Title
Experiments on Health and Education in Developing Economies

Author
Lu, Fangwen

Publication Date
2011

Peer reviewed|Thesis/dissertation

Experiments on Health and Education in Developing Economies

By

Fangwen Lu

A dissertation submitted in partial satisfaction of the

requirement for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jeffrey M. Perloff, Chair
Professor Michael L. Anderson
Professor Enrico Moretti

Abstract

Experiments on Health and Education in Developing Economies

By

Fangwen Lu

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Jeffrey M. Perloff, Chair

Health and education are two important issues in developing economies. Field and natural experiments provided me with great opportunities for identifying the effects of health insurance and incentive on doctors' prescribing behaviors and the peer influences among students.

The first chapter examines whether doctors write more expensive prescriptions for insured patients and if so, why. I conducted a randomized audit experiment using undercover visits to Chinese hospitals. The results show that prescriptions for insured patients are 43% more expensive than those for uninsured patients when doctors expect to obtain a proportion of their patients' drug expenditures. The differences in prescriptions are largely explained by a differential agency problem hypothesis that doctors act of self-interest and prescribe more unnecessary or expensive drugs to insured patients, rather than by a considerate doctor hypothesis that doctors consider the trade-off between drug efficacy and patients' ability to pay.

The second chapter studies peer effects in a natural experiment generated by an unusual change in college admission policy at a prestigious Chinese university. The change in admission policy brought a large number of low-scoring students into several academic departments which only admitted high-scoring students in usual years. Exploiting the large variations in peer characteristics and the strong interactions among peer groups, the analysis finds that specially admitted low-scoring students significantly reduced the performance of regular students in standardized English tests. This detrimental effect of specially admitted students is concentrated among students with English ability below average.

Most research on peer effects in pre-college education focuses on the class or school level and assumes that students are influenced by class- or school-level averages. The third chapter examines peer effects within small groups inside classrooms by exploiting an experiment with random seat assignments inside Chinese classrooms. We define peers as either deskmates or neighboring students who sit at desks directly in front of or behind the students. The results suggest different patterns of impacts from deskmate and other neighboring students: female deskmates improve test scores for both boys and girls while the proportion of females among other neighboring students has strong positive effects on girls but no impact on boys. Overall, the results suggest that organizing students into small groups of homogeneous gender inside a classroom – a small-scale version of single-sex education – can significantly improve test scores.

Contents

List of Figures

List of Tables

Acknowledgement

I have had a wonderful study and research experience at Berkeley. I am deeply indebted to Jeff Perloff, my invaluable mentor. Thank you for the tremendous effort and time you put on me, helping me design research and write papers. Thank you for the encouragement throughout the process!

I also owe deep thanks to Michael Anderson. He makes me feel comfortable to ask all kinds of questions. His constructive comments keep me making progress all the time. The way he comforts me also works well. I won't forget he said "I have seen orals worse than yours" when he tried to ease my anxiety about the orals,.

I'd like to thank Ethan Ligon for invaluable mentoring during the early stage of my study. I have also benefitted from inspiring discussions with Enrico Moretti, Jeremy Magruder, Elizabeth Sadoulet, and Alain de Janvry.

I am very lucky to have fantastic fellow students who have made my experience enjoyable. I enjoyed those happy moments at various class gatherings! The officemates make 317 a place full of fun!

Chapter 1. Insurance Coverage and Agency Problems in Doctor Prescriptions: Evidence from a Field Experiment in China

1.1 Introduction

Health insurance coverage is expanding rapidly in China, the United States, and many other countries. Empirical evidence shows that insurance coverage is associated with rising health expenditure (see Rice and Matsuoka (2004) and Gemmill et al. (2008) for reviews). Because doctors' prescriptions are a major source of health expenditures, exploring whether and why doctors respond to patients' insurance is essential for understanding why expanding insurance coverage leads to rising expenditures. This study uses a randomized field experiment to determine whether Chinese doctors write more expensive prescriptions for insured patients than for uninsured patients. It tests two competing hypotheses explaining such an increase in drug expenditures. The *considerate doctor hypothesis* holds that doctors are concerned about the wellbeing of their patients so that they prescribe medicine taking into account drug efficacy and what they believe that each of their patients can afford to pay. The *differential agency problem hypothesis* contends that doctors act of self-interest, prescribing more expensive drugs to insured patients to increase their own profits.

Arrow (1963) identifies the principle-agent problems between patients and doctors as one of the fundamental market failures in the health care market. In many societies, doctors can pocket profits from selling drugs and services. As patients have limited knowledge about proper treatments, doctors with such financial incentives may recommend treatments to increase their own income rather than their patients' well-being. Although incentives possibly motivate doctors to prescribe more drugs than what are optimal for both insured patients and uninsured patients, insurance coverage allows doctors to increase drug expenditures by a larger amount to insured patients than to uninsured patients. In other words, insurance coverage may exacerbate the extent of agency problems and contribute to the greater health expenditures for insured patients.

Separating these two different motivations is important for evaluating the welfare implications of insurance expansion projects in many countries, but this issue has been overlooked due to identification challenges. First, insurance might be endogenous; patients with insurance may have different health needs from those without insurance, so the difference in health expenditures across insurance statuses may simply reflect different medical needs. Second, doctors' financial incentives also may be endogenous. For example, the markup from prescribing a drug is sometimes used to measure doctors' incentives for prescribing the drug, but the markup may be associated with product characteristics, or marketing strategies. Third, and most challenging, insured patients may request and then receive more drugs or more expensive drugs (e.g., Kravitz et al. 2005), which makes it difficult for studies with observational data to identify whether a doctor may initiate more expensive prescriptions to insured patients.

This study avoids these three identification challenges by using a randomized audit experiment based on undercover hospital visits in Beijing, China. Audit experiments have been used to study a wide range of social life, such as job market, car sale, car repair, sports card trading and also doctor prescribing behaviors (Bertrand and Mullainathan, 2004; Ayres and

Siegelman, 1995; Schneider, 2009; List, 2004 and 2006; Kravitz et al, 2005; Currie, Lin and Zhang, 2010). In the experiment, the same “patients” are randomly presented as having insurance or not having insurance during hospital visits. Further, doctors are randomly told either that the patient will buy drugs at the doctor’s hospital (providing doctors with a financial incentive to prescribe more drugs) or that the drugs will be purchased elsewhere (eliminating the doctor’s financial incentive). The Chinese hospitals are allowed to sell drugs with a 15% markup over the wholesale price and keep the profits. Doctors receive a share of these profits from prescribed drugs sold in their own hospitals; however, doctors do not have a financial incentive for prescribing more drugs if patients buy the drugs elsewhere.

An institutional feature in Beijing helps to circumvent the third identification challenge. In Beijing, doctors in top-rated hospitals are frequently consulted by family members for patients who live elsewhere, and they write prescriptions without seeing patients in person if their family members provide adequate medical information. In the experiment, a “family member” describes the patient problems and provides medical test results using a standard script. Thus, the problems due to differences between patients in observational studies are eliminated.

The analysis shows that when doctors are presented with incentives, insured patients receive prescriptions that cost 43 percent more on average than those of uninsured patients. Perhaps more telling, doctors are more likely to prescribe unneeded drugs to the insured (64%) than to the uninsured (40%). In contrast, doctors without a personal financial incentive are not more likely to prescribe unneeded drugs to the insured, and drug expenditures are similar for both insured and uninsured patients, which suggests that insurance status by itself is not responsible for this behavior.

For an insured patient, doctors with incentives write much more expensive prescriptions than do doctors without incentives; this confirms that agency problems play an important role in drug prescriptions. However, the agency problems are largely constrained when a patient has no insurance. Overall, this study shows that the interaction between insurance coverage and agency problems has significant impacts on medical expenditures. Accordingly, it is doctors’ self financial interest rather than their concerns for patients’ wellbeing that motivates more expensive prescriptions to insured patients.

This study makes three major contributions to the literature on health insurance and agency problems. First, this study provides the clearest picture to date about the causal effects of a patient’s insurance coverage on the decision-making of doctors, especially when doctors’ own earnings are under stake. Most empirical work on the effects of health insurance either has focused the effects via patients or has studied the combined effects from both patients and doctors (Card, Dobkin and Maestas, 2008; Anderson, Dobkin and Gross, 2009; Wagstaff et al., 2009; Zweifel and Manning, 2000; Rice and Matsuoka, 2004; Gemmill et al., 2008; Lundin, 2000). Mort et al (1996) and McKinlay, Potter and Feldman (1996) investigate how insurance affects doctor decisions, but in their studies, doctors were aware that their decisions were under study. This paper uses controlled hospital visits and randomized insurance status to demonstrate that doctors react strongly to patient’s insurance coverage.

Second, this analysis adds new empirical evidence on prescriptions to the agency health literature. McGruire (2000) reviews empirical evidence on agency problems in the health care market, and most studies focus on health care services. As the prescription of drug is more transparent than the prescription of service and therefore is less subject to information asymmetry, it is important to know whether and how agency problems exist in drug prescribing behaviors. In addition to Iizuka (2007) and Dalen, Sorisio and Strom (2010), this study deepens the understanding of drug prescribing behaviors, and shows doctors respond to the randomly presented financial incentives by prescribing more drugs.

Third, this work is the first study to separate the two explanations for increasing drug expenditures under insurance coverage – doctors' self interest versus their consideration for patients, and it demonstrates that insurance coverage can exacerbate doctors' agency problems. Kessel (1958), Feldstein (1970) and Sun et al. (2009) interpret the larger expenditures associated with insurance coverage by referring to the interaction between insurance and agency problems, but they can not provide solid evidence. Iizuka (2007) is an exception that looks at a patient's insurance and a doctor's markup at the same time. This paper uses controlled hospital visits and written prescriptions (as filled prescriptions are subject to patients' decisions) to exclude explanations due to differences across patients. And the randomized insurance and incentive status provide a rigorous identification for illustrating the interaction effects between insurance and incentive on doctors.

The next section describes insurance and doctors' incentives in China. Section 1.3 presents the experimental design and predictions tested. Section 1.4 describes the data, and Section 1.5 presents the main results. Section 1.6 conducts robustness checks. Section 1.7 summarizes the paper and draws conclusions.

1.2 Institutional Background

The Chinese government can not fully subsidize public hospitals, so it allows hospitals to profit from selling drugs, which is called the policy of "compensating hospitals through drug sales". In Beijing, the wholesale prices for drugs are set by the government. Public hospitals (except community-level hospitals) are allowed to charge a retail price for a drug that is 15% above the government-set wholesale price.[1] Every hospital has its own pharmacy. Around 70% of drugs across the country are dispensed through hospital pharmacies.[2] Revenues from drugs account for about half the total revenues for hospitals.[3]

The majority of doctors are affiliated with public hospitals. Although private clinics have emerged in recent years, public hospitals are the dominant medical service provider in China. Doctors work in the outpatient service or the inpatient service in hospitals, and they get salary, bonus, and other benefits from hospitals. How hospitals share profits with doctors may vary across hospitals, but it is widely believed that doctors do share profits.[4] Several online cases suggest that doctors usually have to fulfill a targeted workload first and then get an additional bonus in proportion to work exceeding the target. Only drugs or services sold at their own hospitals can be linked to doctors' bonuses. If patients choose to buy the prescribed drugs elsewhere, doctors cannot share the profit from selling drugs at places other than their hospitals. This feature is important for the design of the experiment. It is possible that drug companies

[1] During the experiment period, the wholesale prices at military-affiliated hospitals and other hospitals were decided by different government agencies, so the wholesale prices are not uniform, but those prices are very close. No violations of the rules for drug pricing appeared during the experiment.

[2] Source: http://finance.ifeng.com/news/special/xylgg/20100419/2070067.shtml (in Chinese)

[3] Source: http://health.sohu.com/20090318/n262859031.shtml (in Chinese)

[4] The share of profits between a hospital and its doctors is said to be a secret rule in hospitals. Source: http://health.newssc.org/system/2009/02/05/011540387.shtml (in Chinese)

would bribe doctors for prescribing their drugs, but this possibility is ignored in this study given the lack of information.[5]

The top-rated hospitals are called Level-3A hospitals, which tend to be concentrated in large cities, such as Beijing. Patients from other areas may visit a doctor at a top-rated hospital in Beijing for treatments or for second opinions. When patients have difficulty in getting to hospitals, and if medical problems can be adequately described by test results, doctors are usually willing to give medical suggestions without seeing the patient in person. This practice is common in top-rated hospitals.[6] Doctors have a fixed phrase for this practice – "patient is not present, family member consults on patient's behalf". Under this circumstance, it is also reasonable for the patient to either buy prescribed drugs near his home or at the Beijing hospital.

The particular type of insurance chosen for this study is the commonly called "public-expenditure" insurance. The favorable feature of the public-expenditure insurance is that hospitals do not verify the insurance status if a patient states that she has public-expenditure insurance. With public-expenditure insurance, patients pay all the medical fees to hospitals as if they did not have any insurance, and then file for reimbursement with the necessary receipts at their affiliated organizations. This feature facilitates the manipulation of insurance status.

Employees of government agencies and government-affiliated non-profit organizations tend to have public-expenditure insurance. Copayment rates vary as they are decided independently by each employer. As patients with public-expenditure insurance pay full medical fees to hospitals, doctors facing public-expenditure insurance do not know the exact copayment rate. The typical copayment rate for employees in urban areas is around 30%, and it is commonly believed that government agencies and government-affiliated organizations tend to offer better benefits, so the copayment rate for public-expenditure insurance is usually expected to be below 30%.

Different insurances in Beijing vary largely in terms of deductibles, copayment rates, and drugs covered by insurance. For the purpose of this study – to compare patients with insurance and patients without insurance, the public-expenditure insurance shares three important features with other insurances: first, insured patients pay a lower out-of-pocket price for drugs on average than uninsured patients pay; second, for many drugs, the copayment rates are the same for the brand-name drugs and their generic equivalent; and third, the regulations on prescriptions are largely similar.

Hospitals are required to clearly label prices for drugs and health care services, and they are prohibited from charging different prices according to insurance status. Hospitals have discretion over which drugs – brand-name or generic – to carry in their pharmacies, as long as they do not price drugs above the allowed level.

1.3 Experiment Design and Predictions

The experiment used undercover hospital visits to detect variations in how doctors prescribe. The experiment employed a 2-by-2 intervention design:

[5] Occasionally, drug companies are exposed secretly bribing doctors for prescribing their drugs. For example, in May 2010, a drug company was exposed for paying 20% of revenue (6.5 out of 32.2 yuan per unit) to doctors who prescribed the drug. Source: http://www.chinanews.com.cn/jk/jk-aqjs/news/2010/05-30/2312025.shtml

[6] In a list of suggestions on how patients should communicate with doctors, written by a doctor in a Level 3A hospital, the first suggestion is to tell doctors clearly who is the patient as people sometimes see doctors on behalf of others. Source: http://www.yafda.gov.cn/article_view.asp?id=2515&type=6&parentID=6

(1) The patient was presented to have public-expenditure insurance ("insured") or not to have any insurance ("uninsured");

(2) Before the prescription was written, the doctor was informed either of the patient's intention to buy medicine from the doctor's hospital, so that the doctor has an "incentive" to prescribe extensively, or of the patient's intention to buy elsewhere ("no incentive").

	Incentive	No incentive
Insured	A	C
Uninsured	B	D

Thus, there were four possible situations, as the chart shows: A. insured-incentive; B. uninsured-incentive; C. insured-no-incentive; D. uninsured-no-incentive. Insurance and incentive status was randomly assigned to doctors.

Hypothetical Patients

Two hypothetical patients were constructed for the experiment (see Appendix 1), and the same hypothetical patients were used for hospital visits across all the interventions. Diabetes, hypertension and abnormal triglyceride were chosen because these problems are among the most common illnesses for the elderly and they are relatively easy to describe with numbers.

Patient 1 was described as a 66 year-old male who recently received medical test results showing high blood pressure, high blood sugar, and elevated triglyceride, and was not yet taking any medication. The problem with triglycerides was included to test for over-prescribing. If a patient were to look at the lab results for blood lipids, he might reasonably conclude that he needed medication for triglycerides. However, according to medical guidelines, the patient should not be prescribed drugs for triglyceride, given his level of triglyceride and the possible side effects of the drugs.[7] Therefore, the prescription of triglyceride drugs indicated over-treatment or inappropriate treatment.

Patient 2 was described as a 65-year-old male with hypertension, already taking the brand-name drug Nifedipine, a control-released tablet for hypertension, but with his blood pressure still abnormal.

The experimenters posed them as family members who visited doctors. They were prepared with answers for the most likely questions that doctors might ask concerning other health problems, such as histories of family illness, smoking, drinking, height, weight, etc. The prepared answers excluded all other risk factors. The experimenters presented the patients' test results directly to the doctors, but they would only give other information if asked by doctors, because knowing the relevance of other information might suggest that the experimenters or the patients were much more knowledgeable than is typical.

Patients were described as living far from where the hospitals were located. If a patient lived nearby, a doctor might suggest that the patient visit the hospital himself. More importantly, in the no-incentive interventions, doctors might get angry if the patient refused to buy drugs at their hospitals, but they would understand if the patient lived far away.

This study was designed to evaluate the effects of health insurance for the social average patient (rather than, say, for a poor patient). Two indicators were provided to doctors to show that the patient was not poor. First, the drug that was currently being taken by Patient 2 for hypertension is a relatively expensive brand-name drug. The brand-name Nifedipine tablet costs

[7] For the guideline in the U.S., see http://www.nhlbi.nih.gov/guidelines/cholesterol/atglance.htm.

about 163 yuan (around 24 dollars) per month,[8] which is 67% higher than the 98 yuan for the generic equivalent.[9] Among drugs for hypertension, the brand-name Nifedipine is one of the most costly. In addition, the monthly expenditure for the brand-name Nifedipine costs more than 10% of the average income of an urban resident.[10] Given the availability of a much cheaper generic equivalent, a patient who pays for the brand-name Nifedipine tablet all out of his own pocket is likely to have decent economic status. Second, if doctors asked about the economic status of an uninsured patient, the doctor was told that the patient had "middle income." In China, "middle income" means that the patient is neither very poor nor very rich.

Experimenters and Hospital Visit Procedure

Hospital visits were conducted by two trained experimenters: the author, a 32-year-old Chinese female, and a 56-year-old Chinese female assistant. Before visiting hospitals, the experimenters participated in mock doctor visits and compared experiences in order to ensure consistent presentations.

Upon arriving at a hospital, the experimenter first went to the registration window, paid the visit fees, and provided information on the patient's name, gender, age and other demographics. Whether the patient had public-expenditure insurance was declared at the registration window. The registration staff either used a computer system to send the demographic and insurance information to the doctor, or printed that information on a registration ticket for the patient or family member to present to the doctor.

When an experimenter saw a doctor, she introduced herself by saying, "I am coming to see you on behalf of my {the relative} who lives in my hometown. He wants a doctor in a top-rated hospital to look at his case." Then the experimenter described the health problems according to the standard script. The experimenter always brought a reference sheet – a piece of paper with medical test results indicating health problems. The experimenter was required to describe the health problems with the assistance of the reference sheet, but was not allowed to directly read the reference sheet. Then the experimenter said either "{the relative} asked me to buy the medicines here for him" for the incentive interventions, or "{the relative} wants to get a prescription and buy drugs at his local store" for the no-incentive interventions. In the end, the experimenter exited the doctor's office with printed or hand-written prescriptions.

Samples of Hospitals and Doctors

All the Level-3A hospitals with separate departments for endocrinology and cardiology in the urban districts of Beijing were included as the sample.[11] Patient 1 was presented to the department of endocrinology, and Patient 2 was presented to the department of cardiology. Each hospital-department received four visits with one visit under each intervention.

The interventions were randomly assigned to doctors within hospitals. All the doctors visited were specialists in the sense that they focus on either endocrinology or cardiology. In Chinese hospitals, there are two types of visits – regular visits and expert visits. Compared to regular visits, expert visits are associated with higher visit fees and doctors with higher titles. As

[8] The exchange rate was \$1 U.S. ≈ 6.8 CNY during the experimental period, the summer of 2010.

[9] A study compares brand-name nifedipine with the generic one, and finds that the two are statistically similar in terms of clinical efficacy and adverse reactions.

Source: http://www.100paper.com/100paper/yiyaoxue/yaoxue/20070623/24750.html

[10] The average annual income per capita among urban residents is 15781 yuan in 2008, so the current drug expenditure of the brand nifedipine tablet is 12% of the average income (163*12/15781≈12%).

