Discussion Paper Series
No. 0902
March 2009

Does Experience Make Better Doctors?

Juan m. Contreras, Beomsoo Kim and Ignez M. Tristao

The Institute of Economic Research - Korea University
Anam-dong, Sungbuk-ku, Seoul, 136-701, South Korea, Tel: (82-2) 3290-1632, Fax: (82-2) 928-4948

Does Experience Make Better Doctors?
Evidence from LASIK Eye Surgeries

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

Beomsoo Kim
Korea University
kimecon@korea.ac.kr

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

September 2008

ABSTRACT

In this paper, 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 other types of surgery, the impact of unobservable underlying patient conditions on outcomes is minimal. Individual learning-by-doing is identified through observations of surgical outcomes over time based on the cumulative number of surgeries each surgeon has performed. Collective learning is identified separately through changes in a group adjustment rule determined jointly by all the surgeons through a structured internal review process. Our unique data set overcomes some of the measurement problems in patient outcomes encountered in other studies, and improves the possibility of identifying the impact of learning-by-doing separate from other effects. Our results do not support the hypothesis that the surgeon's individual learning improves outcomes, but we find strong evidence that experience accumulated by surgeons as a group in a clinic significantly improves outcomes.

Key words: learning, experience, LASIK

JEL codes: I10; I12; I18

We are very grateful to the Maryland Population Research Center at the University of Maryland for providing financial support to fund this research. Beomsoo Kim acknowledges the grant support of Korea University. We are also thankful to the administration and surgeons 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. The analysis and conclusions expressed here are those of the authors and should not be interpreted as those of the Congressional Budget Office.

1. INTRODUCTION

How much does experience enhance a physician's skills and improve patient outcomes? Although economists have tried hard to empirically determine the existence of "learning-by-doing" in medicine and discover the mechanism behind this phenomenon, it has proved elusive.

In this paper, we address those questions using a unique data set of LASIK refractive eye surgeries – operations with well-defined eligibility criteria, precise measures of underlying conditions and postsurgery outcomes, and minimal postsurgical care. We took the data directly from individual patients' charts as part of a two-year longitudinal census of LASIK surgeries from one of the largest ophthalmologic clinics in Colombia.

We find strong evidence that experience accumulated by doctors as a group in a clinic substantially improves outcomes in LASIK surgery. By contrast, we find that the role of individual surgeon's experience in improving patient outcomes is insignificant.

The hypothesis of learning-by-doing postulates that workers can learn through experience in the production process, resulting in economies of scale in future production. Learning curves have been studied and documented in many areas, including economics and operations management. Over the last 30 years, many studies in the medical field have analyzed the relationship between physician (or hospital) volume 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. Around 70 percent of the papers reported a statistically significant association between higher volume and better outcomes. Based on this and other studies, the *Washington Post* wrote that it is now a 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] Debates whether society can derive more benefit from expensive care, such as critical care or cardiovascular surgical services in regional or specialized centers, or from distributing those services among small clinics depends on the existence and magnitude of learning-by-doing (Thompson et al. 1994; Menke and Wray 2001). If learning-by-doing exists in sizable magnitude in medicine society might benefit from regionalized facility even after taking into account some cost from difficult access.

There are, however, 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 those surgeons to show higher volume and better outcomes because of innate ability, not more experience. 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 surgeon's learning curve is at the heart of this issue.

The challenge of measuring a surgeon's learning curve is twofold: data limitations and hard to measure outcomes. Data limitations have prevented previous studies from observing the evolution of a surgeon's learning curve over time. Most studies have relied on annual (or quarterly) volume data for each surgeon or hospital[2] because the exact time of each surgery (with a precision of seconds) is not available. However, in our data we measure a surgeon's

[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).

[2] A notable exception is Huckman and Pisano (2006), who use data from the Pennsylvania Health Care Cost Containment Council (PHC4) and include patient-level records for every individual receiving CABG at a hospital in Pennsylvania in 1994 and 1995. That procedure, however, is not performed by a single surgeon but instead by a surgical team, making it difficult to separate individual from collective learning.

experience precisely from the cumulative number of procedures performed till specific points in time.

The difficulty of defining and measuring patient 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 of capturing both qualitative information and the full range of outcomes.[3] In addition, the effect of a surgeon's skill is hard to isolate from a patient's underlying conditions, which may affect outcomes.[4] Even with the most detailed data, some relevant underlying patient conditions are not recorded. And lastly, patient outcomes also depend on other factors, such as the quality of the surgical team and the postsurgical care provided by the hospital staff. Thus, it is difficult for researchers to measure an individual physician's learning curve.

Another challenge is to separate collective learning from individual learning in the measurement process. Hospitals are complex organizations, and surgery involves not only an individual surgeon but also a surgical team and hospital staff. For example, Coronary Artery Bypass Graft (CABG) surgery usually requires a cardiac surgeon, a surgical assistant or a surgical team, a nursing team, an anesthesiology team, and a perfusionist.[5] In those situations, researchers cannot determine if the final outcome is the result of one particular surgeon or the surgical team. Patient outcomes could depend on team coordination, which is acquired through group learning. It is also possible that final outcomes improve through the interactions of

---

[3] For example, morbidity or quality of life might also be important outcomes of a medical procedure.
[4] For example, in coronary surgery some conditions like smoking behavior or history of previous heart disease can affect the outcome of the surgery.
[5] http://health.yahoo.com/respiratory-resources/health-professionals-involved-with-coronary-artery-bypass-graft-surgery/healthwise--ue4707abc.html

surgeons.[6] For instance, a cardiac surgeon might learn from her peers' experience while interacting with them.

LASIK surgery has common characteristics with other surgical procedures. For example, the two key inputs to any surgery are the surgeon's skill and the technology. The surgeon's skill is important in LASIK in developing the surgical plan and executing that plan during surgery. The laser machine is important because the laser ablation depends on the machine's technology.

LASIK surgeries also have particular characteristics that allow us to mitigate the measurement limitations mentioned above. First, the outcome of this surgery is measured precisely and is relatively unaffected by unobserved underlying patient conditions.[7] Second, the procedure is performed by only one surgeon, so we can measure the effect of each surgeon's experience on outcomes. Third, patients require almost no postsurgical care. Such care can complicate the measurement of the effect of a surgeon's experience on outcomes. And fourth, we observe a clear mechanism of group interaction which allows us to measure collective learning separately from individual learning.

This paper is divided into five sections. Section 1 provides an introduction to the subject. Section 2 provides a brief review of the literature. Section 3 describes the data, the measurement of outcomes, and our empirical methodology. Section 4 presents and discusses our findings, and section 5 concludes with our results and assessment of limitations.

