas FL. (2003). Surgeon volume and operative mortality in the United States. *New England Journal of Medicine*, 349:2117–2127.

Condon PI, Mulhern M, Fulcher T, Foley-Nolan A, O'Keefe M. (1997). Laser intrastromal keratomileusis for high myopia and myopic astigmatism. *British Journal of Ophthalmology* 81:199-206

Epstein AM. (2003). Volume and outcome: It is time to move ahead. *New England Journal of Medicine*, 346:1161–1164.

Gillet P, Goldblum K. (2004). Ophthalmic patient assessment. *The Journal of the American Society of Ophthalmic Registered Nurses*. October-December 23-5

Gruber H. (1992). The learning curves in the production of semiconductor memory chips. *Applied Economics*, 24(8):885–894.

Gruber H. (1994). Learning-by-doing spillovers in the semiconductor industry. *Journal of Political Economy*, 102(6):1201–1227.

Gowrisankaran G, Ho V, Town, RJ. (2006). Causality, learning and forgetting in surgery. (draf, Federal Trade Comission, January 2006).

Halm EA, Lee C, Chassin MR. (2002). Is volume related to outcome in health care? A systematic review and methodological critique of the literature. *Annals of Internal Medicine*, 137(6):511–520.

Hewitt M. (2000). *Interpreting the volume-outcome relationship in the context of health care quality*. Washington DC: Institute of Medicine.

Ho V. (2002). Learning and the evolution of medical technologies: The diffusion of coronary angioplasty. *Journal of Health Economics*, 21:873–885.

Ho V. (2004). Certificate of need, volume, and percutaneous transluminal coronary angioplasty outcomes. *American Heart Journal*, 147(3):442–448.

Holladay, Jack T., Michael J. Lynn, George O. Waring III, Mary Gemmill, Gordon C. Keehn, Brooke Fielding. (1991) Archives of Ophthalmology 109: 70-76.

Hughes, Robert G., Deborah W. Garnick, Harold S. Luft, Stephen J. McPhee, Sandra S. Hunt (1998) Hospital Volume and Patient Outcomes: The Case of Hip Fracture Patients. *Medical Care*, 26(11): 1057-1067

Jollis JG, Peterson ED, DeLong EF, Mark DB, Collins SR, Muhlbaier LH et al. (1994) The relation between the volumeof coronary angioplastry procedures at hospitals treating Medicare beneficiaries and short-term mortality. *New England Journal of Medicine*: 331:1625-9

Keogh BE, Kinsman R. National Adult Cardiac Surgical Database Report 1999–2000. The Society of Cardiothoracic Surgeons of Great Britain and Ireland.

Kimmel, S. E., J. A. Berlin and W. K. Laskey (1995) The relationship between coronary angioplasty procedure volume and major complications*. Journal of American Medical Association.* 274(14)

Lazar, Harold L., Carmel Fitzgerald, Stacy Gross, Tim Heeren, Gabriel Aldea, Richard Shemin. (1995) Determinants of length of stay after coronary artery bypass graft surgery. *Circulation* 92:20-24

Luft HS, Bunker JP, Enthoven AC. (1979). Should operations be regionalized? The empirical relation between surgical volume and mortality. *The New England Journal of Medicine*, 301:1364-1369.

Machat J, Slade S, Probst L. (1999). *The Art of LASIK*, 2nd ed. Thorofare NJ: Slack Incorporated.

Menke TJ, Wray NP. (2001) When does regionalization of expensive medical care save money? *Health Services Management Research*, 14(2):116–124.

Pallikaris I, Siganos D. (1997). *LASIK*. Thorofare NJ: Slack Incorporated.

Philips KA, Luft HS, Ritchie JL. (1995) The association of hospital volumes of percutaneous transluminal coronary angioplasty with adverse outcomes, length of stay, and charges in California. *Medical Care*. 33:502-14.

Ramanarayana, Subramanian. (2006) Does practice make perfect: an empirical analysis of learning-by-doing in cadiac surgery. mimeo

Rathore SS, Epstein AJ, Nallamothu BK, Krumholz HM. (2006). Regionalization of ST-segment elevation acute coronary syndromes care: putting a national policy in proper perspective, *Journal of American College of Cardiology*, 47(7): 1346-9

Roques, F, S.A.M. Nashef, P. Michel, E. Gauducheau, C. de Vincentiis, E. Baudet, J. Cortina, M. David, A. Faichney, F. Gavrielle, E. Gams, A. Harjula, M.T. Jones, P. Pinna Pintor, R. Salamon, L. Thulin. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. *European Journal of Cardiothoraccic Surgery*.15: 816-823

Sarrazin MV, Rosenthal GE. (2004). Hospital volume and outcome after coronary angioplasty: Is there a role for certificate of need regulation? *American Heart Journal*, 147(3):383–385.