[11] This is to exclude Level-3A specialized hospitals which only have departments for particular health problems.

hospitals do not post doctors' names and work shifts for regular visits, and as schedules for expert visits are sometimes inaccurate, it was not possible to randomize by assigning interventions to an ex ante list of doctors, as is commonly done in other field experiments. However, doctors are assigned by either the computer system or staff in regular visits, so a patient has very limited ability to choose a doctor during regular visits. Given that doctors are exogenously decided, a random sequence of the four interventions to a hospital generates a random assignment of interventions to doctors. In the experiment, the sequences for Patient 1 were generated first, and the sequences for Patient 2 were matched to those for Patient 1 by switching the insurance status.

In a few hospitals, there are only one or two doctors available for regular visits. To fulfill the four visits per hospital-patient, the experimenters supplemented with expert visits when necessary. Technically a patient can choose which doctor to visit for expert visits, but the experimenters did not make such a selection. The experimenters always took the next available expert visit if an expert visit was needed. When more than one expert visit was available on the same day, one of them was chosen randomly.

It was difficult to prevent two experimenters from visiting the same doctor during regular visits, because experimenters usually do not know whether a particular doctor has been visited by another experimenter. To limit the possible memory effect, the experimenters left at least a one-week gap between any two regular visits to the same hospital. The doctors consulted while the experiment was under design said that doctors are more likely to remember a patient's face than a patient's health problems. In the experiment, experimenters were asked to avoid visiting the same doctor twice. If there was a high chance that an experimenter would see a doctor she had visited before, the visit was scheduled to another date.

Predictions tested

Following the literature that incorporates a disutility of acting against the best interest of the patient (Evans, 1974; McGuire and Pauly, 1991; Gruber and Owings, 1996), a doctor's utility can be specified as some combination of her own income, her consideration for a patient's financial situation as well as her professional concerns capturing the disutility for deviating from the optimal prescription. In both incentive and no-incentive interventions, the doctor may consider the tradeoff between drug efficacy and patients' other consumptions for patients, and prescribe more drugs or more expensive drugs to insured patients who pay lower out-of-pocket prices. In the incentive interventions, as the doctor's own income is in proportion to the expenditure of prescribed drugs, she wants to increase the drug expenditure up to a patient's ability to pay.[12] Let A, B, C, D in the previous chart be the drug expenditure under the corresponding intervention. The 2-by-2 experiment design allows testing five predictions.

(1) Insurance effect under incentive: A > B. A doctor with incentives writes more expensive prescriptions to insured patients than to uninsured patient for two possible reasons – her consideration on the tradeoff between patients' health and other consumptions or her intention to increase the expenditure up to a patient's budget limit.

(2) Insurance effect under no incentive: C > D. A doctor without incentives prescribes more expensive drugs to insured patients due to her consideration on the tradeoff. This is a test for the considerate doctor hypothesis.

[12] Some insured patients may have very low copayment rate. In that case, a doctor's professional concern may limit the prescription.

(3) Agency problems with insurance: A > C. If a patient has insurance, a doctor with incentives prescribes more expensive drugs than does a doctor without incentives due to the motivation for profit.
(4) Agency problems with no insurance: B ≥ D or B < D. If a patient has no insurance, whether the incentive increases drug expenditures depends on the magnitude that the patient's budget is expected to be and the prescription expenditure when the doctor has no financial incentive.
(5) Interaction effect: A-B > C-D. Insurance coverage exacerbates the effect of incentive due to a doctor's intention for increasing the expenditure up to a patient's budget limit. Alternatively, it is due to the different degree of agency problems between predictions (3) and (4). This is a test for the agency hypothesis, or more completely a differential agency problem hypothesis.

1.4 Data

I visited all Level-3A hospitals with separate departments for endocrinology and cardiology in the urban districts of Beijing between June and August of 2010. Five hospitals were eventually excluded from the sample because they either do not separate the "public-expenditure" insurance from no insurance or do not provide prescriptions unless patients pay for drugs there.[13] Although most doctors were willing to give prescriptions in the absence of the actual patient, some doctors were not. I excluded two hospitals whose doctors in the endocrinology department refused to prescribe without seeing the patient.[14] For Patient 2, one more hospital was dropped due to refusals from doctors in cardiology to prescribe without seeing the patient. Therefore, the final sample includes 25 hospitals for Patient 1 and 24 hospitals for Patient 2. Among the visits to the hospital-departments included in the sample, the experimenters encountered five additional refusals for Patient 1 and four refusals for Patient 2. The successful visits represented 95.4% of the intended visits, without counting refusals in hospitals which were dropped from the analysis; otherwise the success rate was 88% for Patient 1 and higher for Patient 2.[15] The success rate is comparable to the 91% consent rate in the videotape study by McKinlay et al. (1996), and is much better than the 64% consent rate in the survey by Mort et al. (1996) and 53-61% in the study with standardized patients by Kravitz et al. (2005). There was no correlation between refusals and intervention types. (See Appendix 2 for the list of refusals.) When a refusal occurred, a second visit was attempted; in no case was a third try needed.

Visit characteristics include variables on who conducted the visit, whether the visit was an expert visit, and the doctor's gender and age. The doctor's age was estimated based on the experimenter's best guess. If a doctor seemed to be closer to 40 years old rather than 35 or 45

[13] Two hospitals label patients with "public expenditure" insurance if a patient does not have any insurance, and I believe doctors there tend to think patients actually have insurance. In one visit to those hospitals, when I was leaving, I forgot to take the registration ticket, and the doctor said "don't forget it; you need this for reimbursement" although I presented a patient without insurance.

[14] As patient 1 was expected to convey more interesting results, a hospital was excluded from the sample if the doctors in the endocrinology department refused to give prescriptions during the first two visits, although doctors in the cardiology department in the same hospital were willing to prescribe. If two refusals occurred, no additional visits were attempted.

[15] Only two visits were conducted for hospitals which were dropped from the analysis. To calculate the success rate, I assume four refusals in each dropped hospital, which gives a success rate of 88% for Patient 1 (4*2 hospitals + 5 additional refusals divided by 4*27 hospitals). For Patient 2, the first visit was successful in the hospitals which were dropped due to refusals to Patient 1.

years old, her age is recorded as 40. Descriptive statistics are presented in Table 1.1. As there is a random matching between intervention types and doctor visits, Table 1.2 shows that the visit characteristics across intervention groups are largely balanced with data combining the two patients, and none of the F tests is statistically significant. The visit characteristics are also balanced for each individual patient (results not presented).

The raw drug expenditure is the overall amount of payment associated with a prescription. Experimenters requested a prescription of drugs for a month. Depending on the number of drugs per package and doctor's preference, drugs are usually prescribed for 28 days, 30 days or 35 days.[16] In the no-incentive interventions, the prescription is only to give patients information on what drugs to use and how to use them, and it is difficult to know whether doctors intend to prescribe for 28 days or 35 days. Therefore, the raw total expenditure is not calculated for the no-incentive interventions.

Drugs prescribed for Patient 1 can be separated into three mutually exclusive categories: drugs for triglycerides, drugs for diabetes and hypertension, and supplementary drugs. Drugs for Patient 2 include drugs for hypertension and supplementary drugs. Supplementary drugs are aspirin in most cases, except that vitamin B1 was included in one visit. The brand-name aspirin costs about 15 yuan (3% of average raw drug expenditure) and has a small effect on drug expenditure. There is not a clear medical justification whether to prescribe aspirin for the two hypothetical cases.[17] So I do not consider aspirin in the analysis.

Patient 1 has an abnormal triglyceride level, but the level is well below the level requiring drug therapy according to the medical guidelines. Therefore, whether doctors prescribe drugs for triglycerides can serve as an indicator for over-prescription. On average, 43% of doctors in this study prescribed drugs for triglycerides, and there seems to be evidence for over-treatment.

Both diabetes and hypertension require long-term drug therapy, and prescribing drugs for a longer period does not indicate a waste. Therefore, I calculate monthly drug expenditure for diabetes and hypertension on a 30-day basis. The number of drugs and the unit of drugs capture the intensity of prescribing, although admittedly neither is a perfect measure.[18] The number of drugs is a simple count of how many drug names are prescribed.[19] To construct the unit of drugs, a table of "unit dosage" for each drug is compiled, based on the most commonly prescribed quantity or the representative dosage of the brand-name drug (see Appendix 3). The differences between the number and the unit of drugs can be illustrated by an example. Say, a typical usage of Matformin is 500mg each time and 3 times a day (500mg*3), and a doctor prescribes 250mg*3, then Matformin is counted as 1 in terms of the number of drugs, but 0.5 in terms of the unit of drugs. The share of brand-name drugs is the number of brand-name drugs divided by the number of all drugs. A drug is classified as a brand-name drug if (i) it is clearly labeled as brand-

---

[16] In several visits for Patient 2, because the health problem was very simple, doctors reacted quickly and printed out the prescription with drugs for half a month or without drugs currently takenen before a research assistant had a chance to make the request. In these cases, I added necessary drugs and calculated the total drug expenditure with the modified parts based on 28 days or 30 days (not beyond 30 days).

[17] Aspirin is recommended for male hypertension patients above 50 years old if they do not have relevant contraindications for using aspirin, a condition which both Patient 1 and 2 satisfy, and if they are able to reduce their blood pressure below 150. Doctors who prescribe may expect the blood pressure to fall below 150, while those who do not prescribe may wait to prescribe until the blood pressure actually fall below 150. One doctor explicitly said "this time I don't prescribe aspirin, and he needs it when his blood pressure gets back to the normal level."

[18] The variables on drug intensity require a certain aggregation of effects of various drugs, but the precise measure of effects of drugs on blood sugar and blood pressure is complicated and not available.

[19] For Patient 2, two doctors suggested double the amount of the brand nifedipine tablet; it is counted as two drugs under this circumstance.

name drug in the prescription; or (ii) its price is much closer to the price of the corresponding brand-name drugs than to the generic equivalent; or (iii) it is a Chinese patent medicine.[20]

In the no-incentive interventions, several doctors wrote down multiple treatment plans and let patients choose. Under these circumstances, doctors usually suggested that the first one is the best but all the plans work well. So the first plan is used to construct data.

1.5 Results

This section tests the five predictions listed in section 3. Table 1.3 presents the averages of various outcome variables under each intervention. Results in Table 1.4 to 8 are from the linear regression model for the purpose of simple interpretations, but all the results survive the tests of alternative model specifications, which include taking logarithms for expenditure related variables, using a logit model for the binary variable on whether drugs are prescribed for triglycerides, applying a Poisson model for the number of drugs, and taking a Tobit model for the share of brand drugs whose values range between 0 and 1.[21]

1.5.1 Effects of health insurance when doctors have financial incentives

The specification in Equation (1.1) is used to test Prediction 1.

$$Y_{hi} = \alpha_0 + \alpha_1 Insurance_{hi} + X_{hi} + hospital_{hi} + e_{hi} \qquad (1.1)$$

The variable Y_{hi} indicates an outcome variable for hospital h and visit i. The key predictor $Insurance_{hi}$ equals 1 if a patient has insurance and 0 otherwise. The control variables X_{hi} include the four variables for visit characteristics. The analysis is restricted to observations under the two incentive interventions. Table 1.4 presents the estimated coefficients for the key predictor with each coefficient corresponding to a separate regression. The outcome variables are listed on the left. The first two columns present the results combining Patient 1 and Patient 2 together except for the variable on a prescription for triglycerides. All the standard errors are clustered at the hospital level. The first column controls for hospital fixed effects only, [22] and the second column controls for both hospital fixed effects and visit characteristics. The sample size is 50 for prescription on triglycerides and 98 for all other outcome variables. The results for Patient 1 and Patient 2 separately are in column 3 to 6.

Referring to Table 1.3, patients pay 522 yuan in total if they have insurance and 365 yuan otherwise. The 157-yuan difference represents 43% of the amount an uninsured patient pays. Alternatively, if drug expenditures are compared within each hospital, Patient 1 pays for a higher expenditure in 19 out of 25 hospitals if she is insured; the corresponding number for Patient 2 is 20 out of 24.

Drugs for triglycerides should not be recommended for Patient 1. Doctors who did not prescribe tended to say "the level is not very high" or "when the level of blood sugar is reduced, the blood lipid will go down automatically." Overall, 64 versus 40 percent of doctors prescribe drugs for triglyceride under each insurance status respectively, and the 24-percentage difference is weakly significant (t=1.95, p=0.064 in the full model).

[20] The monthly expenditures for Chinese patent drugs are comparable to or higher than those of brand-name western drugs except for one Chinese patent drug.

[21] If the model is featured by the maximum likelihood estimation, the hospital fixed effects are dropped out of the specification. For the number of drugs, I also try the negative binomial model, but the concavity assumption is not satisfied for several specifications.

[22] Since the intervention is fully balanced within hospitals, the model with only hospital fixed effects gives the same results as the model without fixed effects when the standard errors are clustered at the hospital level.

For the health problems which require drug therapy – hypertension and diabetes for Patient 1 and hypertension for Patient 2 – the insured pay 126 yuan more in terms of the monthly drug expenditure, about 42% more than the amount that the uninsured pay. Higher monthly drug expenditures could be driven by two factors: (1) doctors use more intense drug therapy, and (2) doctors prescribe more expensive drugs rather than cheaper drugs. Table 1.4 provides evidence for both channels. On average, doctors prescribed 0.26 more kinds of drugs or 0.44 more units of drugs to the insured.

There is a wide range of drugs suitable for the two patient cases. To manipulate the price level, doctors can choose different brand-name drugs or they can choose between brand-name drugs and their generic equivalence. But it is difficult to capture the intention for manipulating prices by choosing different brand-name drugs since those drugs also vary largely in other dimensions. This study explores the share of brand-name drugs. Doctors did not prescribe 100% of brand-name drugs even for the insured, which may be explained by doctor's habits or doctors' concern for the out-of-pocket payment, as drugs are not totally free to the insured. Table 1.4 shows that doctors prescribed 15 percent more brand-name drugs to the insured. One issue to note is that generic drugs are always much cheaper than the brand-name equivalent, but some generic drugs may have similar price levels as brand-name drugs of other types. Therefore the share of brand-name drugs is only a partial indicator for price manipulation.

The comparison between the insured-incentive intervention and the uninsured-incentive intervention confirms Prediction 1 – a patient's insurance status has strong effects on doctors' prescription decisions when a doctor is incentivized. As postulated earlier, doctors' concern for patient welfare can motivate differential prescriptions across insurance statuses; the following subsection explores this effect.

1.5.2 Effects of health insurance when doctors have no incentives

When doctors know that patients will not buy drugs from their hospitals, they have no financial incentive to prescribe more than what they think is optimal for the patient's health. Prediction 2 suggests that doctors might prescribe less expensive drugs to the uninsured if they care about the patient's out-of-pocket expenditure. I use a similar empirical strategy as in Equation (1.1) but restrict the analysis to the sample under the no-incentive interventions.

During the visits, several doctors showed explicit concerns about the insurance status, although patients will not buy drugs at their hospitals. They said the patient did not have health insurance, so they prescribed inexpensive but effective drugs for the patient, which suggests those doctors were empathetic toward uninsured patients. However, the comparison between means does not provide strong support for the empathy effect. Table 1.5 presents the results for the comparison between the insured-no-incentive intervention and the uninsured-no-incentive intervention. When doctors do not expect profits from prescriptions, none of the outcomes is statistically different across insurance statuses. Doctors are not more likely to prescribe drugs for triglycerides, which is not surprising given that those drugs are not needed. The 17-yuan difference represents 5.5% of the drug expenditure of the uninsured, but the differences are far from being statistically significant.

That doctors prescribe drugs for triglycerides even without a financial incentive may reflect the general over-medication problem, which is common in many Asian countries, such as India (Das-Hummer, 2007). As experimenters explicitly described triglycerides as a concern, doctors might have tried to avoid frustrating the experimenters by suggesting drugs for triglycerides. In addition, habit and competence also may explain the results. However, the overall tendency for over-treatment does not threaten the validity of the analysis, as the conclusions focus on the comparison across interventions so that the common over-treatment tendency cancels out.

When patients do not buy drugs at doctors' hospitals, doctors might leave some decisions for patients to make later, such as the brands of drugs, which may explain why there are no significant differences in drug expenditures and shares of brand-name drugs. However, the same explanation cannot account for the null results on drugs for triglyceride and the intensity of drugs.

Admittedly, due to the small number of observations, the standard errors for estimated insurance effect are large. A simple calculation shows that the hypothesis can be rejected only if doctors without incentive respond to insurance status by increasing prescriptions by 14% difference or more.[23] The 5.5% difference in drug expenditure is comparable to the 7.5% difference in recommendations for medical tests found in Mort et al. (1996). Nevertheless, the difference across insurance statuses when doctors have no incentive is too small to account for the difference under incentives. The interdependent preference among doctors cannot explain the insurance effect under incentives.

1.5.3 Effects of incentives if a patient has health insurance

Equation (1.2) is used to test the effects of incentive where $Incentive_{hi}$ equals 1 if in the incentive intervention and 0 otherwise. This subsection focuses on insured patients.

$$Y_{hi} = \alpha_0 + \alpha_1 Incentive_{hi} + X_{hi} + hospital_{hi} + e_{hi} \tag{1.2}$$

Results in Table 1.6 confirm Prediction 3. Doctors with incentives are much more likely to prescribe drugs for triglycerides than are doctors without incentives if the patient is insured. Incentivized doctors write 100-yuan (31%) more expensive prescriptions for diabetes and hypertension. Evidence also shows that doctors with incentives use drugs more intensively for both patients. However, the share of brand-name drugs is 83% under incentive and 81% under no incentive, and incentivized doctors are roughly equally likely to prescribe brand-name drugs as are doctors without incentives.

1.5.4 Effects of incentives if a patient has no insurance

Table 1.7 presents results on the effects of incentives when a patient does not have insurance coverage. According to Prediction 4, the net effect of incentive on an uninsured patient is not clear, which depends on the expected budget constraint of a patient. When Patient 1 has no insurance, doctors with incentives are not more likely to prescribe drugs for triglycerides. The negative difference is unexpected; as it is not statistically significant, it is treated as random errors. The drug expenditures and intensities of drugs are more or less similar. The differences in the share of brand-name drugs show that doctors with incentive prescribe less brand-name drugs, which have two alternative explanations. First, it may suggest that patients' budgets are expected to be smaller than the amount required for the treatments. Second, in the no incentive case, the names of brands, as comparing to the scientific names of drugs or generic drug names, might be more convenient for doctors to communicate with patients, and the prescriptions of brand-name drugs does not necessarily indicates that doctors want to prescribe brand-name drugs.

---

[23] The minimum detectable effect $MDE = (t_{1-\kappa} + t_{\alpha}) * \sigma / \sqrt{p*(1-p)*N}$. The term σ is the standard deviation in drug expenditure if everyone is uninsured and doctors are under no incentive; N is the sample size; p is percent of treatment; $t_{1-\kappa}$ is the t statistics on power and t_{α} is on significance level. Even if I ignore the requirement on power, MDE = 43, which is around 14% of the average drug expenditure in the uninsured-no-incentive intervention.

1.5.5 Interaction effect of insurance and incentive

The interaction effect of insurance and incentive can be calculated by subtracting coefficients in Table 1.5 from those in Table 1.4 or subtracting coefficients in Table 1.7 from those in Table 1.6. Alternatively, the interaction effect can be expressed as β_1 in Equation (1.3).

$$Y_{hi} = \beta_0 + \beta_1 Insurance_{hi} * Incentive_{hi} + \beta_2 Insurance_{hi} + \beta_3 Incentive_{hi} + X_{hi} + Hospital_{hi} + e_{hi} \qquad (1.3)$$

Table 1.8 presents the estimated interaction effects. Prediction 5 suggests that the interaction between insurance and incentive could have strong effects on drug expenditure if incentivized doctors take advantage of the enlarged ability to pay under insurance coverage. The results on all outcome variables support Prediction 5. Doctors are much more likely to prescribe drugs for triglycerides when insurance and incentive are both present. The interaction effect accounts for a 105-yuan difference in the monthly drug expenditure for diabetes and hypertension, and it is about 80% of the expenditure difference across insurance status when doctors have incentives. Referring to average outcomes by intervention types in Table 1.3, the interaction effects on drugs for triglycerides, drug expenditure and drug intensity for diabetes and hypertension are driven by the fact that prescriptions in the insurance-incentive intervention deviate from those in the other three interventions. However, the interaction effect on drug brand is caused by doctors who prescribe a lower percentage of brand drugs in the uninsured-incentive intervention.

1.5.6 Effects of other variables

Table 1.9 presents estimated coefficients of all variables from Equation (1.3). Several facts are worth noting. First, experimenters do not have an effect on the outcomes, which suggests that the two experimenters did a successful job in conforming doctor visits. Second, none of the control variables has a significant effect on the prescription of drugs for triglycerides, and this indicates over-treatment is not correlated to any particular characteristics of doctors. Third, doctors holding expert visits tend to write more expensive prescriptions, while older doctors tend to prescribe fewer brand-name drugs.