---

[6] Huckman et al. (2006) suggest two explanations for the existence of firm specific performance of workers. The first explanation is the complementarity between a worker and the human, physical or organizational assets held by a given firm. They cite as an example the familiarity between members of a team and cite Pisano et al (2001), who note that "well-developed surgical teams are often capable of performing procedures with minimal verbal communication between members"). The second explanation stresses the influence of an individual within a particular organization, for example, a surgeon demanding specific resources from that hospital for her own procedures. Both cases relate to the observation of outcomes from specific teams rather than outcomes from individuals independent from the team.

[7] 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 Food and Drug Administration (FDA) guidelines on laser surgeries http://www.fda.gov/cdrh/LASIK/when.htm).

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 outcomes in health care was done by Luft et al. (1979).[8] They compared outcomes from low- and high-volume hospitals for 12 surgical procedures and found that for certain operations mortality decreased with increased hospital volume.

More recently, Halm et al. (2002) reviewed 135 papers studying this correlation in the health care industry between 1980 and 2000. Around 70 percent of the papers reported a statistically significant and positive association between higher volume and better outcomes. Ninety of the papers reviewed by Halm et al. examined patient outcomes using the variation in cross-sectional hospital volume used by Luft et al.[9] One shortcoming of this approach is that unobserved hospital characteristics can affect 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 technology, which is often not factored into the data. In addition, the volume-outcome relationship analyzed in these papers measured only correlations, not causality, because volume may be endogenous to outcomes -- better outcomes may lead to higher volume. It is important to note that the volume-outcome relationship recognized in many papers as correlation not as causality.

[8] Learning curves have been studied in other areas. Some examples are Wright (1936) and Asher (1956), who studied the cost-quantity relationships in the aircraft industry. In economics, the first theoretical model of this kind was constructed by Arrow (1962).
[9] This class of studies includes Hughes et al. (1988), Phillips et al. (1995), Jollis et al. (1994), and Kimmel et al. (1995).

To overcome the problem of unobserved heterogeneity in cross-hospital comparisons, some researchers have employed different identification strategies. Ho (2002) examined variation within one hospital over time using the cumulative annual volume of percutaneous transluminal coronary angioplasty surgeries as a measure of experience. She found evidence of learning-by-doing by observing outcome improvements over time using 13 years of data tracking annual volumes of coronary surgery for hospitals in California. Unfortunately, she could not control for changes in technology, which would 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 outcomes over time. In addition, a hospital is a complex organization and postsurgical outcomes depend 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 factors separately, the exact reason for 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 volume of individual surgeons than with hospital-wide volume. However, the direction of causation is unclear. Did surgeon experience improve outcomes (as implied by the learning-by-doing hypothesis)? Or did better outcomes lead to higher volume (as implied by the competing hypothesis of selective referral)? Unfortunately, the evidence presented in the study cannot distinguish between these two hypotheses.

Recently, a number of 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 a variable and assumed that people would go to the closest hospital without a selective referral. Their results indicate that for at least two of the three procedures learning-by-doing played an important role in explaining differences across hospitals in their risk-adjusted outcomes.

Huckman et al (2006) used patient mortality from CABG surgery to show that the quality of a surgeon's performance at a particular hospital improves significantly with increases in procedure volume at that hospital, but does not significantly improves with his volume at other hospitals. In a recent study, Pisano et al. (2001) provided important insights into the learning process of hospitals by analyzing 660 patients in 16 institutions that implemented a new operation for minimally invasive cardiac surgeries. The study showed that cumulative experience is a significant predictor of learning and that the slope of the learning curve varies significantly across organizations.[10] They suggested that the underlying organizational process is the key to explain the differences. In a follow-up study, Edmonson et al. (2001) examined qualitatively how different hospital teams implemented minimally invasive cardiac surgery technology.

Because the surgical procedures analyzed in previous studies were performed by a team of surgeons, it is difficult to separate individual learning from collective learning. For example, the studies by Huckman et al. (2001) and Pisano et al. (2006) examined a particular medical procedure where the surgery is performed by a team of doctors and nurses, and their outcomes may reflect the effect of the team instead of the effect of the surgeon.[11]

In comparison with previous studies, this paper offers several advantages for the study of surgical learning-by-doing. While other studies have used data that identify time of surgery only

---

[10] Pisano et al. (2001) uses procedure time (time required to perform a surgery) as their measure of learning. Unfortunately, this is not a health outcome. They tried to find a correlation between procedure times and higher postoperative patient mortality rates, but they were not successful because of the low mortality rate in the sample due to the small sample size.

[11] They test for the influence versus familiarity of the surgeon with the surgical team. They conclude that familiarity is more important statistically than influence. However, the nature of the surgery and the available data do not give enough statistical power to differentiate between those two explanations

by year – forcing researchers to use annual volume as a proxy for experience[12] – our data identify the exact time of surgery as well as the day, month, and year. Therefore, we can measure individual experience directly from the first to the *n*th surgery for each surgeon. Moreover, we do not need to be concerned about changing technology because the most critical technology – the laser machine – did not change during the study period. A group adjustment rule, which was shared by all surgeons and updated over time following an explicit discussion among surgeons, allows us to measure collective learning separately from individual learning. Using LASIK to analyze surgeons' learning-by-doing has other important advantages, which we will discuss after giving background on this surgery in the following section.

3. BACKGROUND ON LASIK SURGERY

LASIK is an elective refractive surgery that improves visual acuity by reshaping the cornea using a special laser.[13] 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.[14] The flap is folded back, and a laser is used to remove a certain amount of corneal tissue. In this section we explain why LASIK is a good procedure for looking at individual and collective learning.

---

[12] Except for Pisano et al. (2001), in most studies the most disaggregated unit of data that researchers have used is quarterly volume of procedures in each year.

[13] 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 FDA.

[14] 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.

3.1 Why LASIK?

LASIK surgery shares many common characteristics with other surgical procedures. For example, a surgeon's skill and the technology are two of the most important factors determining patient outcomes. A surgeon's skill includes the diagnosis of the patient condition as well as the appropriate surgical plan. On top of these common characteristics 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 because effects from unobserved underlying patient 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 or history of heart disease.

Second, LASIK offers a unique opportunity to measure the outcome of a surgery precisely because we can accurately measure pre- and postsurgical eyesight. We explain this in more detail in section 4.2.

Third, LASIK is performed in most cases by one surgeon, unlike the case in 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 individual learning curve. In addition, LASIK does not require hospitalization, meaning that postsurgical care has a limited impact on the final outcome, compared to other surgeries that require a hospital stay of several days.[15]

Collective learning may occur through interaction, coordination, specialization or knowledge spillover (for example, information exchange or shared experience). In our study we are able to measure collective learning separately from individual learning by taking advantage of the adjustment rule determined jointly by the surgeons. The collective learning mechanism we observe occurs through sharing information and discussing the outcomes in regular meetings and updating the group adjustment rule based on the review of the previous outcomes.