Thompson DR, Clemmer TP, Applefeld JJ, Crippen DW, Jastremski MS, Lucas CE, Pollack MM, Wedel SK. (1994). Regionalization of critical care medicine: Task force report of the American College of Critical Care Medicine. *Critical Care Medicine,* 22(8):1306–1313.

Waring GO III. (2000). Standard graphs for reporting refractive surgery. *Journal of Refractive Surgery*, 16, July/August.

Weatherly SL. (2002). Testing Visual Acuity with the Jaeger Eye Chart. American Society of Nondestructive Testing. Available on the Internet at http://www.asnt.org/publications/.

Figure 1. LASIK surgery procedure

1-3: Cutting the flap in the cornea with the microkeratome.
4. Folding the flap back.
5. Correcting the corneal tissue through laser ablation.
Source: Allaboutvision.com at: http://www.allaboutvision.com/visionsurgery/lasik.htm

Figure 2. Cumulative probability of resurgeries

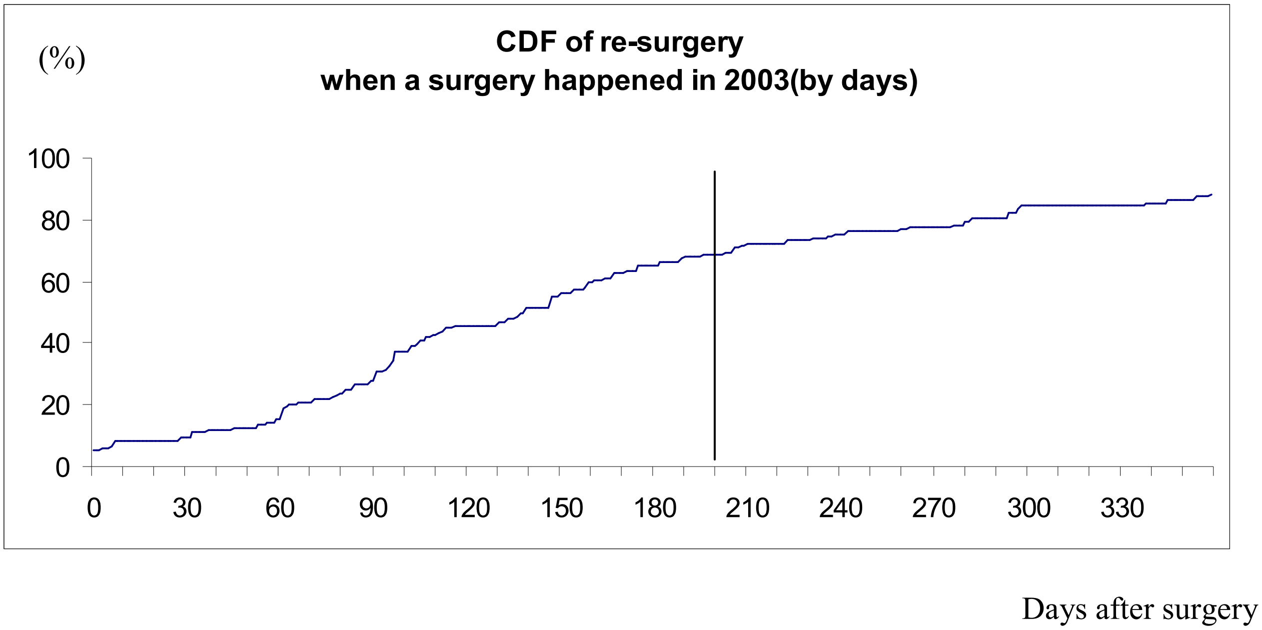

Figure 3. Average presurgery defocus equivalent (by month)