1.6 Robustness

1.6.1 Doctors with multiple visits

During the doctor visits, experimenters were required to avoid visiting the same doctor twice, but the same doctors could be visited by different experimenters more than once.[24] Two concerns might arise: first, although there was at least one-week gap between any two visits to the same doctor, doctors might remember earlier visits and become suspicious during later visits; second, doctors visited more than once might have different characteristics from others and drive the results in particular. To address these concerns, the sample is separated into doctors visited only once and doctors with multiple visits. Table 1.10 shows the results according to Equation (1.4) with the sub-sample of doctors who are only visited once. Only the coefficients on the interaction term are presented, since the individual effects of insurance and incentive are not important even in the full sample. The first column copies the results for the full sample from Table 1.8, and the second column presents the results for the one-visit sub-sample. Given that around 40% of the visits are dropped from the analysis, the significance levels of tests are expected to decrease. The

[24] Two doctors ended up with three visits because experimenters did not recognize the doctors during the second visits. The doctors did not appear to recognize the experimenter either.

purpose of this exercise is to compare the magnitude of the point estimates. The estimated effects on prescription for triglycerides and monthly drug expenditure are roughly the same between the full sample and the sub-sample. There are no statistically significant differences between the coefficients across samples. Overall, the interaction effects of insurance and incentive are not driven by doctors with multiple visits.

On the other hand, the sub-sample of doctors with multiple visits can be used for the within-doctor comparison. The pair-wise comparison between the insured-incentive setting and any of the other three settings shows the following results: (1) in terms of monthly drug expenditure, 15 out of 22 comparisons support the main findings while 4 comparisons go against them, and 3 comparisons par; (2) in terms of drugs for triglycerides, 3 out of 11 comparisons support the main results and none of the comparisons go against them.

1.6.2 Contaminated cases

There are some contaminated cases, in which the actual insurance status or incentive status deviates from the intended ones. In one case intended for the insured-incentive intervention, the insurance status was mislabeled at the beginning, and the experimenter corrected the insurance status before she finished the prescription.[25] In four cases under no-incentive interventions in which the experimenters said they did not intend to purchase drugs at the doctor's hospital, doctors tried to persuade the experimenters to buy drugs there. They were persuading as they were prescribing, so their prescriptions might have been affected by their expectations about purchases. On the other hand, in two cases under uninsured-incentive interventions, doctors suggested that patients purchase drugs locally rather than in their hospitals. Except for the case with mislabeled insurance status, other contaminations are doctor-driven, so all the contaminated cases are kept under their intended intervention types for the main analysis. (See Appendix 4 for a list of contaminated cases.) For the robustness check, the third column of Table 1.10 presents the estimations of the interaction effects by dropping hospitals with contaminated cases, and the results are close to those in the full sample.

1.6.3 Availability of drugs in other regions

Under no incentive, doctors prescribe drugs for patients to purchase in other regions, so the prescriptions might be affected by doctors' expectations on whether certain drugs are available in other regions. A particular concern is whether the possibly limited availability of drugs restricts the prescription of drugs under the insured-no-incentive intervention, which is important for the estimation of agency problems and the interaction effects. However, this concern does not appear to impose any strong threat to the analysis. First, for triglycerides, since the focus is on whether doctors prescribe drugs or not, the concern about availability requires doctors to expect there are no drugs for triglycerides in a certain region, which is very unlikely. Second, the shares of brand-name drugs for diabetes and hypertension are very similar across incentive statuses, which suggest there is no concern along the brand-generic dimension. Third, in terms of varieties of drugs for diabetes and hypertension, except for two Chinese patent drugs, all other drugs prescribed in the experiment are listed in the National Catalog of Basic Drugs under Insurance Coverage (2005). As provinces tend to add more drugs to provincial catalogs and rarely remove drugs, there is no reason to expect those drugs are not available in other regions. The fourth

[25] The visit representative did so by requesting a second copy of the prescription for reimbursement purposes. Although doctors are required to print out two copies of prescriptions for all patients, some hospitals only print out one copy for those uninsured, because only those insured need a second copy for the purpose of reimbursement.

column in Table 1.10 presents the estimation of interaction effects by dropping the two hospital-departments which prescribed the two drugs not included in the national catalog, and the results are similar to those in the full sample.

1.7 Conclusions

This study uses a randomized field experiment to demonstrate that doctors prescribe more expensive drugs to insured patients. The experiment was also designed to distinguish between the agency hypothesis, where doctors act in their self financial interest, and the considerate doctor explanation, where the doctor prescribe taking into account the tradeoff between the efficacy of drugs and patients' ability to pay. The controlled experiment not only solves the usual endogenous problems associated with the insurance and incentive status, but also avoids the confounding effect in observational studies where insured patients possibly request more drugs.

The results show that the doctor is more likely to prescribe inappropriate drugs or more expensive drugs to patients with insurance than to patients without insurance if a doctor expects to receive a proportion of prescribed drug expenditures. The prescriptions to insured patients cost more than 43% of those to uninsured patients on average. However, if the doctor does not have this financial incentive, the prescriptions are similar for insured and uninsured patients. Thus, the increases in drug expenditures are due to an agency problem, where doctors increase their earnings by exploiting the information asymmetry and taking advantage of the greater purchasing power of insured patients, and not to the caring doctor hypothesis. In practice, as a large proportion of patients buy drugs at the hospital where they see a doctor, this study suggests a substantial amount of increases in health expenditures under insurance coverage are in inefficient use due to doctors.

The Beijing hospitals and doctors in the experiment are among the best in China, and they are not a nationally representative sample. And the public expenditure insurance is relatively more generous than many other types of health insurance. Needless to say, the quantitative estimates are highly specific to the particular group of doctors, and the type of insurance. However, the structures of doctors' financial incentives are similar across the country, and different types of insurance share several important characteristics. Therefore, at a broad level, this study illustrates the importance of the interaction between insurance and incentive.

The analysis suggests that although insured patients receive more drugs or more expensive drugs, they may not be receiving better health treatments. It draws cautions in interpreting the welfare consequences of increased health expenditures under insurance coverage in China. It highlights the importance of coordinating an expansion of health insurance and a reform of doctors' incentive structure. In China, as well as in many other developing countries, there are market failures along various dimensions. These market failures interact – the lack of health insurance limits the extent of agency problems while expanding insurance coverage can exacerbate agency problems. This experiment illustrates, dealing with one failure while ignoring the other may actually generate substantial inefficiency.

Chapter 1. References

Anderson, Michael, Carlos Dobkin, and Tal Gross. (2010) The Effect of Health Insurance Coverage on the Use of Medical Services. NBER Working Paper No. 15823.

Arrow, Kenneth. (1963) Uncertainty and the Welfare Economics of Medical Care. The American Economic Review, Vol. 53(5): 941-973.

Ayres, Ian and Peter Siegelman. (1995) Race. Gender Discrimination in Bargaining for a New Car. The American Economic Review, Vol. 85(3): 304-321.

Card, David, Carlos Dobkin, and Nicole Maestas. (2008) The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare. The American Economic Review Vol. 98(5): 2242-2258.

Currie, Janet, Wanchuan Lin and Wei Zhang. (2010) Patient Knowledge and Antibiotic Abuse: Evidence from an Audit Study in China. NBER Working Paper 16602

Dalen, Dag, Enrico Sorisio, and Steinar Strom. (2010) Choosing Among Competing Blockbusters: Does the Identity of the Third-Party Payer Matter for Prescribing Doctors? CESifo Working Paper Series No. 3227.

Das, Jishnu and Jeffrey Hammer. (2005) Which Doctor? Combining Vignettes and Item Response to Measure Clinical Competence. Journal of Development Economics, Vol. 78(2): 348-383.

Das, Jishnu, and Jeffrey Hammer. (2007) Money for nothing: The dire straits of medical practice in Delhi, India. Journal of Development Economics, Vol. 83(1): 1-36.

Evans, R.G. (1974) Supplier-Induced Demand: Some Empirical Evidence and Implications. The Economics of Health and Medical Care, M. Periman, eds. London: Macmillan.

Feldstein, Martin. (1970) The Rising Price of Physician's Services. The Review of Economics and Statistics, Vol. 52(2): 121-133.

Gemmill, Marin, Sarah Thomson, and Elias Mossialos. (2008) What impact do prescription drug charges have on efficiency and equity? Evidence from high-income countries. International Journal for Equity in Health, Vol. 7(1):12-33.

Gruber, Jonathan and Maria Owings. (1996) Physician financial incentives and Cesarean section delivery. RAND Journal of Economics, Vol. 27(1): 99-123.

Iizuka, Toshiaki. (2007) Experts' agency problems: evidence from the prescription drug market in Japan. The Rand Journal of Economics, Vol. 38(3): 844-862.

Kessel, Reuben. (1958) Price discrimination in medicine. Journal of Law and Economics, Vol. 1(1): 20-53.

Kravitz, Richard, Ronald Epstein, Mitchell Feldman, Carol Franz, Rahman Azari, Michael Wilkes, Ladson Hinton, and Peter Franks. (2005) Influence of Patients' Requests for Direct-to-Consumer Advertised Antidepressants: A Randomized Controlled Trial. The Journal of American Medical Association, Vol. 293(16): 1995-2002.

List, John. (2004) The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field. The Quarterly Journal of Economics, Vol. 119(1): 49-89.

Lundin, Douglas. (2000) Moral Hazard in Physician Prescription Behavior. Journal of Health Economics, Vol. 19(3): 632-662.

McGruire, Thomas. (2000) Physician Agency. Handbook of Health Economics, Vol. 1A. Anthony Culyer and Joseph Newhouse, eds. Amsterdam: Elsevier. 461-536.

McGuire, Thomas, and Mark Pauly. (1991) Physician response to fee changes with multiple payers. Journal of Health Economics, Vol. 10(4): 385-410.

McKinlay, John, Deborah Potter, and Henry Feldman. (1996) Non-medical Influences on Medical Decision-Making. Social Science and Medicine, Vol. 42(5): 769-776.

Mort, Elizabeth, Jennifer Edwards, David Emmons, Karen Convery, and David Blumenthal. (1996) Physician Response to Patient Insurance Status in Ambulatory Care Clinical Decision-Making: Implications for Quality of Care. Medical Care, Vol. 34(8): 783-797.

Rice, Thomas and Karen Matsuoka. (2004) The Impact of Cost-sharing on Appropriate Utilization and Health Status: a Review of the Literature on Senior. Medical Care Research and Review, Vol. 61(4): 415-452.

Schneider, Henry. (2009) Agency Problems and Reputation in Expert Services: Evidence from Auto Repair. Johnson School Research Paper Series No. 15-07.

Wagstaff, Adam, Magnus Lindelow, Gao Jun, Xu Ling, and Juncheng Qian. (2009) Extending health insurance to the rural population: An impact evaluation of China's new cooperative medical scheme. Journal of Health Economics 28(1): 1–19.

Zweifel, Peter and Willard Manning. (2000) Moral Hazard and Consumer Incentives in Health Care. Handbook of Health Economics, Vol. 1A. Anthony Culyer and Joseph Newhouse, eds. Amsterdam: Elsevier. 409-459.

Table 1.1 Summary Statistics

Variables	Obs	Mean	Std.	Min	Max
For Patient 1					
Visited by researcher (0/1)	100	0.48	0.50	0	1
Expert visit (0/1)	100	0.33	0.47	0	1
Male doctor (0/1)	100	0.34	0.48	0	1
Doctor age (years)	100	43.00	8.29	30	65
Raw drug expenditure (yuan)	50	534.06	252.76	115.12	1394.86
Prescription for triglycerides (0/1)	100	0.43	0.50	0	1
Monthly expenditure D&H (yuan)	100	374.57	151.04	109.38	762.52
Number of drugs D&H	100	2.39	0.65	1	4
Unit of drugs D&H	100	2.37	0.78	1	5
Share of branded drugs D&H (0~1)	100	0.72	0.32	0	1
For Patient 2					
Visited by researcher (0/1)	96	0.49	0.50	0	1
Expert visit (0/1)	96	0.28	0.45	0	1
Male doctor (0/1)	96	0.42	0.50	0	1
Doctor age (years)	96	44.95	7.56	30	60
Raw drug expenditure (yuan)	48	349.43	143.38	122.63	794.28
Monthly expenditure D&H (yuan)	96	301.45	116.00	102.92	761.84
Number of drugs D&H	96	2.13	0.42	2	5
Unit of drugs D&H	96	2.08	0.47	1.5	4
Share of branded drugs D&H (0~1)	96	0.83	0.26	0	1

Notes: “D&H” represents “for diabetes and hypertension only”.

Table 1.2 Visit Characteristics by Intervention Types

Variables	Insurance Incentive	No Insurance Incentive	Insurance No Incentive	No Insurance No Incentive	F-test	P value
For both patients						
Visited by researcher (0/1)	0.49	0.51	0.51	0.43	0.64	0.59
s.e.	(0.07)	(0.07)	(0.07)	(0.07)		
Expert visit (0/1)	0.29	0.27	0.31	0.37	0.80	0.51
s.e.	(0.07)	(0.06)	(0.07)	(0.07)		
Male doctor (0/1)	0.37	0.33	0.41	0.41	0.53	0.66
s.e.	(0.07)	(0.07)	(0.07)	(0.07)		
Doctor’s age (years)	43.57	43.67	44.08	44.49	0.14	0.95
s.e.	(1.17)	(1.19)	(1.15)	(1.08)		
Observations	49	49	49	49		

Table 1.3 Outcomes by Intention Types

Dependent variables	Insurance Incentive	No Insurance Incentive	Insurance No Incentive	No Insurance No Incentive
For both patients				
Raw drug expenditure (yuan)	522.11	365.14	-	-
s.e.	(35.80)	(23.63)	-	-
Prescription for triglycerides (0/1)	0.64	0.40	0.28	0.40
s.e.	(0.10)	(0.10)	(0.09)	(0.10)
Monthly drug expenditure D&H (yuan)	424.78	298.71	324.50	307.03
s.e.	(23.54)	(15.84)	(18.95)	(15.44)
Number of drugs D&H	2.47	2.20	2.18	2.18
s.e.	(0.10)	(0.08)	(0.07)	(0.06)
Unit of drugs D&H	2.53	2.09	2.16	2.12
s.e.	(0.11)	(0.08)	(0.09)	(0.07)
Share of branded drugs D&H (0~1)	0.83	0.68	0.81	0.80
s.e.	(0.04)	(0.05)	(0.03)	(0.04)
Obs. for triglycerides	25	25	-	-
Obs. for other variables	49	49	49	49

Notes: “D&H” represents “for diabetes and hypertension only”.

Table 1.4 Effects of Insurance When Doctors Have Financial Incentives

	Both patients		Patient 1		Patient 2	
Dependent variables	(1)	(2)	(3)	(4)	(5)	(6)
Raw drug expenditure (yuan)	156.97***	155.49***	202.77***	202.13***	109.26***	99.07***
	(36.35)	(37.67)	(59.01)	(64.56)	(34.81)	(27.38)
Prescription for triglycerides (0/1)	0.24*	0.26*	0.24*	0.26*		
	(0.13)	(0.14)	(0.13)	(0.14)		
Monthly drug expenditure D&H (yuan)	126.07***	125.53***	146.62***	140.80***	104.67***	92.93***
	(25.48)	(25.46)	(31.63)	(31.30)	(33.07)	(24.58)
Number of drugs D&H	0.27**	0.27**	0.32**	0.29**	0.21	0.18
	(0.12)	(0.12)	(0.13)	(0.10)	(0.17)	(0.12)
Unit of drugs D&H	0.44***	0.45***	0.57***	0.58***	0.31*	0.25**
	(0.12)	(0.11)	(0.14)	(0.14)	(0.17)	(0.10)
Share of branded drugs D&H (0~1)	0.15**	0.14**	0.12	0.10	0.18***	0.17***
	(0.05)	(0.05)	(0.10)	(0.09)	(0.06)	(0.05)
Control for:						
Hospital fixed effects	Y	Y	Y	Y	Y	Y
Visit characteristics	N	Y	N	Y	N	Y
Obs for triglycerides	50	50	50	50	-	-
Obs for other variables	98	98	50	50	48	48

Notes: "D&H" represents "for diabetes and hypertension only". The dependent variables are listed on the left, and coefficients are from separate regressions. Standard errors, clustered at the hospital level, are in parentheses. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 1.5 Effects of Insurance When Doctors Have NO Incentives

	Both patients		Patient 1		Patient 2	
Dependent variables	(1)	(2)	(3)	(4)	(5)	(6)
Prescription for triglycerides (0/1)	-0.12	-0.07	-0.12	-0.07		
	(0.11)	(0.09)	(0.11)	(0.09)		
Monthly drug expenditure D&H (yuan)	17.47	16.67	13.70	13.79	21.40	20.21
	(23.00)	(23.38)	(34.54)	(35.82)	(23.09)	(20.06)
Number of drugs D&H	0.00	-0.01	-0.04	-0.04	0.04	0.03
	(0.08)	(0.09)	(0.16)	(0.14)	(0.04)	(0.03)
Unit of drugs D&H	0.04	0.02	-0.09	-0.08	0.18*	0.15
	(0.11)	(0.10)	(0.22)	(0.20)	(0.09)	(0.09)
Share of branded drugs D&H (0~1)	0.01	0.01	0.04	0.03	-0.02	-0.03
	(0.04)	(0.04)	(0.07)	(0.07)	(0.06)	(0.06)
Control for:						
Hospital fixed effects	Y	Y	Y	Y	Y	Y
Visit characteristics	N	Y	N	Y	N	Y
Obs for triglycerides	50	50	50	50	-	-
Obs for other variables	98	98	50	50	48	48

Notes: "D&H" represents "for diabetes and hypertension only". The dependent variables are listed on the left, and coefficients are from separate regressions. Standard errors, clustered at the hospital level, are in parentheses. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 1.6 Effects of Incentives When Patient Has Health Insurance

	Both patients		Patient 1		Patient 2	
Dependent variables	(1)	(2)	(3)	(4)	(5)	(6)
Prescription for triglycerides (0/1)	0.36***	0.35***	0.36***	0.35***		
	(0.11)	(0.10)	(0.11)	(0.10)		
Monthly drug expenditure D&H (yuan)	100.28***	101.22***	104.69**	105.16**	95.68***	110.52***
	(31.65)	(30.93)	(47.58)	(49.43)	(30.97)	(29.33)
Number of drugs D&H	0.29**	0.29**	0.28	0.27*	0.29*	0.29**
	(0.11)	(0.11)	(0.17)	(0.15)	(0.14)	(0.12)
Unit of drugs D&H	0.37***	0.38***	0.47**	0.47**	0.26*	0.30**
	(0.13)	(0.13)	(0.21)	(0.22)	(0.15)	(0.14)
Share of branded drugs D&H (0~1)	0.02	0.02	-0.07	-0.07	0.12*	0.14**
	(0.05)	(0.04)	(0.09)	(0.08)	(0.06)	(0.06)
Control for:						
Hospital fixed effects	Y	Y	Y	Y	Y	Y
Visit characteristics	N	Y	N	Y	N	Y
Obs for triglycerides	50	50	50	50	-	-
Obs for other variables	98	98	50	50	48	48

Notes: "D&H" represents "for diabetes and hypertension only". The dependent variables are listed on the left, and coefficients are from separate regressions. Standard errors, clustered at the hospital level, are in parentheses. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 1.7 Effects of Incentives When Patient Has NO Insurance

	Both patients		Patient 1		Patient 2	
Dependent variables	(1)	(2)	(3)	(4)	(5)	(6)
Prescription for triglycerides (0/1)	0.00	-0.01	0.00	-0.01		
	(0.13)	(0.13)	(0.13)	(0.13)		
Monthly drug expenditure D&H (yuan)	-8.32	-5.09	-28.22	-14.96	12.42	16.38
	(21.19)	(19.38)	(34.97)	(33.26)	(22.30)	(23.83)
Number of drugs D&H	0.02	0.02	-0.08	-0.02	0.13*	0.13*
	(0.09)	(0.09)	(0.15)	(0.16)	(0.07)	(0.06)
Unit of drugs D&H	-0.03	-0.04	-0.19	-0.13	0.14	0.12
	(0.09)	(0.10)	(0.18)	(0.20)	(0.10)	(0.12)
Share of branded drugs D&H (0~1)	-0.12*	-0.11*	-0.16*	-0.14*	-0.08	-0.05
	(0.06)	(0.06)	(0.09)	(0.08)	(0.10)	(0.07)
Control for:						
Hospital fixed effects	Y	Y	Y	Y	Y	Y
Visit characteristics	N	Y	N	Y	N	Y
Obs for triglycerides	50	50	50	50	-	-
Obs for other variables	98	98	50	50	48	48

