Media Attention and Choice of Major: Evidence from Anti-Doctor Violence in China*

Shiyu Bo†, Joy Chen‡, Yan Song§, Sen Zhou¶

This draft: November 7, 2019

Abstract

We examine the effect of media persuasion on educational choice, and find that Chinese newspaper articles on violence against doctors influence students' decisions to study medicine at college. We match articles from over 1,200 newspapers with an administrative dataset on college entrance enrollment from 2005 to 2011, and find that one additional article on anti-doctor violence leads to a 0.6% decrease in the number of students enrolled in medicine-related majors, and this effect is more pronounced for physician and nursing majors. We perform a series of checks to ensure that the effect is driven by exposure to violence-related news rather than violent incidents themselves. An instrumental variable approach that exploits plausibly exogenous variations in local political turnover and province-wide violent incidents helps establish causality. Moreover, we find that exposure to violence-related news reduces the quality of medical students, measured by their rank in the college entrance examination. Our findings suggest that media coverage can change individuals' perception of

*This study was funded by National Natural Science Foundation of China [No. 71803064], Special funds for basic scientific research of centrally-managed universities [No. 17JNQN005], Guangzhou Philosophy and Social Science Planning Grant [No. 2019GZYB24] and the 111 project of China [No. B18026]. We would like to thank the editors, Daniela Puzzello and Nigar Hashimzade, and three anonymous referees, whose comments and suggestions greatly improve this paper. This paper benefited from helpful comments from Sumit Agarwal, Yiwei Chen, Jing Liu, Menghan Shen, Brian Viard, and seminar participants at Asian Meeting of the Econometric Society, China Meeting of the Econometric Society, International Symposium on Contemporary Labor Economics, Young Economist Society, Conference of Educational Finance, Workshop for History and Economic Development, Sun Yat-Sen University, Central University of Finance and Economics, Beijing Normal University, Harbin Institute of Technology Shenzhen and Chinese Academy of Social Sciences. Yujian Chen, Qianyi Lu, Haoran Zhang and Yan Zhou provided excellent research assistance. All errors are our own. We declare no conflict of interest.

†Institute for Economic and Social Research, Jinan University. Zhounghui Building, Room 105, Guangzhou 510632, China. Email: boshiyu@outlook.com.

‡Cheung Kong Graduate School of Business. Tower E1, Oriental Plaza, 1 East Chang An Avenue, Beijing 100738, China. Email: joychen@ckgsb.edu.cn.

§Corresponding author. Institute for Economic and Social Research, Jinan University. Zhounghui Building, Room 112A, Guangzhou 510632, China. Email: kprinceing@gmail.com.

¶China Institute for Educational Finance Research, Peking University. Room 406, Graduate School of Education, Peking University, Beijing 100871, China. Email: zhousen866@gmail.com.

career risks and affect their educational choices.

Keywords: Media attention; College majors; Occupational choice; Health care
JEL Classification: D84; H75; I18; I23; J24; O15

1 Introduction

Media can change individuals' beliefs and affect their decisions. Abundant evidence shows that media persuasion exists (DellaVigna and Gentzkow 2010). For example, DellaVigna and Kaplan (2007) documented how media bias influenced citizens' voting behavior, and Strömberg (2004) discussed how media attention swayed politicians' public spending decisions. However, these studies primarily focus on the political sphere, and relatively little is known outside this area (DellaVigna and La Ferrara 2015). Although limited evidence suggests that spending time on radio and television can produce negative effects on educational achievement (Gentzkow and Shapiro 2008), it is not clear that how the exposure to a particular type of media content affects educational choices.

In this paper, we investigate how media attention affects high school graduates' college major choices by shaping their perceptions of their future career risks. We focus on media coverage of violence against doctors in China. These events, commonly known as *yinao* in Chinese, surged in the early 2000s. They typically escalate from disputes between patients and doctors over treatment procedures or outcomes, and lead to the use of violence by patients against doctors. Anti-doctor violence in China has attracted attention both at home and abroad. An editorial in the Lancet (2010) warned that "China's doctors are in crisis", and an article in the New Yorker (Beam 2014) provided a detailed portrayal of a notable violent event in China. Anecdotal evidence suggests that many doctors perceive this risk to be real and would consider alternative career paths (Xu and Zhang 2014). Thus, a natural question is whether media coverage of these incidents of violence discourage students from pursuing a health-care related career. This question has important policy implications because students' college major choices determine the supply of human capital in the health-care sector.