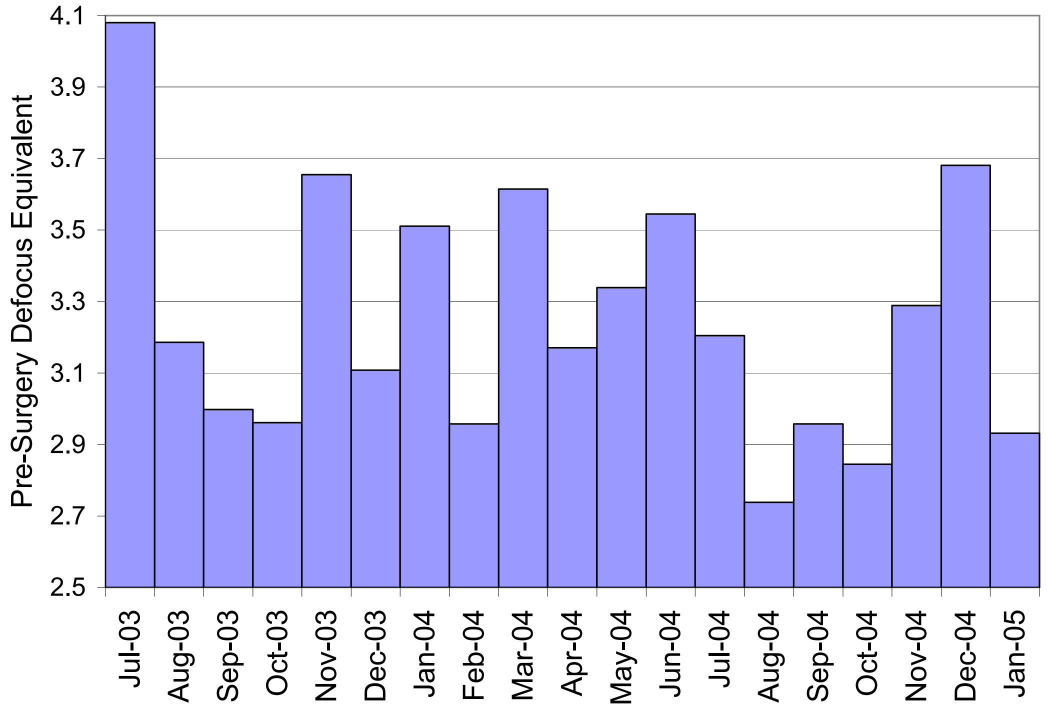

Figure 4. Average postsurgery defocus equivalent (by month)