Notes: "D&H" represents "for diabetes and hypertension only". The dependent variables are listed on the left, and coefficients are from separate regressions. Standard errors, clustered at the hospital level, are in parentheses. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 1.8 The Interaction Effects of Insurance and Incentive

	Both patients		Patient 1		Patient 2	
Dependent variables	(1)	(2)	(3)	(4)	(5)	(6)
Prescription for triglycerides (0/1)	0.36**	0.34**	0.36**	0.34**		
	(0.14)	(0.14)	(0.14)	(0.14)		
Monthly drug expenditure D&H (yuan)	108.60***	103.71**	132.92**	128.49**	83.27**	72.68*
	(37.57)	(38.37)	(56.56)	(56.73)	(38.38)	(37.00)
Number of drugs D&H	0.27**	0.26**	0.36*	0.36**	0.17	0.14
	(0.13)	(0.13)	(0.18)	(0.17)	(0.17)	(0.15)
Unit of drugs D&H	0.40**	0.39**	0.67**	0.64**	0.13	0.11
	(0.16)	(0.16)	(0.27)	(0.28)	(0.18)	(0.16)
Share of branded drugs D&H (0~1)	0.14*	0.13*	0.08	0.07	0.20*	0.18*
	(0.07)	(0.07)	(0.13)	(0.12)	(0.11)	(0.09)
Control for:						
Hospital fixed effects	Y	Y	Y	Y	Y	Y
Visit characteristics	N	Y	N	Y	N	Y
Obs for triglycerides	100	100	100	100	-	-
Obs for other variables	196	196	100	100	96	96

Notes: "D&H" represents "for diabetes and hypertension only". The dependent variables are listed on the left, and coefficients are from separate regressions. Standard errors, clustered at the hospital level, are in parentheses. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 1.9 Effects of All Variables on Prescriptions

	Dependent variables:				
Independent variables	Prescription for Triglycerides	Monthly Expenditure	Number of Drugs	Unit of Drugs	Share of Brand drugs
Insurance*Incentive	0.34**	103.71**	0.26**	0.39**	0.13*
	(0.14)	(38.37)	(0.13)	(0.16)	(0.07)
Insurance	-0.10	21.07	0.01	0.05	0.02
	(0.10)	(24.33)	(0.09)	(0.11)	(0.04)
Incentive	0.02	-3.49	0.03	-0.02	-0.11*
	(0.13)	(21.86)	(0.09)	(0.10)	(0.06)
Visited by researcher	-0.11	-16.17	0.05	-0.00	-0.02
	(0.08)	(20.32)	(0.09)	(0.10)	(0.05)
Expert visit	0.09	44.82**	0.12	0.10	0.13*
	(0.13)	(21.40)	(0.11)	(0.13)	(0.06)
Male doctor	0.07	-1.67	-0.04	-0.03	0.07
	(0.17)	(25.74)	(0.12)	(0.15)	(0.05)
Doctor's age	0.00	-1.14	0.00	0.01	-0.01**
	(0.01)	(1.66)	(0.01)	(0.01)	(0.00)
Hospital fixed effects	Y	Y	Y	Y	Y
Observations	100	196	196	196	196

Notes: Standard errors, clustered at the hospital level, are in parentheses. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 1.10 Interaction Effects of Insurance and Incentive in Different Samples

Dependent variables	Full sample	One-visit sample	No-contamination sample	National-catalog sample
Prescription for triglycerides	0.34**	0.35	0.32**	0.40**
	(0.14)	(0.25)	(0.15)	(0.16)
Monthly drug expenditure D&H	103.71**	97.82*	94.37**	97.82**
	(38.37)	(56.12)	(38.03)	(40.70)
Number of drugs D&H	0.26**	0.44*	0.33**	0.22*
	(0.13)	(0.21)	(0.13)	(0.13)
Unit of drugs D&H	0.39**	0.45*	0.45**	0.35*
	(0.16)	(0.23)	(0.17)	(0.17)
Share of branded drugs D&H	0.13*	0.07	0.11	0.15**
	(0.07)	(0.09)	(0.07)	(0.06)
Control for:				
Hospital fixed effects	Y	Y	Y	Y
Visit characteristics	Y	Y	Y	Y
Obs for triglycerides	100	62	76	92
Obs for other variables	196	114	168	188

Notes: "D&H" represents "for diabetes and hypertension only". The dependent variables are listed on the left, and coefficients are from separate regressions. Standard errors, clustered at the hospital level, are in parentheses. * denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Appendix 1.1: Basic information of hypothetical patients

Patient 1: male, 66 years old.
Basic description of health problems is:
He is recently tested with problems.
Fasting blood sugar is 7.5 mmol/L and 2-hour postprandial blood sugar is 11.5 mmol/L.
Fasting c-peptide is 2.1 ng/ml and 2-hour postprandial c-peptide is 10.2 ng/ml.[26]
Hemoglobin A1C level is 7.8%.
Blood pressure is 160/90.
Triglyceride is 2.3 mmol/L (equivalent to 199 mg/dL). Cholesterin is normal.
Hearst rate is 80.
He does not feel sick.
If asked by doctors, prepared answers include:
Liver function and kidney function are normal.
Height is 175cm and weight is 70kg.
He does not smoke and he drinks little.
He does not eat too much food.
He does not have family history.

Patient 2: male, 65 years old.
Basic description of health problems is:
He has hypertension.
He is taking the brand-name NIFEDIPINE control-released table one tablet per day.
Blood pressure is 155/80.
Heart rate is 75.
He does not feel faint.
If asked by doctors, prepared answers include:
Liver function and kidney function are normal.
Highest blood pressure can reach 170.
He has hypertension for three years.
He has taken Bai Xin Tong for several months.
He does not smoke and he drinks little.

[26] The test of C-peptide is not standardized, and different labs have different normal range for fasting C-peptide. From the internet, we see more three standards: (1) 1.49-3.41 ng/ml; (2) 1.1-4.4 ng/ml; (3) 3.77-4.23 ng/ml. Overall, the constructed C-peptide indicates that the patient is of Type II diabetes, which is obvious even without the C-peptide test given that the patient is already 66 years old. Some doctors ask about the reference range; if so, the answer is from zero point something to three points something. Some doctors may think the C-peptide level is below the normal.

Appendix 1.2: List of doctor refusals

Patient No.	Intended Intervention[1]	Experimenter[2]	Male Doctor	Doctor Age
1	D	0	0	35
1	A	0	0	50
1	C	1	1	50
1	C	1	0	30
1	B	1	0	35
2	D	0	1	55
2	A	0	1	40
2	D	1	0	50
2	B	1	0	40

Notes:
1. A. Insured-incentive; B. uninsured-incentive;
   C. insured-no-incentive; D. uninsured-no-incentive.
2. 0 indicators the assistant, and 1 represents the researcher.

Appendix 1.3: List of unit dosage

Drug names	Instructions (dose*times/day)	Unit
Drugs for diabetes:		
Glipizide	5mg*1	1
Gliquidone	30mg*3	1
Glimepiride	2mg*1	1
Gliclazide	30mg*1	1
Nateglinide	120mg*3	1
Repaglinide	1mg*3	1
Metformin	0.5g*3	1
Pioglitazone	15mg*1	1
Rosiglitazone	4mg*1	1
Acarbose	50mg*3	1
Calcium dobesilate	0.5g*3	1
Chinese Patent Drug 1	#8*3	1
Chinese Patent Drug 2	#1*3	1
Chinese Patent Drug 3	#4*3	1
Chinese Patent Drug 4	#6*3	1
Drugs for hypertension:		
Perindopril	4mg*1	1
Benazepril	10mg*1	1
Fosinopril	10mg*1	1
Enalapril	10mg*1	1
Ramipril	5mg*1	1
Valsartan	80mg*1	1
Irbesartan	150mg*1	1
Telmisartan	80mg*1	1
Losartan Potassium	50mg*1	1
Candesartan	4mg*1	1
Nifedipine	30mg*1	1
Lacidipine	4mg*1	1
Felodipine	5mg*1	1
(Lev)amlodipine	5mg*1	1
Bisoprolol	5mg*1	1
Metoprolol	47.5mg*1	1
Arotinolol	10mg*1	1
Carvedilol	12.5mg*1	1
Indapamide	1.5mg*1	1
Amiloride	25mg*1	1
Antisterone	20mg*1	1
Hydrochlorothiazide	25mg*1	1
Chinese Patent Drug 5	#2*2	1
Chinese Patent Drug 6	#3*3	1
Chinese Patent Drug 7	#3*3	1
Losartan Potassium and Hydrochlorothiazide	(50mg+12.5mg)*1	1.5
Irbesartan and Hydrochlorothiazide	(150mg+12.5mg)*1	1.5

Appendix 1.4: List of contaminated cases

Patient No.	Intended Intervention[1]	Contaminated Intervention[1]	Experimenter[2]	Male Doctor	Doctor Age
1	D	B	1	1	40
1	C	A	1	0	40
1	D	B	0	0	50
1	D	B	0	1	55
1	B	D	1	0	55
1	A	B	1	0	35
2	B	D	0	0	50

Notes:

1. A. Insured-incentive; B. uninsured-incentive;
   C. insured-no-incentive; D. uninsured-no-incentive.
2. 0 indicators the assistant, and 1 represents the researcher.

Chapter 2. Testing Peer Effect among College Students: Evidence from an Unusual Admission Policy Change in China

It has long been believed that peers play an important role in determining an individual's behaviors and educational outcomes. Concern with peer effects can be traced back to two thousand years ago in a Chinese story, "Three Moves of Meng's Mother."[1] Peer influence has been noted in a wide range of policy issues, including neighborhood relocation programs, desegregation, academic tracking, and affirmative action (Kling, Liebman and Katz, 2007; Angrist and Lang, 2004; Duflo, Dupas and Kremer, 2007; Card and Krueger, 2005). Understanding the effects of peers on students is important for school management, admission policy, and school choice.

Despite strong academic interest in peer effects, several issues have contributed to a lack of clear evidence on peer effects, especially among college students. Manski (1993) discusses reflection problems and points out that selection bias is one of several major econometric challenges in identifying peer effects. Selection bias arises because individuals choose their peer group, which makes it difficult to separate true peer effect from the selection effect. To circumvent selection bias, Sacerdote (2001), Zimmerman (2003), and Foster (2006) exploit random dormitory assignments in colleges in the U.S., but their findings disagree with each other: some papers report modest but statistically significant peer effects among roommates, while others find little support for the existence of peer effects among college students. Although random dormitory assignment solves the selection bias problem, it is questionable whether roommates are well defined peers. Stinebrickner and Stinebrickner (2006) discuss that college students establish networks of friends extending beyond the roommate level and that interactions among roommates are limited. In other words, roommates may not be close peers, which may explain the modest or non-existent effects reported in these studies. A further issue in the empirical analysis is whether there is enough variation in peer characteristics for precise estimation. Lyle (2007, 2009) identifies peers as students who study and socialize together in a class of 35 students, relieving the weak-peer identifier problem, and he argues that the limited variation in peer characteristics may explain the null effect of the average peer test scores in his sample. The importance of variation in peer abilities is further confirmed by Carrell, Fullerton, and West (2009) who find strong peer effects when peer groups are not constructed to have an even distribution of academic ability.

An unusual change in admission policy at a prestigious Chinese university provides a rare opportunity for studying peer effects among college students. This design addresses selection bias, and leverages strong peer interaction and wide variation in peer characteristics. The change in admission policy brought a large number of specially admitted low-scoring students into many academic departments which normally only admitted students with much higher scores in college entrance exams. The inflow of specially admitted students imposed an exogenous shock for the

[1] This is documented by Liu, Xiang (BC77–BC6). The story tells that Meng's mother is concerned about their social environments and moves several times to find a good location so that Meng can study well. Eventually, Meng becomes a philosopher famous in Chinese history.

regular students by changing the composition of students' characteristics in the relevant academic departments. The number of specially admitted students varied across departments; some departments did not admit any low-scoring students whereas in other departments the share of low-scoring students reached as high as 40%. The school arrangement further facilitated the exploration of peer effects because all the specially admitted students lived and studied together with regular students, and students in the same department-year interacted intensively with each other.

This paper adopts a difference-in-difference estimation strategy similar to that presented in Imberman, Kugler and Sacerdote (2009), and explores the relationships among the characteristics of specially admitted students and the test outcomes of the other students. We interpret the estimated effects as causal effects of low-scoring students on their peers given that the following two findings: first, evidence indicates that regular students in the same departments are comparable across years, and the inflow of low-scoring students does not seem to alter the backgrounds of regular students in the affected departments; second, evidence also suggests that omitted variables do not seem to contribute to the estimated effects.

Although one important policy implication of peer effect studies pertains to the optimal design of admission policy, to the best of our knowledge, this is the first analysis on peer consequences of an admission policy change among college students. Our data suggest that specially admitted low-scoring students have significantly reduced the performance of the regular students in College English Tests. The detrimental effects from specially admitted students are concentrated among regular students whose English ability was below average.

This paper is organized as follows. Section 2.1 provides background information on the admission policy change and the school arrangement in the sample university. Section 2.2 describes the data and presents the summary statistics. Section 2.3 describes the estimation strategy and presents results. Section 2.4 conducts robustness checks to rule out the selection bias and omitted variable bias. Section 2.5 concludes the paper.

2.1 Background

In China, college programs are ranked into different rounds in the admission process. Programs in earlier rounds are allowed to admit students earlier. Students admitted by programs in earlier rounds leave the applicant pool and are not available for programs in later rounds. Round 1 universities are the universities whose bachelor programs are allowed to admit students in the first round. In normal years, 1994 and 1995 in this study, Round 1 universities only admitted students for their bachelor programs during the first round.

In 1996, the province of Jiangsu implemented an unusual change in the admission policy and created Round 3 bachelor programs, which allowed several Round 1 universities to admit students for their bachelor programs after Round 2 universities completed bachelor program admission.[2] The middle 1990s saw a large movement in the commercialization of higher education in China, and the creation of Round 3 bachelor programs was one of the trials, which was claimed to serve at least two purposes: to help universities to increase funding and to train human resources for local governments. Given that there are about 150 Round 1 and 750 Round 2 universities, and the sample university is a top-ranking Round 1 university, it is not surprising

[2] Round 3 programs remain in the following years, but the way of admission changed largely after 1996.

to see that students admitted through the Round 3 programs differ markedly in educational backgrounds from the regular Round 1 students.

The distribution of Round 3 students varies across departments. Departments related to public affairs, such as public administration, social welfare, and urban planning, were strongly affected; they admitted a larger number of Round 3 students, which conformed to the stated purpose of the Round 3 programs: to train human resources, particularly for local governments. Many departments did not admit any Round 3 students, including both popular departments, such as bio-chemistry and the medical school, and less popular departments, such as geology, astronomy, aerology, and history.

The department-year is the primary unit of affiliation for students. Students are assigned to an academic department in the admission letter before they enter the university. It is difficult for students to change their assigned academic departments, so nearly all students remain with their cohort over their four years of undergraduate study. Round 3 students had much lower scores in the college entrance exam and paid much higher tuition; except that difference, all bachelor students in the same department-year lived and studied together. All the students lived in university dormitories. Separated by gender, bachelor students in the same department-year are assigned to live together in the same room or neighboring rooms for four years.[3] Students in the same department-year took most classes together in the first two years. Many extracurricular activities are organized within the department-year, including class meetings, sports, trips, parties, and so on.

It is worth noting that many departments held associate degree programs during the studied time period. Since the admission order of the associate programs fell after the Round 2 bachelor programs in 1994 and 1995, associate students might be similar to Round 3 students in 1996 in terms of test performance in the college entrance exam. For the sake of our analysis, the major difference between Round 3 bachelor programs in 1996 and the associate programs in earlier years is that Round 3 bachelor students lived and studied together with Round 1 bachelor students while the associate students did not. Therefore, we treat Round 3 students as peers of Round 1 students but ignore associate students in this analysis.

2.2 Data

The data for this study come from archived student academic records, which provide information on National College Entrance Exam (NCEE) scores, College English Test (CET) outcomes, and some demographics, such as age, gender, and home address. We identify whether a student is a Round 3 student using the Admission Approval Table, which documents the program to which a student is admitted. The data cover all the students in 23 selected academic departments entering the university in 1994, 1995, and 1996.[4,5]

[3] Male and female students in the same department-year do not live close to each other.

[4] The Archive Office was reluctant to provide data for all departments, but allowed us to choose departments. We acquired data from departments which admitted many Round 3 students and those which did not admit any Round 3 students. We exclude the school of foreign languages and other departments that we suspect experienced structural changes during the period.

[5] For this analysis, we exclude students who do not study English as their first foreign language, who are admitted via the university but then sent to a joint program abroad, and who come from Hongkong, Macau, or foreign countries. For the presented results, we exclude several students who had NCEE scores too low to be regular Round

Most students, except a group of recommended students, took the National College Entrance Exam to be admitted to college. Two sets of exams are administrated every year: the "art" set and the "science" set, each containing five subject exams. The English test is common to both sets. The total score is the sum of scores in the five subject exams. Since the National College Entrance Exam is administrated by each province separately every year and the raw scores are not comparable across years even within provinces, we standardize all the scores using the averages and standard deviations calculated from the corresponding Round 1 students by year, province, and exam set or exam subject.[6] Round 3 students are mainly those who fail to be admitted by Round 1 and Round 2 programs. As the NCEE total score is the major criterion in college admissions, it is expected that the NCEE scores of Round 3 students differ from those of regular students, as illustrated in Figure 2.1 and Figure 2.2. Since all Round 3 students are from Jiangsu, we compare the NCEE scores among students from Jiangsu. The top panel of Figure 2.1 demonstrates the distributions of the NCEE total scores for students in 1994 and 1995 combined, the middle panel displays the scores of regular Round 1 students in 1996, and the bottom panel is for the specially admitted Round 3 students in 1996.[7] By construction, the standardized scores of regular students in each province-year are scattered around zero, with a standard deviation of 1. And Round 3 students' NCEE total scores fall mainly between -6 and -2. Figure 2.2 illustrates the distributions of the NCEE English scores for the same three groups. The English scores of Round 3 students are also largely different from those of Round 1 students too.

We measure students' performance in College English Test Level 4 (CET-4) and Level 6 (CET-6) exams, which offer several advantages over other possible outcomes. First, the CETs are important. Passing the CET-4 is required for all college students to obtain a bachelor degree certificate.[8] Passing CET-6 serves to further signal a student's English ability. The CET certificates are important documents in the job application package. Most students spend more than four hours per week in class for four semesters. Second, CET test outcomes are comparable across departments. The CETs are organized at the national level and graded at the provincial level. In addition, 90% of the score is determined objectively; only 10% of the score is determined by an essay. In general, the grading is reliable and comparable. Third, all the English teachers are from the Division of College English, and they use identical syllabus. Finally, English classes tend to be organized on the department-year basis. If the size of the department-year is large, students will be separated into several classrooms. Students are ordered "alphabetically" according to their last names,[9] and Round 1 students and Round 3 students are mixed in the classrooms.

Table 1 presents the summary statistics for the CET outcomes in Panel A. The raw scores for the CETs are in the scale of 0 to 100, and a student earns a "Pass" with a score above 60 and an

---

1 students but who nevertheless cannot be identified as Round 3 students; however, excluding them does not affect the results.

[6] After standardization, the scores are still not comparable across provinces. It is arguable whether the scores are comparable across years within provinces, especially for those provinces with a small number of admission quotas for the sample university, because the allocation of quotas to academic departments tends to vary greatly across years in those provinces.

[7] Students who gain admission via recommendation are not included since they do not have official NCEE scores.

[8] Normally, when they graduate, students can obtain two certificates—a graduation certificate and a bachelor degree certificate. Students can still graduate with a graduation certificate if they could not pass the CET-4. Students can be exempt from the CET-4 if they study another foreign language as their first foreign language or if their major is a foreign language. More than 99% students chose English as their first foreign language. Those who chose other foreign languages had to pass similar tests.

[9] The term "alphabetically" is defined in the Chinese way.

“Excellence” if the score is above 85. If students fail, they are allowed to repeat the exam, but they can not repeat a Pass to attempt an Excellent. The data are the eventual CET outcomes when students leave the univeristy, and we can only know whether students earn “Pass” or “Excellent” rather than their raw scores. The CET-4 is compulsory in the sense that all the students are required to pass it to be awarded a bachelor degree certificate, so it is not surprising to see that the average CET-4 passing rate is higher than 96% for regular students. Even for Round 3 students, 76% of them passed the CET-4. The CET-6 is voluntary, and passing CET-4 is the prerequisite for taking the CET-6. Given the importance of the CET-6 certificate and the low monetary cost for taking the test, most students took the CET-6 when they were eligible to, and repeated it if they failed. The passing rate of CET-6 is approximately 62% for regular students and 17% for Round 3 students. The “excellent” rates are 19% and 5% for CET-4 and CET-6 respectively for regular students, and are nearly zero for Round 3 students.