3.2 Factors affecting outcome in LASIK surgery

Outcomes of LASIK surgery depend mostly on two inputs: technology and human capital. Human capital can accumulate through individual learning by the surgeon and collective learning by the clinic.

3.2.1. Technological factors affecting outcomes in LASIK surgeries

Two kinds of technology affect LASIK surgery: hardware and software. In our paper, the hardware used is a Schwind ESIRIS laser machine.[16] The software, provided by the manufacturer, controls the delivery and fluence of the laser beams that reshape the cornea. The software was upgraded once to enhance the performance of the machine.[17] We control for the software version update in our analysis, and test whether this technological improvement affected patient outcomes.

[15] The mean length of hospital stay after CABG was 8.2 days (Lazar et al. 1995).
[16] 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.
[17] 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).

3.2.2. Human capital factors affecting outcomes in LASIK surgeries

Human capital can accumulate at the individual and collective levels. At the individual level, LASIK surgery involves several ophthalmologic skills. First, a surgeon needs to develop a surgical plan based on the patient's age, degree of ametropia,[18] sex, and other factors (Machat et al. 1999). Second, surgeons use a special tool called a microkeratome to make a flap that provides access to the corneal stroma.[19] The surgeon's skill in making a smooth, clean flap is crucial for the surgery because an irregular surface causes unclear vision. The choice of microkeratome type reflects the surgeon's learning because some of the microkeratomes (hansatome, hansatome excelsius, carriazo-barraquer, and pendular) may perform better than others depending on the type of eyesight correction required.

After folding the flap and before using the laser to remove the corneal tissue, 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[20] since dryness alters the absorption rate of the laser on the cornea. Therefore, different surgeons might use different surgical plans for the same patient to get an identical ablation based on his or her own surgical habits (see Figure 1).

Collective human capital accumulation is the second human capital factor that may affect LASIK outcomes. LASIK technology requires an adjustment rule to translate the surgical plan into laser machine parameters. This adjustment rule is known as a nomogram.[21] The

[18] In other words, the degree of an eye abnormality, such as nearsightedness, farsightedness, or astigmatism, resulting from faulty refractive ability of the eye.
[19] The stroma is the thickest layer of the cornea and it is the part that is re-shaped in laser eye surgery.
[20] Laser ablation is defined as the process of removing material from a surface using a laser beam.
[21] 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

manufacturer of the laser machine provides an initial adjustment rule based on its test data. However, it is recommended that doctors develop their own adjustment rule based on their surgical habits and the environmental conditions of the clinic. The disadvantage of this approach is that it takes time to accumulate the data needed to formulate the adjustment rule. 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 surgeons from the clinic where we collected the data use a single adjustment rule. They have a structured process to determine this adjustment rule and share information. They meet regularly to review and discuss patient outcomes, and if necessary, update the adjustment rule. The surgeons changed the adjustment rule twice during the data period we analyzed. This adjustment rule reflects the group cumulative experience due to interaction among surgeons who share the same adjustment rule.

4. EMPIRICAL METHODOLOGY

4.1 Data

The data set used in this study is the population of patients who underwent LASIK surgery in the Clínica Oftalmológica de Antioquia (CLOFAN) in Medellín, Colombia. This clinic opened in July 2003 with a brand new Schwind ESIRIS laser machine and now has the largest market share of LASIK surgeries in Medellín.[22] Before July 2003, the surgeons at CLOFAN performed refractive surgery at two other clinics in Medellín that had older laser technologies. Once the

---

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).

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

new surgical center opened, the surgeons performed all of their LASIK surgeries there. CLOFAN surgeons have an incentive to use their own laser machine because they are all shareholders of the clinic.[23] In addition, the new Schwind ESIRIS machine is the best available technology in the city.

For our study, we are using two years of data from the clinic, from July 2003 to August 2005, encompassing a total of 4,009 surgery cases (surgery on one eye is considered one case). From the patient charts, we collected presurgery eyesight measures, the name of the operating surgeon, and all postsurgery follow-up evaluations. 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 information about the surgery: the date, starting time (to the precision of seconds), operating surgeon and laser control software version. The group adjustment rule was updated in December 2003 and again in May 2004.[24]

We dropped the 6.14 percent of patients living abroad from our analysis because most of them did not stay in Medellín long enough for the follow-up evaluations.[25] In addition, we dropped three other categories of patients: those who had LASIK surgery before (341 cases), those who failed to return after the surgery for follow-ups (337 cases or around 10 percent of LASIK surgery patients) and the optimized refractive keratectomy (ORK) cases (245 cases).

---

[23] In fact, CLOFAN surgeons need to perform a certain number of surgeries per month to pay the equipment costs and generate a profit.

[24] Because we know the month but not the day of the nomogram update, we assume it was on the 15th. We checked the sensitivity of our results by redoing our analysis using the first and the last days of the month. Our results are qualitatively robust.

[25] The majority of patients living abroad were from the United States. Our conversations with the clinic staff confirmed that some patients underwent surgery while visiting family in Colombia. We also learned that most of those patients had already booked their airline ticket and could not stay for the clinic follow-ups.

We did not include these cases for the following reasons. A repeat surgery might be very different from a first surgery, such as the presurgery condition, the goal of the outcome, etc.[26] 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 a 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. ORK is a LASIK surgery based on custom-treated ablation. Because the surgeon knows the complete map of the cornea, the ORK procedure uses a different set of surgeon skills when compared to other LASIK surgeries; also, ORK does not need any adjustment rule. For these reasons, we did not include ORK cases in our analysis.[27]

The ideal approach would be to observe each surgeon's entire history of LASIK surgeries. However, our observations include only the LASIK surgeries performed by the surgeons at CLOFAN using the Schwind ESIRIS laser machine. As a result, each surgeon had a different level of experience at the time he or she performed the first surgery with the new laser machine. To control for these time-invariant surgeon-specific differences, we used a surgeon-specific fixed-effect model that could accommodate time-invariant differences such as ability, education, and so forth.

Our data were censored at the time of collection. For example, a patient operated on in July 2003 had two years for possible follow-ups. However, a patient operated on a day before the data collection began had only one day for possible follow-ups. We use 200 days as a

---

[26] For example, the goal of repeat surgery could be to correct the remaining uncorrected visual impairment. Therefore, doctors might choose a different approach for the follow-up surgery. In addition, the postsurgery outcome from the initial surgery is used as the presurgery visual acuity for the follow-up surgery; therefore, it is no longer exogenous.