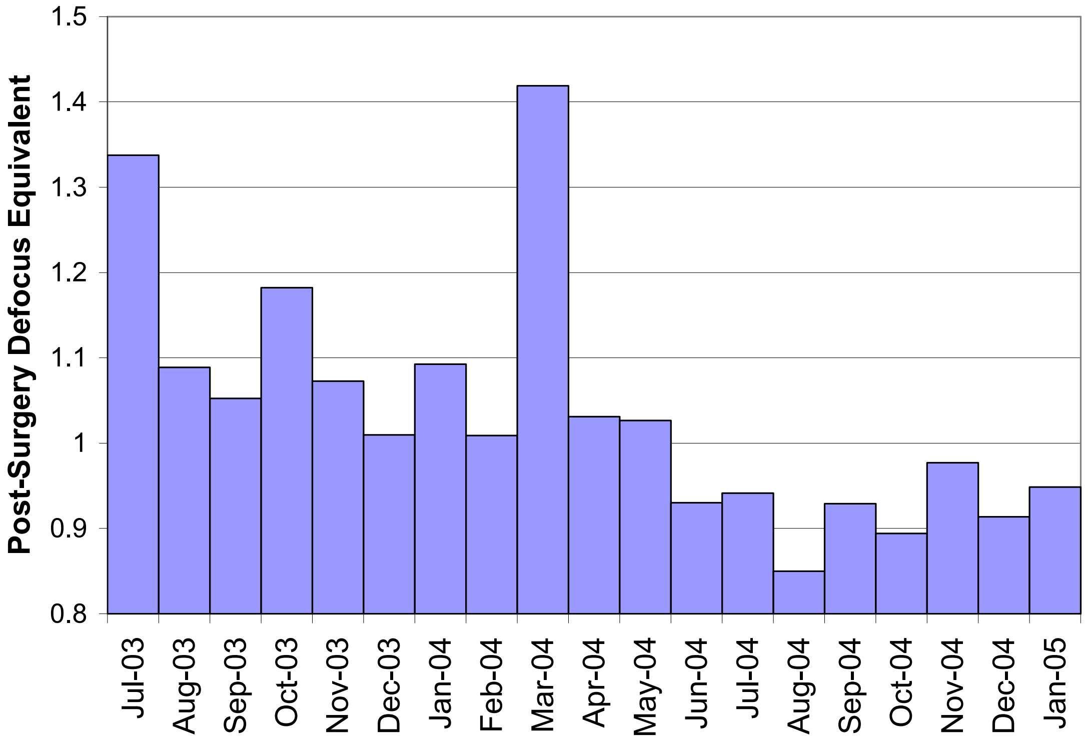

Table 1. Descriptive Statistics

Data period	July 1, 2003 ~ August 8, 2005
Number of doctors	27
Total number of observations	3,892
Observations till Jan. 19, 2005[1]	2,565
Cumulative surgeries by each doctor	108.7 (105.8)
Patients' Characteristics	
Age	38.3 (12.3)
Percent male	35.6 (0.47)
Presurgery defocus equivalent	3.26 (2.29)
25th percentile	1.75
50th percentile	2.50
75th percentile	4.25
Outcome Variables	
Postsurgery defocus equivalent[2]	1.07 (0.83)
percent >0.5[1]	73.7 (0.44)
percent resurgery	8.2 (0.27)
Number of follow-ups	2.68 (1.66)

( ) standard deviations

1. Jan. 19, 2005 is 200 days before the data collection. We keep the window of 200 days throughout data. Therefore, we dropped the observations operated after Jan. 19, 2005.

2. Including only the first surgery. We use the most up-to-date eyesight measure within 200 days from the surgery date.

Table 2. Impact of Learning Vector on Postsurgery Defocus Equivalent

	Without Imputation	With Imputation		
	(1)	(2)	(3)	(4)
Log(cumulative surgeries)	-0.0188	-0.0219	-0.0542	-0.0404
	(0.0393)	(0.0294)	(0.0580)	(0.0663)
1st nomogram update			0.1127	0.1194
			(0.1076)	(0.1029)
2nd nomogram update			-0.1270	-0.1275
			(0.0475)	(0.0465)
Age	0.0087	0.0085	0.0086	0.0084
	(0.0013)	(0.0009)	(0.0009)	(0.0010)
Sex	0.1088	0.1074	0.1068	0.1025
	(0.0660)	(0.0479)	(0.0480)	(0.0471)
Presurgery defocus equivalent	0.0403	0.0235	0.0232	0.0230
	(0.0325)	(0.0264)	(0.0255)	(0.0262)
Presurgery defocus equivalent^2	0.0068	0.0072	0.0072	0.0070
	(0.0031)	(0.0027)	(0.0027)	(0.0027)
Software upgrade	-0.2390	-0.2092	0.0111	0.0225
	(0.1557)	(0.0988)	(0.1266)	(0.1330)
Microkeratome fixed effect	N	N	N	Y
N	1,180	1,729	1,729	1,723

The first column includes observations with postsurgery refraction measure to calculate defocus equivalent. The other columns include imputed defocus equivalent when postsurgery eyesight was measured by Snellen. We use the most up-to-date eyesight measure within 200 days after the surgery. A surgeon-specific fixed effect and type of surgery (myopia, myopic astigmatism, hyperopia, hyperopic astigmatism, astigmatism, custom LASIK, and multizone) fixed effects are included. A microkeratome (hansatome, hansatome excelsius, carriazo-barraquer, and pendular) specific fixed effect is also included in the last column. Robust standard errors are clustered by surgeon and included in parentheses.

Table 3. Impact of Learning Vector on Various Outcomes

	Defocus Equivalent[1]	>0.5 D[1]	Resurgery	Number of Follow-ups
Log(cumulative surgeries)	-0.0404	-0.0303	0.0046	-0.0851
	(0.0663)	(0.0293)	(0.0166)	(0.0866)
1st nomogram update	0.1194	0.0809	-0.0230	-0.0247
	(0.1029)	(0.0555)	(0.0271)	(0.1859)
2nd nomogram update	-0.1275	-0.0734	-0.0813	-0.1014
	(0.0465)	(0.0363)	(0.0374)	(0.1488)
Age	0.0084	0.0017	0.0018	0.0077
	(0.0010)	(0.0012)	(0.0005)	(0.0043)
Sex	0.1025	0.0571	0.0290	0.0843
	(0.0471)	(0.0178)	(0.0164)	(0.0907)
Presurgery defocus equivalent	0.0230	0.0468	0.0268	0.0760
	(0.0262)	(0.0148)	(0.0098)	(0.0374)
Presurgery defocus equivalent^2	0.0070	-0.0015	-0.0011	-0.0037
	(0.0027)	(0.0011)	(0.0008)	(0.0034)
N	1,723	1,723	2,007	2,007

D = diopter

Fixed effects are included for surgeon, type of surgery, and type of microkeratome . Robust standard errors are clustered by surgeon and are shown in parentheses. Linear Probability Model is used for the regression of >0.5 D and resurgery.

[1] We used only the first surgery sample.