Round 3 students are measured in two ways, as shown in Panel B of Table 2.1. “Round 3 Number” measures the total number of Round 3 students in a department-year. It ranges from 4 to 38 in the affected departments in 1996, and zero otherwise. “Round 3 Share” is the percentage of Round 3 students in a department-year, obtained by dividing the number of Round 3 students by the size of a department-year. Given that the size of a department-year ranged from 15 to 131 in 1996, the department-year with the largest number of Round 3 students does not coincide with the department-year with the largest share of Round 3 student.

The College English classes are organized by the Division of College English. An English class usually consists of only students from the same department-year. If the size of a department-year is large, students will be divided into several classrooms. There is no explicit cap for the class size. We could not obtain precise data on historical class size, but we contacted many students in the relevant department-years and obtained the numbers of English classes the department-years were divided into.[10] We estimated class size by dividing the number of students in a department-year with the number of classes. Class size averaged 27, and ranged from 13 to 37. In 1996, as more students entered the university, the class size increased on average in both affected and unaffected departments. After controlling for department and year fixed effects, class size is significantly correlated with the number and the share of Round 3 students at the conventional level (p=0.05 and 0.02 respectively), so we controlled for the effect of class size in the estimation of peer effects.

2.3 Estimation

We estimate the impact of Round 3 students on the test performance of regular students in the same department-year as follows,

$$Y_{idt} = \gamma^{d} + \lambda^{t} + \beta\, \mathrm{Round3}_{dt} + \eta\, \mathrm{Class}_{dt} + \delta\, X_{idt} + e_{idt} \tag{2.1}$$

where the left-hand variable Y_{idt} is the academic outcome of interest (CET 4 Pass, CET 6 Pass, CET4 Excellence, or CET 6 Excellence) of student i at department d in entering year t; $\mathrm{Round3}_{dt}$ is one of the Round 3 indicators; Class_{dt} is the size of the College English class; X_{idt} are the

[10] Although it has been more than ten years, the contacted students can report the number of students in their department-years which roughly agree with the numbers in the dataset, so we believe the numbers of classes they report are correct.

pretreatment characteristics of individual student i at department d of year t, including age, gender, prefecture type, home province, and NCEE scores. Since the NCEE scores are not comparable across provinces, we interacted NCEE English scores with the 26 province dummies to allow for a differential effect of province on NCEE score. For the group of recommended students, who do not have official NCEE scores, their NCEE scores are coded as 0, and we include a dummy variable indicating whether the student is a recommended student. The coefficients γ^d and λ^t capture department and year fixed effects. The addition of department and year effects makes this a difference-in-differences specification in which changes in outcomes before and after 1996 in departments that admitted more Round 3 students are compared to changes in departments that received fewer or no Round 3 students. We report standard errors clustered by departments to account for the possible correlation in the residual e_{idt} in the same department over time. Since all the CET outcomes are binary, and both the CET-4 passing rate and the CET-6 excellence rate are at an extreme, we logistical models for the analysis.[11]

Table 2.2 reports the empirical estimates of the effect of Round 3 students on the CET-4 passing rate of the regular Round 1 students. The dependent variable equals 1 if a student has passed the CET-4 and 0 otherwise. Each coefficient is from a different regression. In other words, each column contains estimates from two separate regressions. The coefficients are in log-odds ratios; we calculate the marginal effects at 0.967, the average CET-4 passing rate of regular students in 1994 and 1995. The marginal effects multiplying with 100 can be interpreted as the changes in percentage points. In column 1, we report estimates controlling for the department and year fixed effects only. A one-student increase in the number of Round 3 students in a department-year reduces the rate of passing CET-4 by .14 percentage points. Similarly, increasing the share of Round 3 students from 0 to 1 reduces the passing rate by 13 percentage points. In column 2, we add class size as a control. The effects of Round 3 indicators remain stable in magnitude, but the addition of class size slightly inflates standard errors of Round 3 indicators. Class size does not seem to affect the CET-4 passing rate in any regressions (not reported in the table). From column 3 to column 5, we gradually add individual characteristics, home province dummies, and province-specific NCEE English scores as additional control variables. All the coefficients remains relatively stable as more control variables are added; in the specification with all the control variables, the coefficients are statistically significant at 0.1 and 0.05 respectively for the number and the share of Round 3 students. By multiplying the per-unit reduction with the average levels of Round 3 indicators in the year 1996, we see that Round 3 students have reduced the CET-4 passing rate of the regular students by 1.5 to 2 percentage points, which is comparably large given that the overall failure rate in passing CET-4 is less than 4 percentage points.

Table 2.3 presents the estimates of the effect of Round 3 students on the CET-6 passing rate of regular students. Unlike CET-4, a pass of which is required for a bachelor certificate, CET-6 is taken voluntarily. The variable "CET-6 Pass" equals 1 if a student takes the exam and passes it and 0 otherwise. The marginal effect is calculated at 0.640, the average CET-6 passing rate of students in 1994 and 1995. A one-student increase in the number of Round 3 students in a

[11] We are aware of the potential incidental parameter problem in the panel data setting, but we do not think it is likely in our settings. Our fixed effect is at the department level, and the smallest department contains 45 students. We check the marginal effects estimated from the logistic model with those from the linear probability model. For the CET-6 passing rate, both the magnitudes and the significance levels are similar; for the CET-4 passing rate, the magnitudes are similar, but the logistic model provides more precise estimates, which confirms our expectation that the linear-probability model tends to give less precise estimates when the probability rate is at the extreme end.

department-year significantly reduces the CET-6 passing rate by about 0.5 percentage points. On average, Round 3 students reduce the CET-6 passing rate in 1996 by more than 5 percentage points. Table 2.3 also suggests the measurement of peer groups is important. Although all the estimates for Round 3 Number are statistically significant, the share of Round 3 students is not significantly correlated with CET-6 passing rates.

Table 2.4 and 5 present the estimation results for the excellence rates of CET-4 and CET-6 respectively. All the estimates are negative but none of them is statistically significant. The overall effects of Round 3 students, calculated by multiplying the marginal effect with the average of Round 3 indicators in 1996, are around one half of the standard deviations of the excellence rates, so we interpret these results as little evidence for the effects of Round 3 students on the excellence rates.

As all the CET outcomes are binary, and students with abilities around the cutoffs tend to be affected, the effects of Round 3 students have a local-effect interpretation, even though "local" cannot be not precisely determined. Given that the CET-4 and CET-6 passing rates are 96.2% and 62.4%, and both rates are significantly affected by Round 3 students, we can generally say that Round 3 students have strong negative impacts on the English test performance of regular students whose English ability is distributed around the bottom 5% to 40%. Nothing restricts us to extend the domain of effects to regular students from the lowest to the bottom 40%, or roughly speaking below average. However, for students with better academic backgrounds in English, we do not find such detrimental effects.

We explore whether there are heterogeneous gender effects along two dimensions. The first is whether male Round 3 students tend to exert more negative impacts than female Round 3 students do. The second is whether male Round 1 students are more affected. Lavy and Schlosser (2007) suggest that boys tend to be more disruptive and violent in Israeli schools, and also boys tend to be more affected by the disruption of other boys. However, we do not find differential effects along either dimension (results are not shown).

2.4 Robustness Checks

The estimates for Round 3 students only indicate causal interpretations in the difference-in-difference model if we can rule out the following two possible biases: selection bias and omitted variable bias. We explore both of this biases below.

2.4.1 Selection Bias

Selection bias is a major econometric obstacle in identifying peer effects. Selection bias may have arisen in our sample if the Round 3 admission process discouraged applications for the relevant academic departments and led to the admission of worse regular students in the affected departments in 1996. But the college admission and application system largely rules out this possibility for the following three reasons.

First, Round 3 admission did not affect the application and admission processes in provinces other than Jiangsu. Students' academic department in a college was decided before they enter the college. Applicants had to specify their preferred academic department when they submit the college application. After that, they are not given any further choice and had to accept the university and department to which the admission system assigns them. Thus, only information available at the application stage could possibly affect students' applications. Every year, each province publishes an official brochure containing all the admission quotas allocated to that province.[12] As the Round 3 admission was held only in Jiangsu province in 1996, such information was not available in the brochures published by other provinces. In addition, news about Round 3 admission was not widely disseminated by the media, and even for students who were already on campus before 1996, most of them were unaware of Round 3 admission until Round 3 students eventually entered the campus. Applicants in other provinces were unlikely to know about Round 3 admission when they submitted their applications. Therefore, Round 3 admission was not likely to affect applications and admissions outside of Jiangsu.[13]

Second, many Round 1 universities normally held associate program admissions after the Round 2 bachelor programs, and Round 3 programs in 1996 were placed in a status similar to that of associate programs in previous years. Round 3 programs were deemed merely as additional programs a university or department tended to have, and therefore the introduction of Round 3 programs was not likely to discourage potential applicants from applying for the bachelor programs.

Third, the "120% rule" in the admission process limited potential selection bias even if potential applicants switched out of the affected departments. The "120% rule" dictates that a university has to use the NCEE total score as the major admission criterion. In particular, it mandates that if a university wants to admit 100 students from a certain province, it can only admit students whose NCEE total scores ranked among the top 120 students of all applicants from that province. Therefore, the variations of students' NCEE scores across academic departments are largely restricted. Then, the university places applicants into academic departments based on students' interests as specified in the application form. Quotas in popular departments are always filled earlier, and less popular departments usually have to admit students who apply for the university but not for their particular department, if those applicants do not refuse. As the strongly affected departments are not among the most popular departments, and they tend to admit students who don't apply for them even in normal years, so Round 3 programs were not likely to have actual effects on admission of affected departments.

To identify whether Round 3 programs have actually affected the characteristics of entering Round 1 students, we conduct the following placebo falsification test,

$$X_{idt} = \gamma^{d} + \lambda^{t} + \beta\ Round3_{dt} + \varepsilon_{idt} \tag{2.2}$$

where X_{idt} represents a background variable of Round 1 student i in department d of the entering year t, and it could be NCEE total score, NCEE English score, or other demographic characteristics. $Round3_{dt}$ indicates Round 3 Number or Share. The coefficients γ^{d} and λ^{t} are the fixed effects for departments and years respectively.

If high-scoring potential applicants would switch from an affected department to another department or university, the NCEE total score and English score would be lower in the affected departments in 1996, since their positions have to be filled by other students with lower scores. Because we are particularly concerned about the selection bias among students from Jiangsu, the first two columns in Table 2.6 focus on regular students from Jiangsu. The admission of Round 3

[12] If a student applied to a university that does not have a quota allocated for the student's province, the application would not be considered by that university. Information on admission quotas includes the name and the round number of a college and how many students each academic department will admit from that province.

[13] The estimated effects of Round 3 students on regular students' CET outcomes are similar regardless ofwhether we use the subsample of students outside of Jiangsu or the full sample of all students, except that the former gives smaller estimates for the effects on the CET-4 passing rate.

students is not significantly associated with the NCEE total scores and English scores for regular students from Jiangsu, and all the estimates are positive. The third and fourth columns repeat the same tests for all the students. The NCEE scores are not comparable across provinces and they can be very volatile for students from provinces which only send a small number of students to the sample university each year, so we interpret these results with caution. The number of Round 3 students is positively related to the NCEE total scores at the 5 percent level, which suggests that the affected departments attracted better students. Nevertheless, the effect is opposite to what selection bias would indicate. We also test whether there are changes in students' demographics, including living area, age, and gender. In addition, we check whether the number of recommended students and the number of students from Jiangsu are significantly correlated with Round 3 indicators; these two numbers are not subject to the choice of applicants, but they may affect the quality of attending Round 1 students. Column 5 to 9 in Table 2.6 show that Round 3 indicators are not significantly related to those characteristics, which suggests that after controlling for the university yearly effect, students in the same department are comparable across years, and there does not seem to be a selection bias along those aspects either.

If students interested in public administration, social welfare, or city planning decide to switch to other universities because of the Round 3 programs, the introduction of Round 3 programs may increase the mismatch between students' academic interest and their academic department, but we don't think this mismatch is not likely to affect our estimates. English is not associated with any specific academic department and is universally required for all college students, so interest in studying English is not likely to change if a student is placed in a less interested department. Furthermore, in regular years, the majority of students admitted to strongly affected departments do not apply for those departments, so there is limited possibility for additional mismatch. One concrete example is that among 23 entering students in public administration in 1995, fewer than 5 students had specified public administration as their department of interest.

2.4.2 Omitted Variable Bias

If other phenomena associated with Round 3 admission also occurred in 1996 but went unobserved by our research, we may wrongly attribute the correlation to the causal peer effects from Round 3 students. Suppose, in 1996, the government announced that it would increasingly recruit staff from college students who studied public administration, social welfare, and city planning. The announcement may first motivate the creation of the Round 3 programs. At the same time, the announcement might also affect students' efforts in the relevant departments. For example, regular students entering in 1996 may think that they could find a job more easily than previous students in the same department might think, and therefore they might put less effort into their studies than the previous students did. In that case, even in absence of Round 3 students, the government announcement would lower academic performance of regular students.

Students were expected to take the CET-4 and the CET-6 by the end of their second and third years respectively. Most students who entered the university in 1994 already passed the CET-4 before the Round 3 admission in 1996. But students entering in 1995 would study for one more year to take the CET-4 after the entry of the Round 3 students. Similarly, for the CET-6, students entering in 1994 studied for one year under the influence of potentially omitted variables if any, while students entering in 1995 were under the two-year influence. For either CET-4 or CET-6, students of 1995 would have experienced the effects from omitted variables, if any, for one more year, compared to students of 1994. Neither 1994 nor 1995 students should be directly affected by the Round 3 students, however, as these students would primarily interact with their 1996 classmates.[14] Therefore, we should be able to detect any possible omitted variable bias with the following equation.

$$Y_{idt} = \gamma^{d} + \lambda^{t} + \beta \text{ pseudo\_Round3}_{dt} + \eta \text{ Class}_{dt} + \delta X_{idt} + \varepsilon_{idt} \quad (2.3)$$

where pseudo_Round3$_{dt}$, is generated by assuming that the same Round 3 students were admitted in 1995 rather than in 1996. For students in 1994, pseudo_Round3$_{dt}$ is 0. The year t is restricted to 1994 or 1995, and students in 1996 are dropped from the analysis.

Table 2.7 shows that Round 3 admissions by departments in 1996 do not predict the test performance of students admitted in 1995 compared to those admitted in 1994 for the same departments. All the results are estimated with the full set of control variables. None of the coefficients is significant at the level of 0.1. For CET-4 and CET-6 passing rates, all the four coefficients are positive, which suggests that the null effects between pseudo_Round3$_{dt}$ and the passing rates are not simply due to a lack of statistical power. Equation (2.3) will be a valid two-side test for the omitted variable bias under the assumption that there are no spillover effects across students of different years in the same department. Since the spillover effects across years and the omitted variable effects point in the same direction and do not cancel out each other, the null effect of pseudo_Round3$_{dt}$ sustains the assumption of null spillover effects and also indicates that there is no omitted variable bias.

5 Conclusions

This study exploits an unusual change in the admission policy in a Chinese university, and examines the impact of specially admitted low-scoring students on the academic performance of their classmates. The arrangements in the sample university facilitate the study of peer effects because the specially admitted students lived and studied together with regular students in the same department-year. The exogenous inflow of specially admitted students, the large differences in academic backgrounds between specially admitted students and regular students, and the strong interactions among these two groups overcomes the well-documented empirical problems associated with identifying peer effects.

Our estimates suggest that specially admitted low-scoring students strongly reduced the College English Test passing rates of regular students in the same department-years. On average, specially admitted students reduced the CET-4 passing rate by 1.5 to 2 percentage points, a sizable proportion given that the overall failure rate of the CET-4 is less than 4%. The specially admitted students also significantly affected whether regular students passed the CET-6 and reduced the passing rate of that test by more than 5 percentage points. However, the presence of specially admitted students does not seem to have significantly impacted the overall attainment in terms of of excellence rates.

To our best knowledge, this is the first study on peer effect consequences of an admission policy change among college students, and it sheds some light to potential outcomes of college

[14] If there are any spillover effects between classes, they should go in the same direction as the omitted variables bias, as we discuss below. A null result it thus evidence against both omitted variables bias and spillover effects between classes.

admission policy changes. In addition, this study analyzes peer effects in a developing country setting, and extends our understandings of peer effects among college students to a larger context.

References

Angrist, Joshua D., and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." American Economic Review, 94(5):1613-1634.

Card, David, and Alan B. Krueger. 2005. "Would the Elimination of Affirmative Action Affect Highly Qualified Minority Applicants? Evidence from California and Texas. Industrial and Labor Relations Review, 58(3):416-434.

Carrell, Scott E., Richard L. Fullerton, and James E. West. 2009. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." Journal of Labor Economics, 27(3):439-464.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2008. "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." National Bureau of Economic Research Working Paper 14475.

Foster, Gigi. 2006. "It's Not Your Peers, and It's Not Your Friends: Some Progress toward Understanding the Educational Peer Effect Mechanism." Journal of Public Economics, 90(8–9): 1455–75.

Imberman, Scott, Adriana D. Kugler, and Bruce Sacerdote. 2009. "Katrina's Children: Evidence on the Structure of Peer Effect from Hurricane Evacuees." National Bureau of Economic Research Working Paper 15291.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." Econometrica, 75(1):83-119.

Lavy, Victor, and Analia Schlosser. 2007. "Mechanisms and Impacts of Gender Peer Effects at School." National Bureau of Economic Research Working Paper 13292.

Lyle, David S. 2007. "Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point." Review of Economics and Statistics, 89(2): 289–99.

Lyle, David S. 2009. "The effects of Peer Group Heterogeneity on the Production of Human Capital at West Point." American Economic Journal: Applied Economics, 1(4):69-84.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." Review of Economic Studies, 60(3): 531–42.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." Quarterly Journal of Economics, 116(2): 681–704.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2006. "What Can Be Learned About Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds." Journal of Public Economics, 90(8–9): 1435–54.

Zimmerman, David J. 2003. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment." Review of Economics and Statistics, 85(1): 9–23.