[27] We count ORK when we calculate the cumulative number of surgeries. If we include the ORK cases for the analysis of individual learning, results do not change.

window to follow the patient. We dropped the surgeries that occurred less than 200 days from the end of our data period. We chose 200 days (just over 6 months) because 70 percent of repeat surgeries occurred within 200 days of the first surgery. Figure 3 presents the cumulative density function of the duration between the original surgery and resurgery. If we chose a shorter window, we would have had more observations but incomplete outcomes. On the other hand, a longer window would probably have yielded better value for the outcome but would have resulted in fewer observations.

Our data allowed us to test for the existence of learning-by-doing and to differentiate between individual learning and collective learning. We tested for individual learning by examining surgeons' outcomes as they accumulated LASIK surgeries over time. We also tested for collective learning by examining whether the changes in the laser adjustment rule made by the surgeons improved average patient outcome.

4.2 Measures of Eyesight

A main advantage of using LASIK surgery to examine learning-by-doing is the availability of precise measures of eyesight to define previous patient condition and surgical outcomes 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.[28] Although the Snellen measure is informative, ophthalmologists

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

need to perform a refraction assessment in order to determine the refractive error and prescribe corrective lens.[29] The results are expressed as sphere, cylinder, and axis.[30] 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 astigmatism.[31] 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.[32] A value of zero indicates a perfect sphere or cylinder, meaning that the patient does not have myopia/hyperopia or astigmatism.[33]

Ophthalmologists use a standard metric called the defocus equivalent to obtain a composite measure based on the refraction measure of the eye.[34] The defocus equivalent is obtained by the following formula:

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

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 astigmatism of 3.5 diopters, measured on 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. In our data set we use the defocus equivalent to measure visual acuity because the Snellen test cannot measure severe visual acuity problems.

---

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

[30] 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.

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

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

[33] In order to measure the cornea and obtain values for the sphere, the cylinder, and the axis, the surgeon has several options; one is to measure the cornea directly with an automated refractometer; 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.

[34] Spherical equivalence (SE) is used more widely and is calculated as |cylinder/2| + |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 associated with any specific meaning. Holladay et al. (1991) proposed the defocus equivalent to overcome these shortcomings of SE.

Visual acuity is recorded several times for each patient. After the surgery, surgeons typically evaluate a patient's eyesight with the Snellen measure; for patients that demonstrate good eyesight, surgeons often do not perform a refraction evaluation. Based on this criterion, 21 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 used only the Snellen measure.[35]

4.3 Measures of Outcomes

We use three outcome measures. The first outcome is the postsurgical defocus equivalent, measured in the last follow-up observed within 200 days after the surgery. The second measure is a dummy for the success or failure of the procedure. We consider a surgery to be successful if the defocus equivalent is within ±1.0 diopters of the desired result.[36] The third outcome measure is an indicator of whether the patient needed at least one repeat surgery.[37]

Table 1 presents descriptive statistics of the data. The average number of surgeries per doctor through January 19, 2005, with the Schwind ESIRIS laser machine is 98. The mean age of patients in the sample is 39 years, and 70 percent are female. At the bottom of Table 1, we present the mean and standard deviation of our three measures of outcomes. Mean postsurgery defocus equivalent is 1.1 diopters; 42.9 percent of surgeries had a defocus equivalent higher than 1.0 diopters; 8.4 percent of patients required resurgery. There are five different types of procedures based on the presurgery eyesight. Myopia, hyperopia, myopic astigmatism,

[35] 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, had 20/20 or better eyesight. We ran a regression with age, sex, and Snellen measure as independent variables.
[36] Waring (2000) discusses standard measures for reporting refractive surgery outcomes.
[37] It is worth noting that not all surgeries outside ±1.0 diopters of the desired result required a repeat surgery.

hyperopic astigmatism and mixed astigmatism.[38] The average presurgery eyesight measure is 3.1 diopters for myopia and 4.3 diopters for myopic astigmatism. The presence of astigmatism increases the defocus equivalent. The most common correction is the myopic astigmatism (40 percent of LASIK surgeries). Hyperopia cases show significantly worse postsurgery eyesight measured in defocus equivalent than myopia cases. This occurs in part because patients with hyperopia tend to be older, and in part because hyperopia is harder to correct than myopia. Although hyperopia and mixed astigmatism cases had a worse outcome measured in defocus equivalent, they had lower resurgery rates than other type of presurgery conditions such as myopia.

4.4 Econometric Model

The main questions addressed by this paper are whether there is learning-by-doing in LASIK eye surgery and whether learning-by-doing happens individually, collectively, or both. We define individual learning-by-doing as an improvement in surgery outcomes due to the surgeon's accumulated experience in performing LASIK procedures. We define group learning as an improvement in surgery outcomes due to changes in the group adjustment rule. Our empirical strategy aims to identify and measures these effects. 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 surgeon 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. We

[38] Mixed astigmatism means that the two curvatures in the cornea defined by the sphere and the cylinder measures are irregular and have meridians placed differently than the normal (90º and 180º).

consider only the first surgeries in our sample, although we accumulate all surgeries (first and repeat surgeries) when calculating surgeon 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.[39]

In equation (1), L = [log(cumulative surgeries$_k$), dummies for versions of the group adjustment rule] is a vector of learning effects containing individual learning (represented by the cumulative number of LASIK surgeries for each surgeon *k*) and collective learning (represented by the versions of the group adjustment rule). 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 version of the group adjustment rule to identify the collective learning.[40] δ = [δ_{INDIV} , δcollective] is a vector that measures the slope of the individual learning curve and the contribution of each version of the group adjustment rule to the final surgery outcome. If there is individual learning-by-doing, surgeons should get better outcomes (i.e., defocus equivalent measures closer to zero) in the n^{th} surgery than in the $n\text{-}1^{th}$ surgery. If the experience accumulated by surgeons as a group in the clinic matters, surgeons should get better outcomes after updates in the group adjustment rule. We should expect a negative sign in δ when there is individual or collective learning since our measures of outcome are defined as adverse.

The group adjustment rule is implemented using an Excel spreadsheet. The inputs are the surgical plan and the outputs are the machine parameters. The initial version of the group adjustment rule came with the machine, and it corrected the cylinder and the sphere in all the astigmatism, myopia and hyperopia cases. The first update of the group adjustment rule (second

[39] The results are also robust to higher-order terms.
[40] We also tried the square root and a quadratic functional form in the vector of learning. Results are robust to all these functional forms.

version) was implemented only for the astigmatism cases. The second update of the group adjustment rule (third version) adds more correction for the astigmatism cases.

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 procedure 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).