Table 4. Impact of Learning Vector on Various Outcomes, by Surgeon Volume

	Defocus Equivalent[1]	>0.5 D[1v]	Resurgery	Number of Follow-ups
	(1)	(2)	(3)	(4)
	All doctors			
Log(Cumulative surgeries)	-0.0404	-0.0303	0.0046	-0.0851
	(0.0663)	(0.0293)	(0.0166)	(0.0866)
2nd nomogram update	-0.1275	-0.0734	-0.0813	-0.1014
	(0.0465)	(0.0363)	(0.0374)	(0.1488)
N	1,723	1,723	2,007	2,007
	Doctors with ≥ 100 volume			
Log(cumulative surgeries)	-0.0618	-0.0472	-0.0086	-0.0394
	(0.1313)	(0.0539)	(0.0199)	(0.1560)
2nd nomogram update	-0.2112	-0.1164	-0.0787	-0.0688
	(0.0906)	(0.0432)	(0.0701)	(0.1832)
N	858	858	991	991
	Doctors with < 100 volume			
Log(cumulative surgeries)	-0.0206	-0.0169	0.0204	-0.0660
	(0.0730)	(0.0304)	(0.0250)	(0.0850)
2nd nomogram update	-0.0127	-0.0183	-0.0791	-0.1157
	(0.0584)	(0.0556)	(0.0364)	(0.2530)
N	865	865	1,016	1,016

D = diopter

See notes for Table 3.

Table 5. Impact of Learning Vector on Defocus Equivalent, by PreSurgery Defocus Equivalent

	Whole Sample	Easy cases [1]	Severe cases [2]
	(1)	(2)	(3)
Log(cumulative surgeries)	-0.0404	-0.0293	-0.0297
	(0.0663)	(0.0595)	(0.1061)
1st nomogram update	0.1194	0.2026	0.0583
	(0.1029)	(0.1399)	(0.1579)
2nd nomogram update	-0.1275	-0.0525	-0.2016
	(0.0465)	(0.0847)	(0.0974)
Age	0.0084	0.0129	0.0062
	(0.0010)	(0.0020)	(0.0020)
Sex	0.1025	-0.0283	0.1764
	(0.0471)	(0.0736)	(0.0724)
N	1,723	719	1,004

Fixed effects are included for surgeon, type of surgery, and type of microkeratome. Robust standard errors are clustered by surgeon and are shown in parentheses.
1. <50th Percentile of Presurgery Defocus Equivalent
2. ≥50th Percentile of Presurgery Defocus Equivalent

Table 6. Impact of Learning Vector on Various Outcomes, by Presurgery Defocus Equivalent

	Whole Sample	Easy cases [1]	Severe cases [2]
	(1)	(2)	(3)
	Outcome: >0.5 D		
Log(cumulative surgeries)	-0.0303	-0.0178	-0.0373
	(0.0293)	(0.0492)	(0.0299)
2nd nomogram update	-0.0734	-0.1276	-0.0289
	(0.0363)	(0.0954)	(0.0639)
Postsurgery defocus equivalent	1.07	0.95	1.07
	Outcome: Resurgery		
Log(cumulative surgeries)	0.0046	-0.0090	0.0018
	(0.0166)	(0.0198)	(0.0284)
2nd nomogram update	-0.0813	-0.0200	-0.0928
	(0.0374)	(0.0272)	(0.0483)
	Outcome: Number of follow-ups		
Log(cumulative surgeries)	-0.0851	-0.1406	-0.1677
	(0.0866)	(0.1707)	(0.1605)
2nd nomogram update	-0.1014	0.1448	-0.0718
	(0.1488)	(0.1991)	(0.2034)

D = diopter

See notes for Table 5.
1. <50th Percentile of Presurgery Defocus Equivalent
2. ≥50th Percentile of Presurgery Defocus Equivalent

Table 7. Impact of Learning Vector on Various Outcomes After Controlling Monthly Time Trend

	Defocus Equivalent[1]	>0.5 D[1]	Resurgery	Number of Follow-ups
Log(cumulative surgeries)	-0.0210	-0.0276	0.0052	-0.0285
	(0.1732)	(0.0228)	(0.0210)	(0.0858)
1st nomogram update	0.1547	0.0858	-0.0217	0.0856
	(0.1008)	(0.0704)	(0.0223)	(0.1779)
2nd nomogram update	-0.0998	-0.0695	-0.0803	-0.0163
	(0.0540)	(0.0463)	(0.0408)	(0.1628)
Monthly time trend	-0.0118	-0.0016	-0.0004	-0.0366
	(0.0113)	(0.0093)	(0.0050)	(0.0182)
N	1,723	1,723	2,007	2,007

D = diopter

See note for Table 3.