Figure 2.1 Distribution of NCEE Total Scores

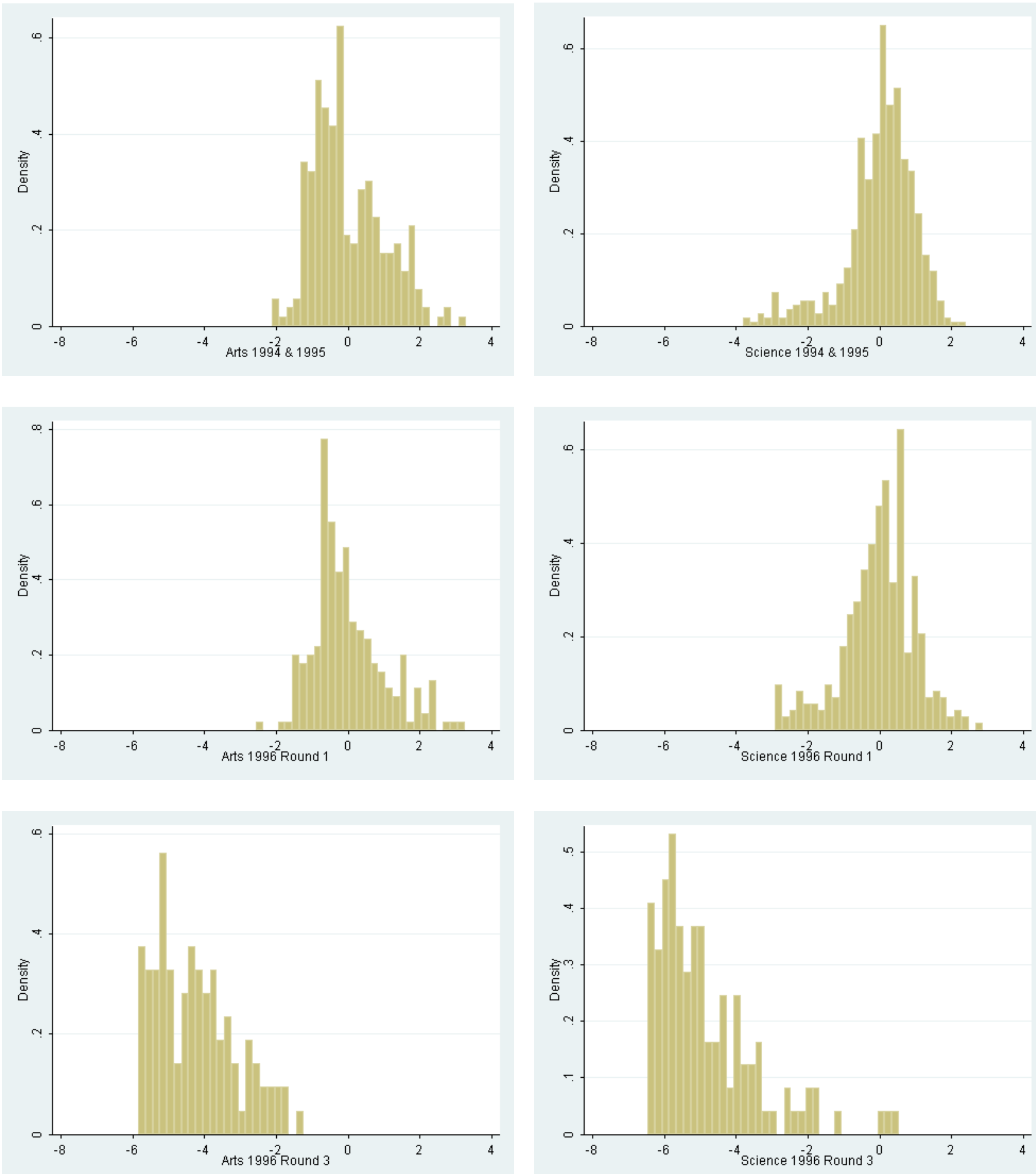


Figure 2.2 Distribution of NCEE English Scores

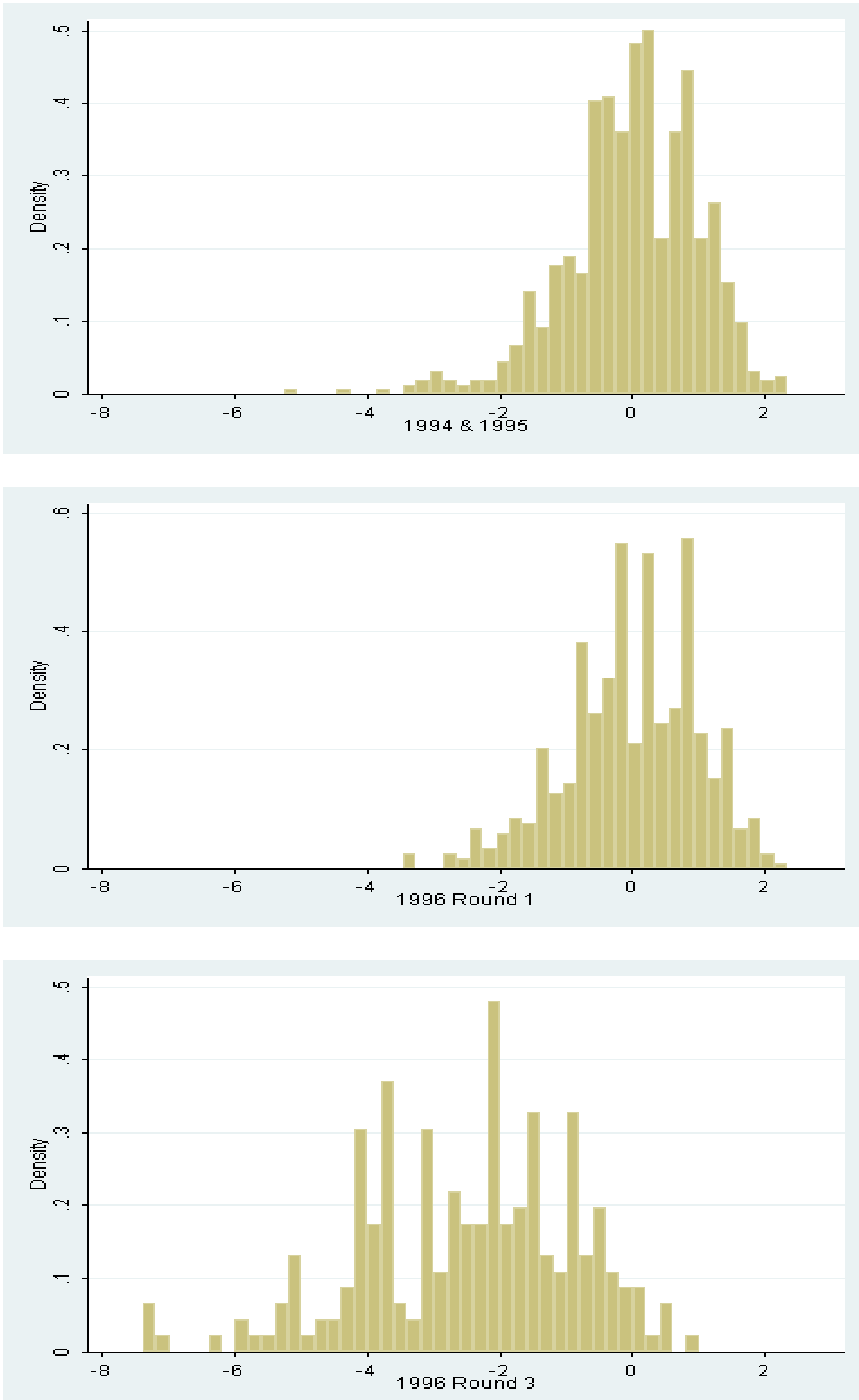


Table 2.1 Summary Statistics

Panel A: CET Performance (Academic Achievements)	Students[1]	Mean	SD	Minimum	Maximum
Round 1 students					
CET-4 Pass	3020	0.962	0.191	0	1
CET-6 Pass	2966	0.624	0.485	0	1
CET-4 Excellence	2689	0.189	0.391	0	1
CET-6 Excellence	2966	0.047	0.211	0	1
Round 3 students					
CET-4 Pass	229	0.760	0.428	0	1
CET-6 Pass	229	0.166	0.373	0	1
CET-4 Excellence	213	0.005	0.069	0	1
CET-6 Excellence	229	0	0	0	0

Panel B: Class Characteristics	Department years	Mean[2]	SD[2]	Minimum	Maximum
Round 3 Number in 1996	23	12.74	11.42	0	38
Round 3 Share in 1996	23	0.151	0.119	0	0.403
Class size in all years	69	26.91	5.501	13	37

Panel C: Baseline Characteristics	Students	Mean	SD	Minimum	Maximum
Round 1 students					
Age entering college	3020	18.45	0.795	15.5	22.75
Being male	3020	0.678	0.467	0	1
From Rural areas	3020	0.279	0.449	0	1
From Jiangsu	3020	0.503	0.500	0	1
Being recommended student	3020	0.572	0.232	0	1

Note:
1. The changing number of students is due to the data availability. The CET-6 outcomes are missing for students in the department of biology-medicine in 1996. The department of computer science does not provide data on the CET-4 excellence if a student passes CET-6, therefore we code CET-4 excellence to be missing for all the students in computer science.
2. They are weighted by the number of regular students in the department-years.

Table 2.2 Effects of Round 3 Students on CET-4 Pass of Regular Students

	Dependent Variable: CET-4 Pass				
Key Predictors	(1)	(2)	(3)	(4)	(5)
Round 3 Number	-0.0435**	-0.0446**	-0.0435*	-0.0410	-0.0438*
	(0.0205)	(0.0223)	(0.0243)	(0.0265)	(0.0254)
Marginal Effect	-0.0014	-0.0014	-0.0014	-0.0013	-0.0014
Round 3 Share	-4.0090***	-4.1790**	-4.2993**	-4.1179**	-4.6274**
	(1.5319)	(1.6511)	(1.9075)	(2.0429)	(2.0589)
Marginal Effect	-0.1279	-0.1334	-0.1372	-0.1314	-0.1477
Observations	2540	2540	2540	2528	2528
Control Variables					
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Class size	-	Yes	Yes	Yes	Yes
Individual characteristics	-	-	Yes	Yes	Yes
Province Fixed Effects	-	-	-	Yes	Yes
NCEE English scores	-	-	-	-	Yes

Notes: Each estimate is from a different logistic regression. The coefficients are in log odds ratios. The dependent variable is whether a regular Round 1 student passed the CET-4, which is equal to 1 if pass and 0 otherwise. The unit of observation is a student entering the sample university in the years 1994, 1995 and 1996. Individual characteristics include age, age squared, gender and living in rural or unban areas. NCEE English scores are interacted with home province dummies to allow differential effects of NCEE English scores from provinces on CET performances. Standard errors clustered by department in parenthesis. Marginal effects are calculated at p=0.967, the average CET-4 passing rate of students in 1994 and 1995.

* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

Table 2.3 Effects of Round 3 Students on CET-6 Pass of Regular Students

	Dependent Variable: CET-6 Pass				
Key Predictors	(1)	(2)	(3)	(4)	(5)
Round 3 Number	-0.0223***	-0.0250***	-0.0255***	-0.0242**	-0.0238**
	(0.0074)	(0.0093)	(0.0097)	(0.0103)	(0.0105)
Marginal Effect	-0.0051	-0.0058	-0.0059	-0.0056	-0.0055
Round 3 Share	-0.9867	-1.1392	-1.1316	-1.0148	-0.9326
	(0.7030)	(0.7608)	(0.9471)	(0.9583)	(0.8387)
Marginal Effect	-0.2273	-0.2625	-0.2607	-0.2338	-0.2149
Observations	2966	2966	2966	2966	2966
Control Variables					
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Class size	-	Yes	Yes	Yes	Yes
Individual characteristics	-	-	Yes	Yes	Yes
Province Fixed Effects	-	-	-	Yes	Yes
NCEE English scores	-	-	-	-	Yes

Notes: Each estimate is from a different logistic regression. The coefficients are in log odds ratios. The dependent variable is whether a regular Round 1 student passed the CET-6, which is equal to 1 if pass and 0 otherwise. The unit of observation is a student entering the sample university in the years 1994, 1995 and 1996. Individual characteristics include age, age squared, gender and living in rural or unban areas. NCEE English scores are interacted with home province dummies to allow differential effects of NCEE English scores from provinces on CET performances. Standard errors clustered by department in parenthesis. Marginal effects are calculated at p=0.640, the average CET-6 passing rate of students in 1994 and 1995.
* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

Table 2.4 Effects of Round 3 Students on CET-4 Excellence of Regular Students

	Dependent Variable: CET-4 Excellence				
Key Predictors	(1)	(2)	(3)	(4)	(5)
Round 3 Number	-0.0093	-0.0139	-0.0128	-0.0129	-0.0158
	(0.0080)	(0.0097)	(0.0100)	(0.0109)	(0.0119)
Marginal Effect	-0.0015	-0.0022	-0.0020	-0.0021	-0.0025
Round 3 Share	-0.7974	-1.2957	-1.1671	-1.1984	-1.5089
	(0.7661)	(0.8333)	(0.9247)	(0.9758)	(0.9978)
Marginal Effect	-0.1273	-0.2069	-0.1864	-0.1914	-0.2410
Observations	2689	2689	2689	2665	2665
Control Variables					
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Class size	-	Yes	Yes	Yes	Yes
Individual characteristics	-	-	Yes	Yes	Yes
Province Fixed Effects	-	-	-	Yes	Yes
NCEE English scores	-	-	-	-	Yes

Notes: Each estimate is from a different logistic regression. The coefficients are in log odds ratios. The dependent variable is whether a regular Round 1 student obtained an Excellence in the CET-4, which is equal to 1 if pass and 0 otherwise. The unit of observation is a student entering the sample university in the years 1994, 1995 and 1996. Individual characteristics include age, age squared, gender and living in rural or unban areas. NCEE English scores are interacted with home province dummies to allow differential effects of NCEE English scores from provinces on CET performances. Standard errors clustered by department in parenthesis. Marginal effects are calculated at p=0.200, the average CET-4 excellence rate of students in 1994 and 1995.

* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

Table 2.5 Effects of Round 3 Students on CET-6 Excellence of Regular Students

	Dependent Variable: CET-6 Excellence				
Key Predictors	(1)	(2)	(3)	(4)	(5)
Round 3 Number	-0.0166	-0.0163	-0.0161	-0.0154	-0.0211
	(0.0179)	(0.0180)	(0.0201)	(0.0203)	(0.0231)
Marginal Effect	-0.0009	-0.0009	-0.0009	-0.0009	-0.0012
Round 3 Share	-1.0528	-0.9954	-1.0158	-0.9478	-1.8736
	(1.0389)	(1.3505)	(1.4898)	(1.5269)	(1.9538)
Marginal Effect	-0.0594	-0.0561	-0.0573	-0.0535	-0.1057
Observations	2679	2679	2679	2420	2420
Control Variables					
Department Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Class size	-	Yes	Yes	Yes	Yes
Individual characteristics	-	-	Yes	Yes	Yes
Province Fixed Effects	-	-	-	Yes	Yes
NCEE English scores	-	-	-	-	Yes

Notes: Each estimate is from a different logistic regression. The coefficients are in log odds ratios. The dependent variable is whether a regular Round 1 student obtained an Excellence in the CET-6, which is equal to 1 if pass and 0 otherwise. The unit of observation is a student entering the sample university in the years 1994, 1995 and 1996. Individual characteristics include age, age squared, gender and living in rural or unban areas. NCEE English scores are interacted with home province dummies to allow differential effects of NCEE English scores from provinces on CET performances. Standard errors clustered by department in parenthesis. Marginal effects are calculated at p=0.060, the average CET-6 excellence rate of students in 1994 and 1995.

* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

Table 2.6 Testing for Exogeneity of Round 3 Students

	Pretreatment Variables								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Key Predictors	NCET Total Score Jiangsu	NCET English score Jiangsu	NCET Total score All	NCET English score All	Rural Area	Entering Age	Being Male	From Jiangsu	Recommend students
Round 3 Number	0.0052	0.0017	0.0061**	-0.0035	-0.0010	0.0009	0.0008	0.0019	-0.0004
	(0.0048)	(0.0043)	(0.0028)	(0.0026)	(0.0012)	(0.0011)	(0.0025)	(0.0012)	(0.0006)
Round 3 Share	0.7232	0.3286	0.5156	-0.2244	-0.0894	0.0978	-0.0954	0.1832	-0.0392
	(0.5027)	(0.4472)	(0.4077)	(0.2888)	(0.1258)	(0.1128)	(0.2757)	(0.1407)	(0.0579)
Observations	1409	1409	2847	2847	3020	3020	3020	3020	3020
Controlled for:									
Department Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each estimate is from a different linear regression. The unit of observation is a student entering the sample university in the years 1994, 1995 and 1996. Standard errors clustered by department in parenthesis.
* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

Table 2.7 Placebo Testing

	Dependent Variables			
Key Predictors	CET-4 Pass	CET-6 Pass	CET-4 Excellence	CET-6 Excellence
Round 3 Number	0.0351	0.0057	-0.0070	0.0154
	(0.0377)	(0.0098)	(0.0100)	(0.0231)
Round 3 Share	5.5657	1.3942	-1.1653	1.0935
	(3.5670)	(1.1969)	(1.3002)	(2.1866)
Observations	1283	1847	1650	1448
Control Variables				
Department Effects	Yes	Yes	Yes	Yes
Year Effects	Yes	Yes	Yes	Yes
Class size	Yes	Yes	Yes	Yes
Individual Characteristics	Yes	Yes	Yes	Yes
Home province	Yes	Yes	Yes	Yes
NCEE English scores	Yes	Yes	Yes	Yes

Notes: Each estimate is from a different logistic regression. The coefficients are in log odds ratios. The unit of observation is a student entering the sample university in the years 1994 and 1995 but not in 1996. Individual characteristics include age, age squared, gender and living in rural or unban areas. NCEE English scores are interacted with home province dummies to allow differential effects of NCEE English scores from provinces on CET performances. Standard errors clustered by department in parenthesis.
* significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent

Chapter 3. Micro-groups, Peer Effects, and Classroom Organization: Evidence from a Chinese Field Experiment

Social interactions are generally believed to play an important role in students' academic achievement. Most studies on peer effects in primary and secondary education define peers at the classroom or school level and assume that students are influenced by classroom or school-level averages. However, recent work by Carrell, Sacerdote and West (2011) finds that students may form subgroups within classrooms, implying that peer effect analyses at the classroom or school level may overlook important interactions within sub-classroom groups.

This paper examines peer effects among subgroups of Grade 7 students by exploiting an experiment with random seat assignment in a Chinese middle school. As in many Chinese schools, students in this school stay at a fixed seat in a fixed classroom for most classes, while teachers rotate through the classrooms. Students are typically assigned to seats loosely according to height, so that shorter students occupy front rows. In this experiment, however, students were assigned to blocks of rows based on height and then randomly assigned to seats within blocks. This within-block randomization controls for non-random sorting of students into groups and allows an exploration of peer effects in a micro-environment.

We find that the gender of nearby students influences a student's performance, but the effects vary according to the student's gender. Having a female deskmate (a student who shares the same desk) increases a student's test scores by 0.05 to 0.08 standard deviations regardless of the student's gender. Being surrounded by five females rather than five males increases a female student's test scores by 0.3 standard deviations but has no effect on a male student's test scores. In contrast, the academic performance of surrounding students displays little relationship with a student's test scores. The differential effects of neighboring students across genders, however, suggest significant opportunities for improving classroom arrangements.

In addition to estimating the effects of nearby students on academic performance, we also explore how students evaluate the influence of deskmates on study habits. Students report that deskmates with higher baseline test scores positively influence their own study habits. Likewise, students report that they positively influence the study habits of their deskmates if they themselves have higher baseline scores. However, the same measures are not significantly affected by one's own gender or the deskmate's gender. The discrepancy between the self-reported impacts of peer test scores and gender and the actual impacts of peer test scores and gender highlights the risks of using subjective evaluations as proxies for actual outcomes.

The next section reviews the existing literatures. Section 3.2 describes school environment. Section 3.3 presents the experimental. Section 3.4 describes the data and verifies the randomness of seat assignment, and Section 3.5 presents the main results. Section 1.6 summarizes the paper and draws conclusions.

3.1 Literature Review

A rich set of empirical studies on peer effects leverage variation in peer groups at the classroom or school level (Hanushek et al., 2003; Angrist and Lang, 2004; Arcidiacono and Nicholson, 2005; Hoxby and Weingarth, 2005; Ammermueller and Pischke, 2009; Gould, Lavy and Paserman, 2009; Wang, 2010). Others explore living arrangements among college students (Sacerdote, 2001; Zimmerman, 2003). To the best of our knowledge, however, no studies leverage experimental or quasi-experimental variation to estimate peer effects within sub-classroom groups.

A parallel literature explores the specific effects of student gender on peer outcomes. The US Department of Education (2005) and Morse (1998) review studies comparing students in single-sex and coeducational classes; some studies suggest the single-sex schooling may be beneficial while others indicate no difference. Hoxby (2000) and Lavy and Schlosser (2011) explore plausibly exogenous variations in the gender composition in coeducational schools and find that the proportion of female students has positive effects on students' cognitive achievements. However, gender composition it does not have differential effects on boys and girls. Whitmore (2005) finds that students assigned to classrooms with higher proportions female do better in kindergarten and second grade, with some evidence of differential effects on boys and girls.

Our study extends the academic peer effects literature to sub-classroom groups. The results reveal that even within micro-level environments, there can be strong peer effects. This finding has policy relevance because teachers have significant discretion in organizing groups within classrooms. Implementing single-sex groups within classrooms, for example, seems less controversial than implementing single-sex classrooms or single-sex schools. Changes to classroom arrangements thus represent a low-cost way to potentially improve academic performance.

3.2 School Environment

This experiment was implemented in a coeducational public middle school in Jiangsu, China. Students involved in the study were in Grade 7, the starting grade at the middle school. At the beginning of the school year, students were assigned to a fixed classroom. They stayed in the same classroom for most classes over the semester, while teachers rotated from classroom to classroom. This arrangement is standard for this middle school and many other schools in China.

Desks and benches are provided in classrooms, typically arranged as in Figure 3.1. Each desk seats two students, and there are four desks per row with aisles between desks. There are six, seven or eight rows in each classroom depending on the number of students. All students are assigned to fixed seats, and they are required to stay in the assigned seat during class time. The practice of assigning students to fixed seats helps teachers catch absentees and misbehaving students during class.

An administrative teacher assigns seats for the classroom.[1] When assigning seats, student height is the major consideration. Classrooms are typically crowded, and taller students sitting in the front may block the view of shorter students behind them. In a non-experimental setting, the

[1] An administrative teacher is a regular teacher with additional managerial responsibilities, which include arranging class events, disciplining misbehavior, communicating with parents, and assigning student seats.

administrative teacher may have personal preferences for assigning seats. For example, some administrative teachers like to put students of the same gender together while others tend to mix genders. Seats may also be dynamically adjusted during the school year; as administrative teachers learn more about students, they may rearrange seats. In addition, some parents may request to have their children moved to the front of the classroom or near high performing students.