Additionally, we control for the software version in our regressions. The software for the Schwind ESIRIS laser 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.[41]

4.5 Random Assignment

Selection across surgeons and over time is a possible source of bias.[42] Selection across surgeons can occur if some treat more difficult cases than others. One possible scenario is that more experienced (or better performing) surgeons would be assigned to more serious cases. Outcomes for surgeons treating severe cases would be underestimated. However, selection across surgeons is not an issue for this paper because we do not use variation across surgeons.[43] Selection over

[41] 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 because they are not regarded as an important part of the patient charts by the surgeon who is the usual record keeper; 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.

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

[43] We checked selection across surgeons by running 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 surgeons, 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.

time is still a source of bias and it may occur if a surgeon treats more (or fewer) severe cases over time.

In order to check for selection in our data, we test whether patients' observable characteristics vary over time by surgeon. As mentioned above, one of the advantages of using LASIK surgery is that unobservable underlying conditions are minimal. To check for selection over time, we test whether presurgery eyesight changes over time or with the surgeon's experience by using a separate regression. The monthly time trend and the log cumulative number of surgeries were the key independent variables for each analysis, and we also include a surgeon-specific fixed effect. The result is statistically insignificant at the 5 percent level, meaning that the severity of the patients' condition did not change over time or with the individual surgeon's level of experience.

Our conversations with surgeons and managers at CLOFAN also revealed that a patient is assigned almost randomly to the surgeon. When a patient contacts the clinic for consultation without requesting a specific surgeon which is most of the cases, the receptionist assigns the patient randomly based on availability. Because each surgeon is an equal shareholder of the clinic, the receptionist assigns patients so that each surgeon has the same workload. Nevertheless, LASIK surgeries were performed disproportionately across surgeons. The reason is that, although LASIK surgery is regarded as a general procedure that can be handled by any surgeon in the clinic, some specialize in performing certain procedures, such as cataract surgery, and as a result have less time available for LASIK cases. Unfortunately, we cannot observe this effect directly because our data include only LASIK patients.

5. RESULTS

If there is a learning curve in LASIK surgeries, outcomes will improve either with individual surgeon's experience (measured as the logarithm of the cumulative number of surgeries) or with their group experience (measured using the changes in the group adjustment rule). For all our measures of outcome, this translates into a negative slope for both the log of cumulative number of surgeries and the changes in the group adjustment rule, given that better outcomes are associated with smaller numerical values.

In order to illustrate graphically how outcomes evolve over time, Figure 3 presents the average pre- and postsurgery defocus equivalent by calendar month. In reading this figure, it is useful to bear in mind that the group adjustment rule was updated in December 2003 and May 2004, and that the larger the average postsurgery defocus equivalent, the worse the outcome.

This figure shows two important things: first, we observe no significant variation or pattern by the average presurgery defocus equivalent over the data period; and second, we observe a decrease in the average postsurgery defocus equivalent. In fact, with the exception of three isolated months (July 2003, March 2004 and December 2004), the average presurgery defocus equivalent fluctuated mostly between 3 and 3.5 diopters. During the LASIK machine's first month of operation, we observe a high average postsurgery defocus equivalent -- that is, more adverse outcomes, although average patient presurgery defocus equivalent was also high in the first month. After that, outcomes improved slightly and remained stable until October 2003; thereafter, there was a slight decrease in average postsurgery defocus equivalent until the last month of our sample, with the exception of March 2004. The average postsurgical defocus equivalent is particularly high in March 2004 because of four surgeries with bad outcomes. Although there is more variation in outcomes after May 2004, the overall level is lower after May 2004 than before.

In Table 2, we show the impact of individual learning (measured by the log of cumulative surgeries) on postsurgery eyesight (measured by defocus equivalent). In the first column, we only include the log of cumulative surgeries as an independent variable. As expected, the sign is negative, which means that as surgeons perform more surgeries, their patients' postsurgery eyesight gets closer to perfect. The magnitude of the coefficient implies that 1% increase of the number of surgeries would improves the outcome by 0.001 diopters, a minuscule though statistically significant improvement at 1% confidence level. In the second column, we add demographic characteristics as controls. In the third and fourth columns, we include presurgery eyesight to control for the patient's underlying conditions. In column four, the coefficient for hyperopia cases are positive and statistically significant, meaning that patients with hyperopia have worse outcomes when compared to patients with myopia. As we add more controls, the point estimate for individual learning decreases (in absolute value) from 0.101 to 0.084. This constitutes a very small improvement. In the final column, we control for technological change reflected by the installation of a new version of the software in the same machine. The laser machine software version was upgraded once during our data period.[44] The coefficient for the software update dummy is small and statistically insignificant. However, once we control for this technological change, the individual learning coefficient decreases by 40 percent in terms of magnitude and also becomes statistically insignificant. As a result, we conclude that there is no statistically significant individual learning in LASIK surgery and controlling for technological change can be important.

삭제됨: one more surgery

삭제됨: One hundred additional surgeries, which in our sample take one and half years on average to accomplish, will improve the outcome by 0.08 diopters.

In all regression specifications we include a surgeon-specific fixed effect, and a type of surgery fixed effect to control for time-invariant factors across surgeons and types of procedures.

[44] The CLOFAN clinic bought the new software from the manufacturer of the laser machine. Besides some screen appearance changes, the new software changed the ablation, the transition zone and the calculation of the depth of ablation. We got this information directly from the manufacturer in the "Schwind Circular letter", produced by the Technical Department of Schwind eye-tech-solution GmbH, Rev. 3 / 2003-08-28 / TM/AL/SB, Germany.

Standard errors are clustered by surgeon to capture nonlinear common factors associated with a particular surgeon.

As we mentioned in an earlier section, we are able to identify clinic-wide collective learning in our data. In Table 3 we use those group adjustment rules as key independent variables, with the initial group adjustment rule as the omitted category. In the first column we control only for the second and the third group adjustment rules in addition to surgeon fixed effects. Both adjustment rules improve the outcome. The second group adjustment rule is imprecisely estimated but shows a negative sign, which means that once the initial rule was changed, outcomes improved. The third adjustment rule shows a larger point estimate and is precisely estimated. In comparison to the initial rule, the third adjustment rule that doctors implemented improved the average post surgery eyesight by 0.24 diopters.

In the second, third and fourth columns of Table 3 we add demographic controls, presurgery eyesight and type of eyesight, respectively. The effect of the second group adjustment rule is negative and statistically insignificant throughout the columns and the point estimates were sensitive to controls. On the other hand, the third group adjustment rule shows a robust point estimate and is statistically significant at 5 or 1 percent level depending on specifications. In the last column, we include a dummy for changes in the software technology used by the clinic. In this specification, the third group adjustment rule lowers the postsurgery eyesight to 0.16 diopters, which is equivalent to a substantial 15 percent improvement on the average postsurgery eyesight. This point estimate is statistically significant at 10 percent confidence level. Age is consistently significant across specifications, meaning that older patients have significantly worse outcomes. If a patient's age increases by one year, the postsurgery defocus equivalent is higher by 0.01 diopters. If the patient is a woman, the postsurgery defocus equivalent is 0.1 diopter higher, and this variable shows statistical significance in all columns except the second one.

To control for the nonlinear effect of the presurgery eyesight measure, we include 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 last column.

In Table 4 we present the estimates of individual learning and collective learning using the specifications in the last column from Table 2 and 3, respectively, where all controls are present. In the third column of Table 4 we consider individual learning and collective learning together in the same regression. As a result, the magnitude of individual learning changed from -0.05 to -0.01, which is a much smaller estimate. The group adjustment rules estimates, however, did not change much after controlling for individual learning. Although the second group adjustment rule improved the outcomes, it is not precisely estimated. The third group adjustment rule shows a larger point estimate in absolute value, which means a larger improvement compared to the initial adjustment rule, and it is statistically significant at a 10 percent confidence level.

In Table 5, we present the results for two other outcome variables. In the first column, the dependent variable is whether the postsurgery eyesight is greater than 1 diopter (i.e., a failure). When we use postsurgery eyesight as an outcome, a single outlier could have a large 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 is imprecisely estimated. The group adjustment rules presented in the second column show a monotonic improvement effect on outcomes, and the third group adjustment rule has an effect on outcomes statistically different from zero at 10 percent confidence level. The third group adjustment rule improved the outcome by 9.1 percentage point. In the third column, when we estimate individual and collective learning simultaneously, the point estimates of the effect of the third group adjustment rule on outcomes is robust. The coefficients measuring the effects on outcomes of individual learning and of the second group adjustment rule changed some and remained statistically insignificant.

In the fourth column we use repeat surgery as the adverse outcome. The overall repeat surgery rate is 8.4 percent. The coefficient of individual experience for repeat surgery is also not precisely measured. Both collective learning variables show a monotonic improvement effect compared to the initial adjustment rule, but they are not precisely estimated. When we estimate together individual and collective learning in the last column, the magnitude of the third group adjustment rule is robust throughout different specifications, even though it is insignificant.

In Table 6 we examine whether individual learning and collective learning might work differently for easy cases and severe cases. This may occur because of nonlinear effects of the presurgery eyesight on the outcomes. The upper block of Table 6 repeats the results reported in Table 4 and 5. The middle block shows the results for easy cases, and the bottom block shows the results for severe cases. We consider easy cases to be the ones below the 25 percentile in severity of presurgery eyesight, while severe cases are the ones above the 75 percentile.

In the first column, we use the postsurgery eyesight as the measure of outcome. The severe cases show larger individual learning compared to the easy cases, although they are imprecisely estimated. However, the magnitude of the coefficient for the third group adjustment rule is twice the value in the easy cases compared to the severe cases and is statistically significant at 5 percent confidence level. We show in the second column the measure of outcome which looks at a failure of surgery. The results pattern in the second column is similar to the results pattern in the first column. The third group adjustment rule for easy cases is statistically significant at 5 percent level and 20 percent improvement in terms of magnitude. In the third column, we show the resurgery measure of outcome. The results for this measure did not have any consistent pattern in previous outcomes. For resurgery, the third group adjustment rule improves some for severe cases.

6. CONCLUSIONS

In this paper, we examine the existence of individual and collective learning-by-doing for LASIK eye surgeries. LASIK surgery shares many common characteristics with other surgeries. In addition, it has some advantages over other medical procedures that allow us to measure individual and collective learning-by-doing.

In comparison with previous studies, the distinguishing features of this paper are, first, the use of a longitudinal data set with good measures of surgeons' experience and precisely defined medical outcomes, and, second, the analysis of a surgical procedure that allows us to separate individual from collective learning. Past studies have used data that often confounded measures of outcomes with unobserved underlying patient conditions, and most of those studies used only annual surgical volume as an indicator of surgeon experience. In addition, in past studies it was difficult to isolate the effect of learning-by-doing from other effects, such as selective referral, and to separate the individual from the collective learning effects on outcomes, because those studies analyzed surgical procedures performed by a surgical team rather than by individual surgeons.

The main question addressed in this paper is whether patient outcomes improve with surgeon experience. We use two measures of learning in LASIK procedures. First, we measure individual learning using the cumulative number of surgeries for each surgeon. Second, we measure group learning by using updates of the group adjustment rule, which is the result of a periodical and structured review process by the surgeons. We do not find evidence that as surgeons increase the number of surgeries performed, they obtain better outcomes. However, we do find evidence of collective learning, because outcomes significantly improved after the third group adjustment rule update. Importantly, if we did not control for measures of collective learning or for measures of technological change (reflected by updates in the software of the laser machine), we would have found spurious evidence of individual learning-by-doing.

We use three different measures of outcomes: postsurgery defocus equivalent, a dummy variable for success or failure, and repeat surgery. The first and second measure of outcome show a consistent and statistically significant 16 percent and 9.1 percent improvement after the third group adjustment rule update, respectively. Patients with easy cases benefited more as a result of the third adjustment rule update. We find that collective learning plays a substantial role in LASIK surgery performance, while individual learning does not.

Our study has some limitations. We analyzed learning-by-doing for a particular procedure – LASIK eye surgery. The relative importance of individual learning compared to collective learning may change across organizations and across procedures. This also implies that our results might not generalize to other surgical procedures. However, previous studies also examined particular medical procedures. For example, it is not clear whether the learning-by-doing that some studies have found in the case of CABG surgery applies to other procedures. Notwithstanding these qualifications, our results suggest that collective learning may be more important than individual learning for LASIK eye surgeries.

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.

Arrow K. J. (1962). The Economic Implications of Learning by Doing, *Review of Economic Studies*, Vol. xxix: 155-73.

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

Edmondson, Amy, Richard Bohmer and Gary Pisano. (2001) Disrupted Routines: Team Learning and New Technology Implementation in Hospitals, Author(s): Amy C. Administrative Science Quarterly, Vol. 46(4): 685-716

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. (draft, Federal Trade Commission, 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.

Huckman, Robert and Gary Pisano. (2006) The firm specificity of individual performance: evidence from cardiac surgery. Management Science, April

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 volume of coronary angioplasty 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.

Pisano, Gary, Richard M. J. Bohmer and Amy Edmondson. (2001) Organizational differences in the rate of learning: evidence from the adoption of minimally invasive cardiac surgery. Management Science vol 47(6).

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 Cardiothoracic 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.