A typical day consists of a 30-minute reading session in the early morning, four 45-minute lecture sessions in the late morning, three 45-minute lecture or self study sessions in the afternoon, and one 40-minute self-study or physical exercise session in the late afternoon. During most sessions, students must stay in their own seats. In lectures, chatting is generally prohibited. During self-study sessions, students choose what to study for themselves. Usually students are not prohibited from talking in low voice with neighboring students during self-study sessions.

Neighboring students have many opportunities to interact with each other, and different peer groups may influence students in different ways. For example, students can talk with their deskmates without moving at all, but they generally have to turn their bodies to talk with students at adjacent desks. Deskmates can always observe each other with ease, but it is difficult to observe details across rows. Though students may interact with other students across aisles, the columns of desks in this experiment rotated every several weeks. Students thus had few opportunities to form lasting peer groups across aisles.

Since each column of desks remained fixed during this experiment, we define peer groups within columns of desks. The first peer group is the desk – each student has a single deskmate. The second peer group consists of “neighbor-4 students.” A student’s neighbor-4 peers are the two students sitting at the desk directly in front of her and the two students sitting at the desk directly behind her. The last peer group, “neighbor-5 students,” consolidates the first two groups. A student’s neighbor-5 peers are her neighbor-4 peers plus her deskmate.

To see a concrete example of these peer groups, consider Student 1 in the second row and second column of Figure 3.1. Student 2 is his deskmate, Students 3 through 6 are his neighbor-4 peers, and Students 2 through 6 are his neighbor-5 peers. For Student 3, as no students sit in front of her, her neighbor-4 peers and neighbor-5 peers include only two and three students respectively.

3.3 Experimental Design

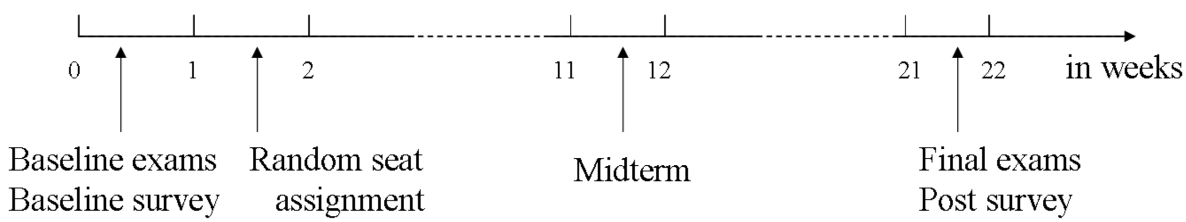
In this experiment, a research group in the Jiangsu Department of Education randomly assigned students’ seats. During the first week of the Fall 2009 semester, the Department of Education requested information on students’ names, gender, and heights in each classroom. The basic mechanism for assigning seats is as follows. First, students were sorted from shortest to tallest by gender within each classroom. Then, the first sixteen students were placed in Block 1 (corresponding to Row 1 and 2), the next 16 students were placed in Block 2 (Rows 3 and 4), and so on until all students were assigned to blocks. Students taller than 5 feet, 6.5 inches (equivalently 169 centimeters) were put in a separate block. Finally, a random sequence was generated, and students were randomly permuted and assigned to seats within each block. The size of the last two blocks varies depending on the number of students and distribution students’ height within classrooms. Students in shorter groups always sit in front of students in taller groups, but within a block taller students may sit in front of shorter students. This did not present challenges in the classroom as all students within the same block are of roughly similar height. Due to the one-child family planning policy and a frequent preference for sons, the ratio of boys to girls was 1.27 in the sample school. As boys and girls were of similar height in Grade 7, [2] we have nine boys and seven girls (9/7=1.28) in a block until it becomes infeasible.

Some students required favorable seat assignments due to near-sightedness, and in some cases parents lobbied for favorable seat assignments. To increase compliance rates, the researchers allowed administrative teachers to list several student names for favorable seat treatments; students in the favored list accounted for 9% of all students. Students on this list received a seat assignment in either a front row or a middle column.[3]

Administrative teachers were asked to cooperate by adopting the random seat assignments and avoiding seat adjustments over the semester. During the semester, however, administrative teachers might adjust seats. There were no financial incentives for administrative teachers to cooperate in this experiment, and students were not informed of the research project.

3.4 Data and Validation of Random Seat Assignment

The data for this study consist of three rounds of test scores and two rounds of surveys for students of Grade 7 in 2009. We illustrate the timeline for collecting the data as follows.



The baseline test and baseline survey were administrated during the first week of the semester before random seat assignment. The random seat assignment was announced during the second week. Students sat according to the random assignment unless the administrative teachers made adjustments. For the midterm and final exams, due to the school's efforts to prevent cheating, students were seated such that students in the same classroom were generally spread over more than ten rooms and no student sat immediately adjacent to another student from the same grade. Two teachers monitored the exams in each classroom. The post survey was administrated right after the final exam when students were still seated according the seat arrangement for the final exam. As students took exams and surveys in different seats than their experimental assignments, any correlations in outcomes among randomly assigned peers are not likely generated by communication among students when taking the exams or surveys.

Grading was rigorously conducted. Teachers in the same subjects allocated exam questions among themselves so that the same teacher always graded the same question. In addition, students' names were hidden during the grading process. In the baseline test, the school tested students on three major subjects – Chinese, English and math. In the midterm and final, the

[2] Boys were 0.4 inches taller than girls on average.

[3] For students on this list, if they were originally assigned to the first four rows, they were moved to the middle columns in the same row. If they were originally in Row 5, they were moved to the middle columns in Row 4. If they were originally behind Row 5, they were moved to Row 5.

school tested seven subjects – Chinese, English, math, politics, history, geography, and biology. Each of the three major subjects accounted for 150 points in the raw scores, and the other four subjects accounted for approximately 50 points each. The "Core score" represents the sum of the scores for the three major subjects, and the "Total score" represents the sum of all seven scores. All test scores are standardized to have mean 0 and standard deviation 1. Figure 3.2 presents the kernel density of students' baseline total scores by gender. The test scores are left skewed for both boys and girls, and girls had higher scores overall. Figure 3.3 illustrates the kernel density of the Core scores and Total scores for the midterm and final exams.

Besides the administrative data on students' gender, height and test scores, the surveys also provide rich information on students' family backgrounds, subjective interests, and aspirations, both before and after the random seat assignment. The surveys also report students' evaluations of peer influences. Panel A of Table 3.1 presents baseline summary statistics, while Panel B presents post-experiment summary statistics.

We use three types of peer groups in this study – deskmates, neighbor-4 peers, and neighbor-5 peers. For each type of peer group, we construct two measures of the peer characteristics, gender composition (whether the deskmate is female or the proportion of females among neighbor-4 and neighbor-5 peers) and baseline total test score. Panel C of Table 3.1 presents summary statistics of peer characteristics.

Random seat assignments allow us to identify the casual effects of neighboring students. The traditional method for verifying random assignment is to regress self baseline characteristics on peer characteristics. However, since sampling is done without replacement, a simple bivariate regression may generate mechanical negative correlations – if student *i* has a high share of female peers, for example, then it is more likely that student *i* is male. Wang (2009) proposes two solutions for this issue. One is to control for each student's own value of the corresponding characteristic in the regression. For example, when testing whether deskmate gender is correlated with a student's academic background, controlling for the student's own gender in the regression eliminates the mechanic correlation. However, this approach is not feasible when testing whether a student's gender is correlated with deskmate gender. An alternative solution is to jointly test whether the baseline characteristics are different across peer groups. To do so, we compare two models. The constrained model contains only the block fixed effects and the status in the favor list, while the full model includes additional peer group fixed effects. By testing whether the peer group fixed effects are jointly significant, we can test whether there is non-random sorting in to peer groups.

Panel A of Table 3.2 presents results from regressions controlling for block fixed effects, status in the favor list, and self baseline characteristics whenever possible. For the estimates in italics, we are unable to control for self baseline characteristics. Regressions of gender on peer gender and baseline test scores on peer baseline test scores display weak evidence of negative mechanical correlation. There is no evidence of any correlation between gender and peer baseline test scores or baseline test scores and peer gender.

Panel B presents results from testing for non-random sorting in to peer groups. The *F*-statistics test whether the peer group fixed effects explain a significant amount of the variation in gender or baseline test scores. There is no evidence of any non-random sorting into peer groups.

Panel C of Table 3.2 tests for correlations between peer characteristics and other self characteristics. The results are consistent with null effects for all tests. Of the 72 tests in Panel C, five tests are weakly significant at 10 percent level, and one test is weakly significant at the 5 percent level (p = 0.03 for this test). This rejection rate is exactly what we would expect under the null hypothesis that all correlations are zero. In summary, there is no evidence of non-random sorting into peer groups. Nevertheless, since one characteristic, ideal education level, appears to be more imbalanced than most, we include it as a covariate in our regressions to ensure that the imbalance does not bias our results.

3.5 Results

3.5.1 Main Effects of Peers on Academic Performance

Given the within block randomization of seats, we estimate the main effects of peers using the following equation

$$Y_{cbi} = Peer_{cbi} + X_{cbi} + \lambda^{cb} + e_{cbi} \qquad (3.1)$$

The outcome Y_{cbi} represents standardized test scores for class c, block b, and student i. The regressor of interest, $Peer_{cbi}$, represents the gender or baseline test score of student i's deskmate or neighbor-4 students. The term X_{cbi} includes student i's own characteristics – gender, baseline test score, being on the favored list, and ideal education level.[4] The term λ^{cb} contains block fixed effects; these fixed effects are important for identification since randomization occurs within blocks. Table 3.3 presents results from estimating equation 3.1. Overall, a female deskmate increases a student's test scores by 0.05 to 0.08 standard deviations. The effects for the total midterm score are statistically significant, but the effects for the total final score are only marginally significant. A deskmate's baseline test score, however, does not seem to affect a student's midterm or final score. The inclusion of neighbor-4 gender or baseline test scores does not affect the coefficients on deskmate gender or baseline test scores, which further confirms that deskmates and neighbor-4 students are uncorrelated under the random seat assignment. The coefficient for the proportion of females among neighbor-4 students is positive but statistically insignificant. The academic backgrounds of neighbor-4 students are positively related to a student's midterm test scores, but the coefficients are only marginally significant.

3.5.2 Heterogeneous Treatment Effects

Heterogeneous treatment effects may provide opportunities for improving welfare by rearranging students. Columns (1) through (4) in Table 3.4 present regressions in which the effects of peer gender composition are estimated separately for males and females. The coefficients on the interaction between being female and having a female deskmate are small and statistically insignificant; female deskmates do not appear to affect male and female students differently. Although the gender composition of neighbor-4 peers does not have an overall effect on students (see Table 3.3), the coefficients on the interaction between being female and the proportion females among neighbor-4 students imply that female students benefit significantly from sitting near other females. The main effect on gender composition of neighbor-4 peers, however, is negative and statistically insignificant, suggesting that males do not benefit from sitting near females.

[4] We include the ideal education level because it is appears correlated with some of the key regressors.

Columns (5) through (8) in Table 3.4 examine whether peer gender composition affects students with different academic backgrounds differently. The results for total midterm and final scores suggest that students with better academic backgrounds benefit less from having female deskmates. If a student's baseline score is one standard deviation above the average, she does not benefit from having a female deskmate – the main effect and interaction effect cancel out. The interaction effects between academic background and gender composition of neighbor-4 peers, however, are small and statistically insignificant.

To summarize the effects of deskmates and neighbor-4 students, Table 3.5 estimates the effects of neighbor-5 students for each gender (recall that a student's neighbor-5 peers are her deskmate plus her neighbor-4 peers). For every outcome, female students have positive effects on neighboring female students but no effects on neighboring male students. Moving a female student from an all-boy microenvironment to an all-girl microenvironment, for example, increases her test scores by approximately 0.3 standard deviations. Male students, in contrast, display no similar response to gender peer composition.

3.5.3 Effects of peers on self-reported peer relationships

The survey questionnaire includes four specific questions on the relationship between the surveyed student and his deskmate: (1) what is the effect of the deskmate on the surveyed student's study habits; (2) what is the effect of surveyed student on the deskmate's study habit; (3) how well do the surveyed student and the deskmate get along with each other; (4) how strongly does the surveyed student wish to change seats.

Table 3.6 presents the effects of peer gender and academic background on survey responses.[5] The table also presents coefficients for student *i*'s gender and baseline test scores for the purpose of comparing Columns 1 and 2. Column 1 presents the effects of deskmate gender and academic background on the surveyed student's study habits. A higher scoring deskmate has a positive and statistically significant effect on the surveyed student's study habits, but a female deskmate has no statistically significant effect on the surveyed student's study habits. Column 2 demonstrates that when the surveyed student evaluates his own effect of on his deskmate's study habits, the surveyed student believes that he is more helpful if he has a better test score. This result is consistent with the results in Column 1, which show that students believe their deskmates' positive influence is increasing in deskmate test scores. However, they are inconsistent with the actual effects of high scoring deskmates, which are small and statistically insignificant (see Table 3.3).

It is commonly presumed in Chinese culture that good students have positive impacts on others. The discrepancy between the perceived effects of high scoring peers and the actual effects of high scoring peers thus suggests that the survey responses may be biased towards common beliefs. Our results therefore cast doubt on the use of subjective evaluations as proxies for actual outcomes.

[5] Although the survey responses are categorical variables, we choose to present results from linear regression models rather than ordered logit models. This is because the within block randomization requires us to control for block fixed effects, and the ordered logit model may suffer from the incidental parameter problem with a large number of fixed effects. Nevertheless, the coefficient signs and significance levels are similar if we instead estimate an ordered logit model.

Columns (3) and (4) examine whether deskmate gender or academic background affect deskmate relationships. Peer characteristics have no significant effects on either the surveyed student's relationship with her deskmate or her desire to change seats. Students with better academic backgrounds, however, report better relationships with their deskmates.[6]

3.6 Conclusion

We identify peer effects within subgroups inside classrooms by exploiting the random assignment of seats in a Chinese middle school. The results suggest that having a female deskmate is beneficial for both boys and girls, while having more female neighbors has strong positive effects on girls but no impacts on boys. The differing patterns between the deskmate results and the neighbor-4 results may be due to differences in interactions – interactions between deskmates are easier and to some degree unavoidable, while interactions with neighboring students are subject to self selection.

Our results demonstrate clear opportunities for improving classroom arrangements. In our sample, placing females in small groups with other females increases their test scores by 0.3 standard deviations. Placing males in small groups with other males has no negative impact on their test scores.[7] Overall, students placed in small groups of homogeneous gender within a classroom – a small-scale version of single-sex education – can significantly outperform students in groups of mixed genders. This finding suggests a low cost way to improve test scores within the world's largest primary education system.

Our analysis of the survey responses indicates that students believe that sitting near students with high test scores will improve their own study habits. The gender of neighboring students, in contrast, is not perceived to have any influence on study habits. The actual results, however, suggest that both beliefs are wrong – neighbors' genders matter, while neighbors' baseline test scores do not. This finding suggests that subjective evaluations may be biased towards common beliefs and highlights the risks of using subjective evaluations as proxies for actual outcomes.

[6] We do not view this relationship as having a causal interpretation; it is simply a descriptive fact.

[7] Female deskmates have a modest positive effect on male scores, but female neighbors have a modest negative effect on male scores. These two effects cancel out in the "neighbor-5" regressions in Table 3.5.

References

Ammermueller, Andreas, and Jorn-Steffen Pischke. 2009. "Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study." *Journal of Labor Economics*, 27(3): 315–48.

Angrist, Joshua D., and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review*, 94(5): 1613–34.

Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics*, 114(2): 533–75.

Arcidiacono, Peter, and Sean Nicholson. 2005. "Peer Effects in Medical School." *Journal of Public Economics*, 89(2–3): 327–50.

Black, Sandra E. 1999. "Do Better Schools Matter? Parental Valuation of Elementary Education." *Quarterly Journal of Economics*, 114(2): 577–99.

Boozer, Michael A., and Stephen E. Cacciola. 2001. "Inside the 'Black Box' of Project Star: Estimation of Peer Effects Using Experimental Data." Yale University Economic Growth Center Discussion Paper 832.

Campbell, Patricia B., and Jo Sanders. 2002. "Challenging the System: Assumptions and Data behind the Push for Single-Sex Schooling." In *Gender In Policy and Practice: Perspectives on Single-Sex and Co-educational Schooling*, ed. Amanda Datnow and Lea Hubbard, 31–46. New York: RoutledgeFalmer.

Carrell, Scott E., Richard L. Fullerton, and James E. West. 2009. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." Journal of Labor Economics, 27(3):439-464.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2011. From Natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects. NBER Working Paper 16865.

Cullen, Julie Berry, Brian A. Jacob, and Steven Levitt. 2006. "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." *Econometrica*, 74(5): 1191–1230.

Figlio, David N. 2007. "Boys Named Sue: Disruptive Children and Their Peers." *Education Finance and Policy*, 2(4): 376–94.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2008. "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." National Bureau of Economic Research Working Paper 14475.

Foster, Gigi. 2006. "It's Not Your Peers, and It's Not Your Friends: Some Progress toward Understanding the Educational Peer Effect Mechanism." Journal of Public Economics, 90(8–9): 1455–75.

Gould, Eric D., Victor Lavy, and M. Daniele Paserman. 2009. "Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence." *Economic Journal*, 119(540): 1243–69.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2004a. "Disruption versus Tiebout Improvement: The Costs and Benefits of Switching Schools." *Journal of Public Economic*s, 88(9–10): 1721–46.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2004b. "Why Public Schools Lose Teachers." *Journal of Human Resources*, 39(2): 326–54.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin. 2003. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics*, 18(5): 527–44.

Harwath, Irene, Mindi Maline, and Elizabeth DeBra. 1997. *Women's Colleges in the United States: History, Issues and Challenges.* Washington, DC: U.S. Government Printing Office.

Hoxby, Caroline. 2000. "Peer Effects in the Classroom: Learning from Gender and Race Variation." National Bureau of Economic Research Working Paper 7867.

Imberman, Scott, Adriana D. Kugler, and Bruce Sacerdote. 2009. "Katrina's Children: Evidence on the Structure of Peer Effect from Hurricane Evacuees." National Bureau of Economic Research Working Paper 15291.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1): 83–119.

Krueger, Alan B. 1999. "Experimental Estimates of Education Production Functions." *Quarterly Journal of Economics*, 114(2): 497–532.

Lavy, Victor, and Anal ́a Schlosser. 2005. "Targeted Remedial Education for Underperforming Teenagers: Costs and Benefits." *Journal of Labor Economics*, 23(4): 839–74.

Lazear, Edward P. 2001. "Educational Production." *Quarterly Journal of Economics*, 116(3): 777–803.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531–42.

Morse, Susan. 1998. *Separated by Sex: A Critical Look at Single-Sex Education for Girls*. Washington,DC: American Association of University Women Educational Foundation.

Sacerdote, Bruce. 2001. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *Quarterly Journal of Economics*, 116(2): 681–704.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2006. "What Can Be Learned About Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds." Journal of Public Economics, 90(8–9): 1435–54.U.S.

Department of Education. 2005. Single-Sex Versus Coeducational Schooling: A Systematic Review. Office of Planning, Evaluation and Policy Development Policy and Program Studies Service. Washington, DC, September.U.S.

General Accounting Office. 1994. *Elementary School Children: Many Change Schools Frequently, Harming Their Education*. Health, Education, and Human Services Division. Washington, DC, February.

Whitmore, Diane. 2005. "Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment." *American Economic Review*, 95(2): 199–203.

Zimmerman, David J. 2003. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment." *Review of Economics and Statistics*, 85(1): 9–23.