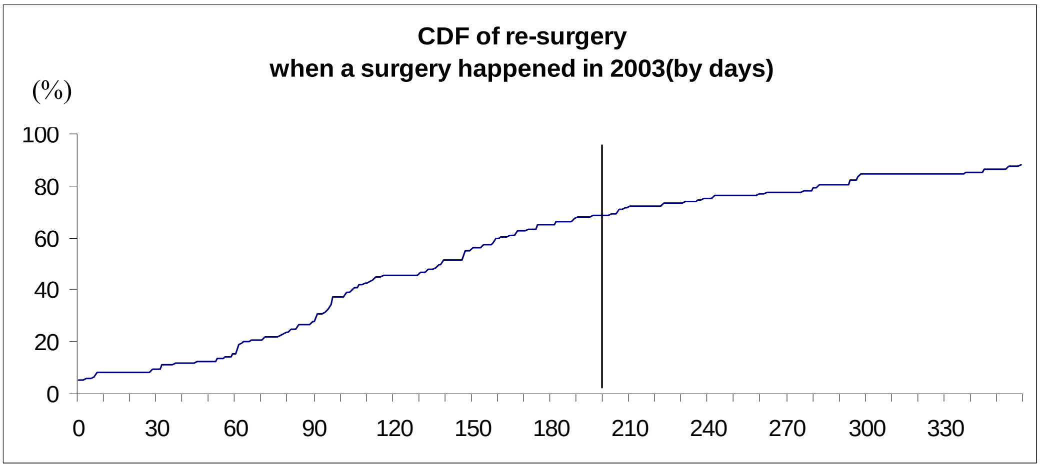
Weatherly SL. (2002). Testing Visual Acuity with the Jaeger Eye Chart. American Society of Nondestructive Testing. Available on the Internet at http://www.asnt.org/publications/.

Wright, Theodore. 1936. "Factors Affecting the Cost of Airplanes", *Journal of the Aeronautical Sciences*, 3(4): 122-128.

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



Days after surgery

Figure 3. Presurgery and Postsurgery Eyesight (by month)

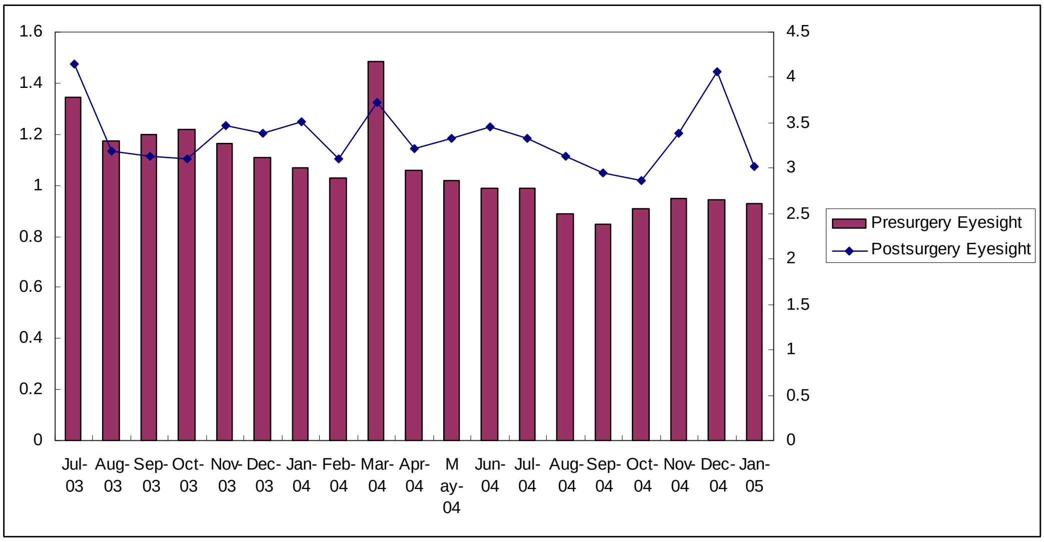


Table 1. Descriptive Statistics

	Full	By Type of Presurgery Eyesight				
		Myopia	Myopic Astigmatism	Hyperopia	Hyperopic Astigmatism	Mixed Astigmatism
Age	38.5	32.0	32.4	49.6	44.9	34.8
	(13.5)	(10.1)	(10.2)	(11.3)	(13.8)	(11.7)
Female	0.7	0.7	0.6	0.7	0.7	0.7
Presurgery Eyesight (DE)	3.32	3.05	4.28	1.96	3.19	2.38
	(2.5)	(2.18)	(2.93)	(1.45)	(2.04)	(1.43)
Postsurgery Eyesight (DE)	1.06	0.90	0.94	1.29	1.17	1.03
	(0.95)	(1.03)	(1)	(0.84)	(0.9)	(0.85)
Defocus Equivalent >1.0 D (%)	42.90	26.46	31.68	64.48	54.52	43.70
Need Resurgery (%)	8.36	9.42	7.53	10.15	8.49	5.88
	(1.9)	(1.5)	(1.6)	(2.2)	(2)	(1.6)
Average Surgeries per Surgeon	98.2					
	(95.4)					
Observations	1746	223	704	335	365	119

( ) standard deviations

Full sample includes LASIK patients operated from July 1 2003 when the new laser machine was available to Jan. 19, 2005.

We keep the 200 days observation period after the operation throughout data. We use the most up-to-date postsurgery eyesight measure within 200 days after the operation. We drop repeat surgeries from the sample even though we count it when we calculate individual cumulative number of surgeries since the objective of repeat surgery might be different.

Table 2. Impacts of Individual Learning on Postsurgery Eyesight

	(1)	(2)	(3)	(4)	(5)
Log(cumulative surgeries)	-0.101	-0.095	-0.091	-0.084	-0.050
	(0.027)	(0.029)	(0.025)	(0.025)	(0.032)
Age		0.011	0.013	0.009	0.009
		(0.002)	(0.001)	(0.001)	(0.001)
Female		0.099	0.105	0.095	0.095
		(0.057)	(0.048)	(0.045)	(0.045)
Presurgery Eyesight			-0.006	0.014	0.013
			(0.032)	(0.033)	(0.033)
Presurgery Eyesight^2			0.008	0.007	0.007
			(0.003)	(0.003)	(0.003)
Myopic Astigmatism				-0.078	-0.070
				(0.086)	(0.084)
Hyperopia				0.263	0.269
				(0.080)	(0.081)
Hyperopic Astigmatism				0.106	0.110
				(0.107)	(0.109)
Mixed Astigmatism				0.069	0.080
				(0.084)	(0.082)
Software update					-0.115
					(0.071)
Observations		1746	1746	1746	1746

Robust standard errors are clustered by surgeon and included in parentheses.
We use the most up-to-date eyesight measure within 200 days after the surgery.
A surgeon-fixed effect is included for all the columns.
Among the type of presurgery eye condition myopia is the omitted category.