Figure 3.1 Arrangement of a classroom

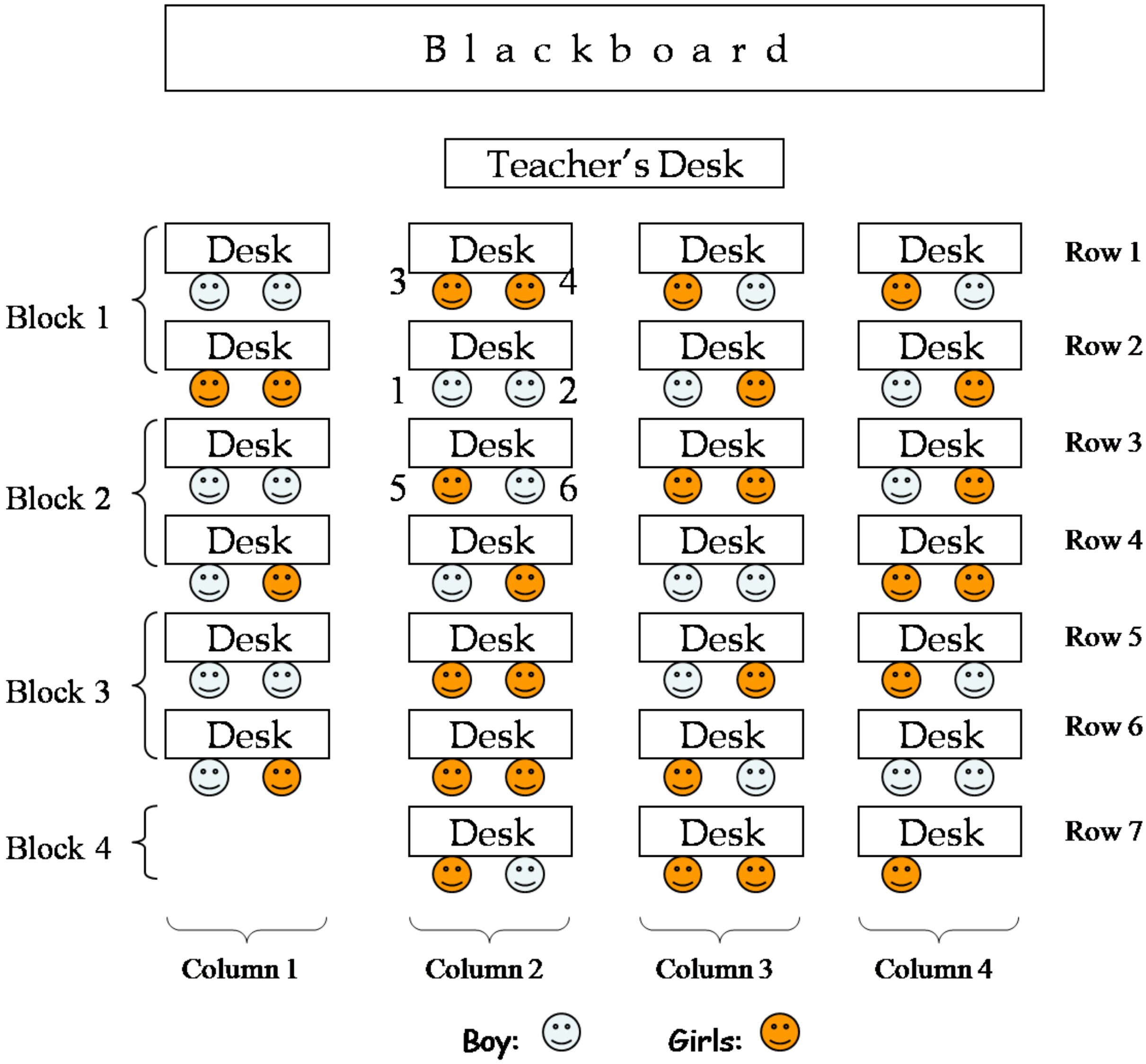


Figure 3.2 Distribution of baseline scores by genders

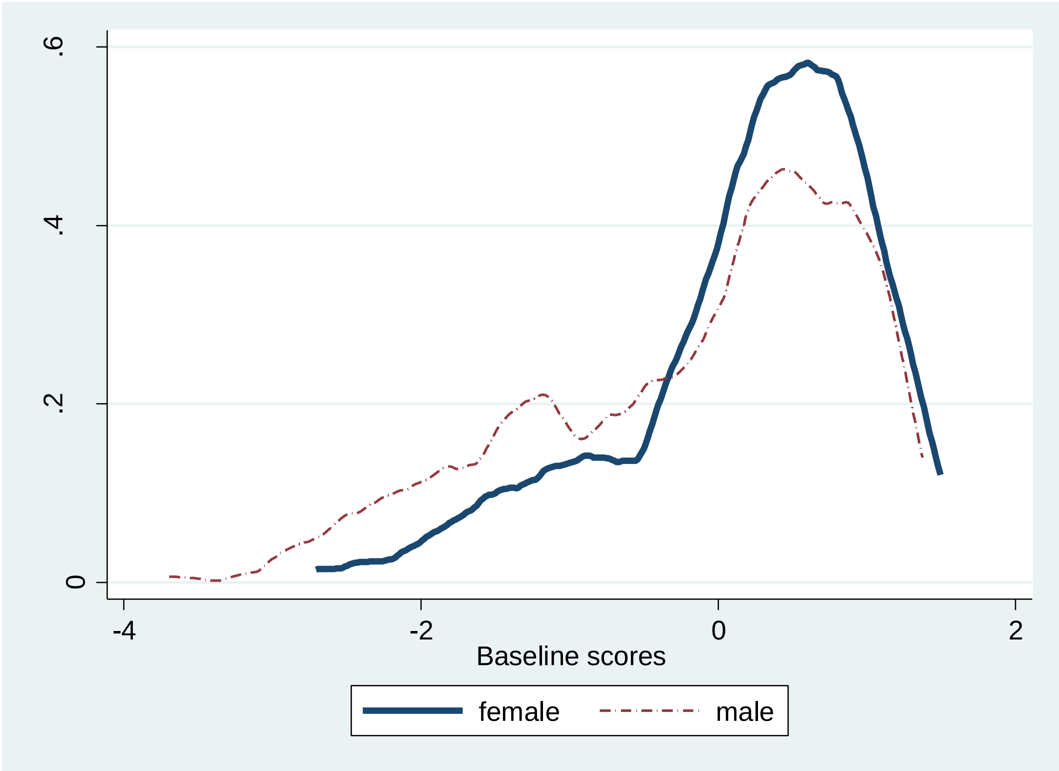


Figure 3.3 Distributions of midterm and final scores by genders

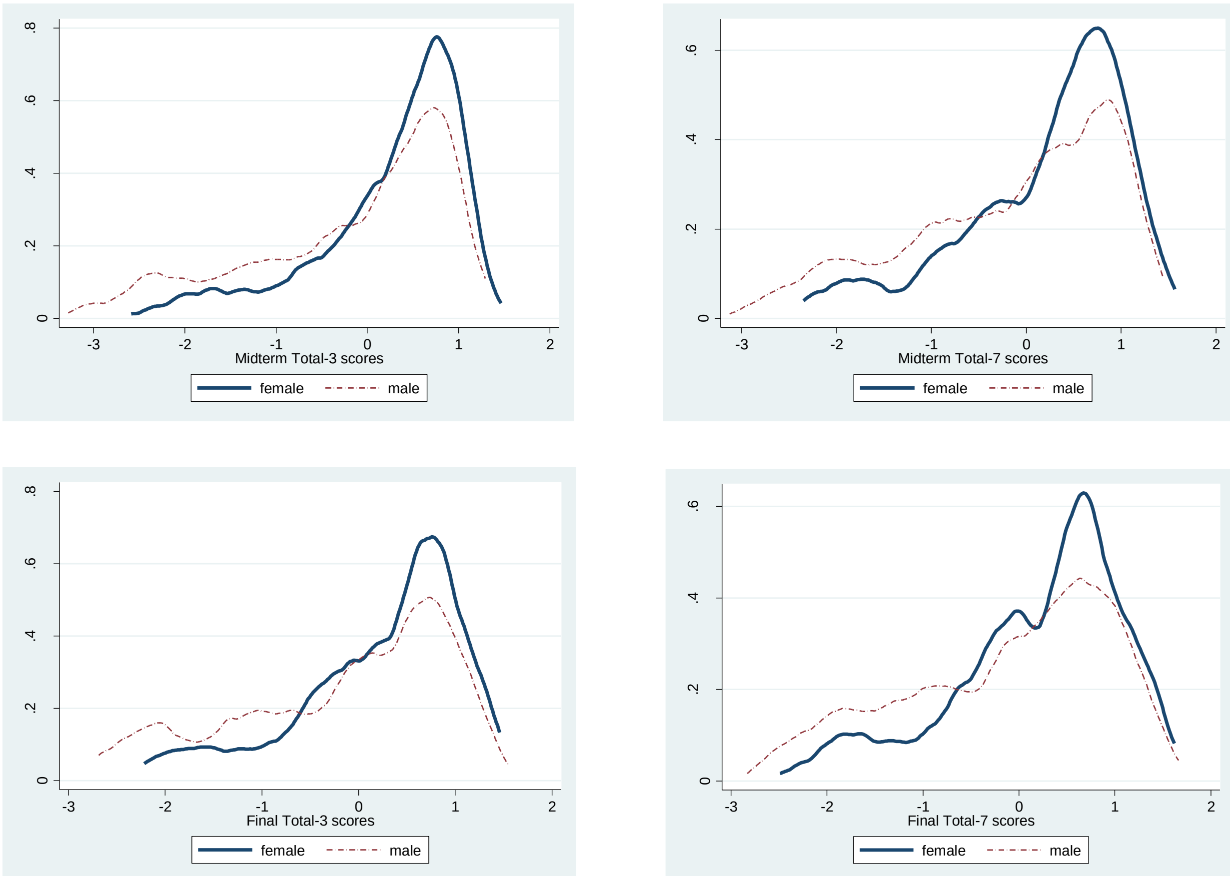


Table 3.1 Summary Statistics

Variables	Obs	Mean	S.D.	Max	Min
Panel A: baseline characteristics					
Being a female	664	0.44	0.50	0	1
Being in favorable list	664	0.08	0.28	0	1
Body height (centimeters)	664	156.22	6.51	135	178
Age (years)	638	12.46	0.54	10.17	14.5
Number of children	635	1.83	0.77	1	7
Father's education	633	11.65	3.06	3	19
Mother's education	629	10.15	3.68	3	19
Family income relative to others	635	3.16	0.76	1	5
Living in school	603	4.17	0.74	1	5
How much like school (pre)	626	4.02	0.81	1	5
Interest in Chinese (pre)	621	3.95	1.04	1	5
Interest in English (pre)	626	4.05	0.86	1	5
Interest in Math (pre)	641	18.65	2.92	9	22
Ideal education level (pre)	639	78.41	21.37	0	100
Chance of getting ideal level (pre)	664	0.44	0.50	0	1
Panel B: post experiment characteristics					
How much like school (post)	648	4.10	0.84	1	5
Interest in Chinese (post)	590	4.03	0.86	1	5
Interest in English (post)	588	4.01	1.09	1	5
Interest in Math (post)	592	4.05	0.93	1	5
Ideal education level (post)	641	18.63	3.01	9	22
Chance of getting ideal education (post)	649	73.15	21.93	0	100
Effects of desk-mate on self study	643	3.33	1.04	1	5
Effects of self on desk-mate study	641	3.48	0.86	1	5
Relation between self and desk-mate	641	3.74	1.13	1	5
Desire to change seats	648	3.34	1.34	1	5
Panel C: characteristics of peers					
Female desk-mate	664	0.44	0.50	0	1
Percent of females in neighbor 4 students	664	0.44	0.27	0	1
Percent of females in neighbor 5 students	664	0.44	0.23	0	1
Baseline score of desk-mate	664	0.00	1.00	-3.69	1.50
Average baseline score of neighbor 4 students	664	-0.01	0.58	-2.33	1.12
Average baseline score of neighbor 5 students	664	0.00	0.49	-1.73	1.08

Table 3.2 Tests for Exogeneity between Peer Characteristics and Self Backgrounds

	Desk-mate female	Neighbor 4 female	Neighbor 5 female	Desk-mate score	Neighbor 4 score	Neighbor 5 score
Panel A:						
Being female	*-0.055*	*-0.156*	*-0.256**	-0.007	0.012	0.001
	(0.064)	*(0.093)*	*(0.134)*	(0.021)	(0.034)	(0.045)
Baseline test score	-0.041	0.103	0.055	*-0.075*	*-0.103*	*-0.197*
	(0.082)	(0.117)	(0.158)	*(0.046)*	*(0.126)*	*(0.157)*
Panel B:						
SSR of restricted model	155.7	707.8	871.0	587.4	2872.1	3502.7
SSR of full model	84.8	575.6	745.4	324.4	2363.8	3017.6
K1	63	63	63	63	63	63
K2	332	663	663	332	663	663
N	664	2964	3628	664	2952	3616
F statistics	1.032	0.881	0.833	1.000	0.820	0.791
P value	0.39	>0.99	>0.99	0.50	>0.99	>0.99
Panel C:						
Body height (centimeters)	0.117	0.583	0.791	-0.042	0.012	-0.026
	(0.177)	(0.672)	(0.715)	(0.100)	(0.246)	(0.294)
Age (years)	-0.030	0.013	-0.040	0.000	0.008	0.005
	(0.030)	(0.090)	(0.105)	(0.021)	(0.046)	(0.043)
Number of children	-0.041	-0.161	-0.242*	0.003	-0.007	-0.016
	(0.056)	(0.137)	(0.127)	(0.035)	(0.056)	(0.060)
Father’s education	0.019	0.138	0.230	-0.059	0.123	0.068
	(0.129)	(0.248)	(0.236)	(0.046)	(0.101)	(0.145)
Mother’s education	-0.017	0.214	0.231	-0.041	0.064	0.034
	(0.117)	(0.182)	(0.208)	(0.036)	(0.129)	(0.167)
Family relative income	-0.046	0.041	-0.008	-0.021	0.001	-0.035
	(0.042)	(0.099)	(0.104)	(0.031)	(0.048)	(0.067)
How much like school (pre)	-0.001	-0.021	-0.043	0.016	0.026	0.057
	(0.042)	(0.105)	(0.113)	(0.028)	(0.053)	(0.055)
Interest in Chinese (pre)	0.036	0.088	0.109	-0.061*	0.100*	0.023
	(0.063)	(0.157)	(0.123)	(0.029)	(0.047)	(0.049)
Interest in English (pre)	0.101	-0.113	-0.032	-0.050	0.001	-0.058
	(0.071)	(0.121)	(0.144)	(0.030)	(0.049)	(0.038)
Interest in Math (pre)	-0.032	-0.168	-0.257*	-0.013	0.019	-0.000
	(0.047)	(0.110)	(0.143)	(0.023)	(0.066)	(0.074)
Ideal education (pre)	0.578**	-0.765*	-0.080	-0.008	-0.094	-0.073
	(0.188)	(0.349)	(0.458)	(0.109)	(0.134)	(0.161)
Chance of getting ideal education (pre)	2.153	0.656	1.912	-0.549	-0.781	-1.512
	(1.695)	(3.546)	(4.567)	(0.833)	(1.765)	(2.285)

Notes:
Coefficients in Panel A and C are from separate regressions. Regressions in Panel A and C control for block fixed effects, self gender, self baseline score, and status in the favor list whenever possible. Standard errors, clustered at the classroom level, are in parentheses. There are 12 classrooms.
* denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3.3 Main Effects of Peer Characteristics

Independent variables	Midterm Core score (1)	Midterm Core score (2)	Midterm Total score (3)	Midterm Total score (4)	Final Core score (5)	Final Core score (6)	Final Total score (7)	Final Total score (8)
Female desk-mate	0.066*	0.066*	0.079**	0.079**	0.052	0.054	0.066*	0.067*
	(0.035)	(0.033)	(0.034)	(0.032)	(0.039)	(0.040)	(0.031)	(0.033)
Baseline score of desk-mate	-0.006	-0.003	-0.004	-0.002	0.009	0.009	0.005	0.005
	(0.013)	(0.013)	(0.016)	(0.015)	(0.013)	(0.013)	(0.014)	(0.015)
Proportion of females in Neighbor 4		0.037		0.073		0.048		0.041
		(0.045)		(0.041)		(0.065)		(0.069)
Baseline score of Neighbor 4		0.072*		0.082*		-0.002		0.022
		(0.036)		(0.039)		(0.031)		(0.034)
Block fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Other controls	Y	Y	Y	Y	Y	Y	Y	Y
Observations	656	656	655	655	658	658	658	658

Notes:
Other controls include self gender, self baseline score, status in the favor list and ideal education. Standard errors, clustered at the classroom level, are in parentheses. There are 12 classrooms.
* denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3.4 Heterogeneous Effects of Peer Characteristics

Independent variables	Midterm Core score (1)	Midterm Total score (2)	Final Core score (3)	Final Total score (4)	Midterm Core score (5)	Midterm Total score (6)	Final Core score (7)	Final Total score (8)
Self Female*Female desk-mate	-0.046	-0.013	-0.027	-0.008				
	(0.062)	(0.067)	(0.053)	(0.055)				
Self Female* Proportion of females in Neighbor 4	0.307**	0.322**	0.270**	0.319**				
	(0.139)	(0.125)	(0.108)	(0.111)				
Self baseline score*Female desk-mate					-0.070	-0.072*	-0.057	-0.072**
					(0.043)	(0.034)	(0.036)	(0.032)
Self baseline score*Proportion of females in Neighbor 4					0.061	0.005	0.071	0.022
					(0.084)	(0.087)	(0.102)	(0.109)
Female desk-mate	0.083	0.082	0.063	0.068	0.067*	0.080**	0.056	0.068*
	(0.051)	(0.049)	(0.054)	(0.049)	(0.034)	(0.033)	(0.039)	(0.031)
Baseline score of desk-mate	-0.002	-0.000	0.010	0.007	-0.003	-0.002	0.010	0.006
	(0.012)	(0.014)	(0.013)	(0.014)	(0.013)	(0.015)	(0.014)	(0.015)
Proportion of females in Neighbor 4	-0.100	-0.073	-0.073	-0.104	0.037	0.072	0.048	0.040
	(0.085)	(0.072)	(0.097)	(0.107)	(0.047)	(0.045)	(0.065)	(0.071)
Baseline score of Neighbor 4	0.070*	0.081*	-0.003	0.021	0.071*	0.080*	-0.003	0.021
	(0.035)	(0.038)	(0.031)	(0.033)	(0.036)	(0.038)	(0.031)	(0.033)
Block fixed effects	Y	Y	Y	Y	Y	Y	Y	Y

Other controls	Y	Y	Y	Y	Y	Y	Y	Y
Observations	656	655	658	658	656	655	658	658

Notes:

Other controls include self gender, self baseline score, status in the favor list and ideal education. Standard errors, clustered at the classroom level, are in parentheses. There are 12 classrooms.

* denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3.5 Heterogeneous Effects of Neighbor 5 Students by Gender

Independent variables	Midterm Core (1)	Midterm Core (2)	Midterm Total (3)	Midterm Total (4)	Final Core (5)	Final Core (6)	Final Total (7)	Final Total (8)
	Female	Male	Female	Male	Female	Male	Female	Male
Proportion of females in Neighbor 5	0.294***	-0.075	0.339***	-0.038	0.263***	-0.016	0.307***	-0.046
	(0.068)	(0.085)	(0.070)	(0.068)	(0.064)	(0.139)	(0.067)	(0.151)
Baseline score of Neighbor 5	-0.039	0.163*	-0.014	0.151*	-0.036	0.027	-0.032	0.065
	(0.053)	(0.076)	(0.060)	(0.076)	(0.035)	(0.049)	(0.049)	(0.056)
Block fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Other controls	Y	Y	Y	Y	Y	Y	Y	Y
Observations	289	367	289	366	291	367	291	367

Notes:
Other controls include self gender, self baseline score, status in the favor list and ideal education. Standard errors, clustered at the classroom level, are in parentheses. There are 12 classrooms.
* denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3.6 Peer Effects on Subject Evaluations

Independent variables	Desk-mate effects On self study	Self effects On desk-mate study	Relation with desk-mate	Desire to change seat
Female desk-mate	0.154	-0.012	0.020	0.226
	(0.107)	(0.087)	(0.090)	(0.147)
Baseline score of desk-mate	0.111**	0.033	0.047	0.042
	(0.048)	(0.037)	(0.051)	(0.081)
Proportion of females in Neighbor 4	0.215	0.074	-0.010	0.012
	(0.140)	(0.117)	(0.153)	(0.206)
Baseline scores of Neighbor 4	0.069	0.002	0.020	-0.023
	(0.082)	(0.067)	(0.093)	(0.131)
Self female	0.160	0.073	0.008	0.100
	(0.099)	(0.080)	(0.091)	(0.087)
Self baseline score	0.116	0.186**	0.199***	-0.059
	(0.078)	(0.061)	(0.053)	(0.056)
Other controls	Y	Y	Y	Y
Block fixed effects	Y	Y	Y	Y
Observations	643	641	643	648

Notes:
Other controls include status in the favor list and ideal education. Standard errors, clustered at the classroom level, are in parentheses. There are 12 classrooms.
* denotes significance at the 10% level, ** at the 5% level, and *** at the 1% level.