Table 3. Impacts of Collective Learning on Postsurgery Eyesight

	(1)	(2)	(3)	(4)	(5)
2nd Group Adjustment Rule	-0.066	-0.014	-0.098	-0.125	-0.126
	(0.059)	(0.055)	(0.056)	(0.074)	(0.074)
3rd Group Adjustment Rule	-0.239	-0.242	-0.196	-0.196	-0.155
	(0.089)	(0.086)	(0.073)	(0.070)	(0.083)
Age		0.011	0.013	0.009	0.009
		(0.001)	(0.001)	(0.001)	(0.001)
Female		0.086	0.091	0.091	0.092
		(0.055)	(0.045)	(0.043)	(0.044)
Presurgery Eyesight			0.011	0.018	0.018
			(0.032)	(0.033)	(0.033)
Presurgery Eyesight^2			0.007	0.007	0.007
			(0.003)	(0.003)	(0.003)
Myopic Astigmatism				0.117	0.101
				(0.087)	(0.099)
Hyperopia				0.295	0.293
				(0.077)	(0.080)
Hyperopic Astigmatism				0.288	0.272
				(0.115)	(0.127)
Mixed Astigmatism				0.272	0.256
				(0.098)	(0.105)
Software update					-0.047
					(0.069)
Observations	1746	1746	1746	1746	1746

See notes for Table 2.
The first group adjustment rule which is the initial adjustment rule provided by the manufacturer was the omitted category.

Table 4. Impacts of Learning Vector on Postsurgery Eyesight

	(1)	(2)	(3)
Log(cumulative surgeries)	-0.050		-0.010
	(0.032)		(0.047)
2nd Group Adjustment Rule		-0.126	-0.111
		(0.075)*	(0.109)
3rd Group Adjustment Rule		-0.155	-0.157
		(0.083)*	(0.083)*
Age	0.009	0.009	0.009
	(0.001)**	(0.001)**	(0.001)**
Female	0.095	0.092	0.092
	(0.046)*	(0.044)*	(0.045)*
Presurgery Eyesight	0.013	0.018	0.017
	(0.033)	(0.033)	(0.033)
Presurgery Eyesight^2	0.007	0.007	0.007
	(0.003)*	(0.003)*	(0.003)*
Myopic Astigmatism	-0.070	0.101	0.089
	(0.085)	(0.100)	(0.112)
Hyperopia	0.269	0.293	0.290
	(0.081)**	(0.081)**	(0.083)**
Hyperopic Astigmatism	0.110	0.272	0.260
	(0.110)	(0.128)*	(0.155)*
Mixed Astigmatism	0.080	0.256	0.243
	(0.083)	(0.105)**	(0.120)*
Software update	-0.115	-0.047	-0.038
	(0.072)	(0.069)	(0.081)
Observations	1746	1746	1746

See notes for Table 3.

Table 5. Impacts of Learning Vector on Various Outcomes

	Failure of Surgery			Need Resurgery		
	(1)	(2)	(3)	(4)	(5)	(6)
Log(cumulative surgeries)	-0.013		0.005	-0.011		-0.007
	(0.014)		(0.023)	(0.012)		(0.012)
2nd Group Adjustment Rule		-0.037	-0.045		-0.018	-0.008
		(0.039)	(0.060)		(0.031)	(0.031)
3rd Group Adjustment Rule		-0.091	-0.091		-0.032	-0.033
		(0.048)*	(0.048)*		(0.025)	(0.025)
Age	0.006	0.006	0.006	0.002	0.002	0.002
	(0.001)**	(0.001)**	(0.001)**	(0.001)*	(0.001)**	(0.001)**
Female	0.055	0.055	0.054	0.027	0.026	0.027
	(0.031)*	(0.030)*	(0.030)*	(0.016)*	(0.016)*	(0.016)*
Presurgery Eyesight	0.044	0.046	0.046	0.017	0.018	0.018
	(0.012)**	(0.012)**	(0.012)**	(0.010)*	(0.010)*	(0.010)*
Presurgery Eyesight^2	0.000	0.000	0.000	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Myopic Astigmatism	-0.018	0.052	0.058	-0.032	-0.003	-0.011
	(0.034)	(0.040)	(0.053)	(0.032)	(0.043)	(0.047)
Hyperopia	0.267	0.276	0.277	-0.065	-0.061	-0.062
	(0.044)**	(0.042)**	(0.041)**	(0.031)*	(0.032)*	(0.032)*
Hyperopic Astigmatism	0.156	0.222	0.228	-0.064	-0.037	-0.044
	(0.051)**	(0.053)**	(0.064)**	(0.036)*	(0.042)	(0.045)
Mixed Astigmatism	0.137	0.208	0.215	-0.043	-0.013	-0.021
	(0.052)**	(0.057)**	(0.071)**	(0.033)	(0.039)	(0.043)
Software update	-0.037	0.013	0.008	-0.040	-0.029	-0.023
	(0.024)	(0.035)	(0.037)	(0.029)	(0.027)	(0.032)
Observations	1746	1746	1746	1746	1746	1746

See notes for Table 3.

Table 6. Impacts of Learning Vector on Various Outcomes, based on Severity of Cases

	Postsurgery Eyesight (1)	Failure of Surgery (2)	Need Resurgery (3)
Log(cumulative surgeries)	-0.010	0.005	-0.007
	(0.047)	(0.023)	(0.012)
2nd Group Adjustment Rule	-0.111	-0.045	-0.008
	(0.109)	(0.060)	(0.031)
3rd Group Adjustment Rule	-0.157	-0.091	-0.033
	(0.083)*	(0.048)*	(0.025)
Observations	1746	1746	1746
		Easy Cases	
Log(cumulative surgeries)	0.059	-0.008	-0.023
	(0.046)	(0.037)	(0.022)
2nd Group Adjustment Rule	0.009	-0.052	-0.028
	(0.113)	(0.084)	(0.029)
3rd Group Adjustment Rule	-0.316	-0.199	0.022
	(0.121)**	(0.070)**	(0.029)
Observations	529	529	529
		Severe Cases	
Log(cumulative surgeries)	-0.143	0.003	-0.005
	(0.148)	(0.050)	(0.035)
2nd Group Adjustment Rule	-0.023	0.062	0.018
	(0.325)	(0.186)	(0.095)
3rd Group Adjustment Rule	-0.177	-0.125	-0.088
	(0.254)	(0.137)	(0.053)*
Observations	429	429	429

See notes for Table 3.
Easy cases are defined as those below the 25th percentile of presurgery eyesight. Severe cases are defined as those above the 75th percentile of presurgery eyesight.