We utilize two unique datasets to answer this question: a news dataset and an administrative dataset on China's college entrance examinations. We construct the news dataset by collecting news articles on anti-doctor violence that were published between 2000 and 2012 in more than 1,200 newspapers. The administrative dataset contains information on the college entrance examination scores and college majors of over 15 million high school graduates, and covers all students entering Chinese colleges between 2005 and

2011. Combining these two datasets enables us to provide new insights into the impact of media persuasion on individual perceptions in the context of educational choices.

We operationalize the concept of media attention by computing the total number of articles reporting anti-doctor violence that are published in newspapers circulating in each city and each academic year. The circulation rules of Chinese newspapers enable us to exploit geographical variations in media attention across cities and over time to identify the persuasion effect. Using a generalized difference-in-differences (DID) framework, we find that one additional article leads to an approximately 0.6% decrease in the number of students choosing medicine majors. This effect almost doubles when we focus on the sub-disciplines in which physicians and nurses are trained. Besides the intensive margin, we find that the extensive margin of media reports is also effective at discouraging students from choosing medicine majors. We validate the identification assumption of the baseline regression using an event-study approach and falsification tests.

To disentangle the media persuasion effect from the direct effect of violent incidents, we conduct a series of tests exploiting the rich information contained in our unique news dataset. First, we show that the students from cities in which no local violent incidents occur are also discouraged by violence-related articles from studying medicine, and that our baseline results are robust to explicitly controlling for the cumulative number of past local violent events. Second, we perform a textual analysis on each news article by applying a machine-learning algorithm to its title to derive a sentiment score denoting its negativity. We then group the news articles into two categories based on whether their sentiment scores are above or below the sample average, and show that effects in the baseline regressions are entirely driven by the number of articles with above-average negativity. Similarly, we extract information on which page an article appears in the newspaper, and find that baseline results are driven by articles on the front pages. These results demonstrate that changes in the students' choices of major are driven by their perceptions of news articles rather than the direct effect of the violent incidents.

To further address the endogeneity concerns, we use an instrumental variable approach to support a casual interpretation of our findings. Our instrument is analogous to the strategy exploiting media pressure from other newsworthy events, which is commonly used in studies on media impact (Eisensee and Strömberg 2007). Our measure for news pressure is a dummy variable that is equal to one if there is a turnover of a city's political leaders in a given year. During political turnover years, the increasing need to cover the background, track record and new policies of newly appointed leaders as well as the achievements of departing leaders may crowd out reports on anti-doctor events. We follow Nunn and Qian (2014) and Sequeira et al. (2019) and construct a DID style in-

strument by interacting the political turnover dummy with the total number of violent incidents in the same province and year. The intuition for this instrument is that political turnover should produce a smaller effect on media coverage of anti-doctor violence when there are fewer available incidents to report.

The validity of this instrument relies on the interaction term being exogenous conditional on the control variables. One concern is that a city's political cycle may be correlated with the economic conditions in the city (Li and Zhou 2005; Chen and Kung 2016), which could influence students' choices of major through channels other than news reports. Therefore, we explicitly control for local economic performance such as GDP level and growth rate, fiscal revenues and industrial outputs in all regressions. Other important factors relating to political selection in China such as connections to higher level officials (Jia et al. 2015; Landry et al. 2018) are unlikely to be associated with local conditions that affect students' college enrollment decisions. One additional concern is that the number of violent incidents in the same province may be endogenous because it includes the number of local events. We address this concern by explicitly controlling for the number of local events in the regressions. The number of events in all other cities in the province should be uncorrelated with any unobserved confounders at the local level, conditional on the number of local events and the province-by-year fixed effects.

We further explore the welfare implications by investigating the consequences of discouraging students from choosing medicine-related majors. We first confirm that discouraged students are those who attend elite colleges and achieve above-median test scores. We then find that these students tend to substitute towards majors that are related to medicine, such as pharmacy, biology, chemistry, engineering, and other science majors. Finally, we estimate a generalized triple-difference model to study the effects on the quality of students choosing medicine-related majors to evaluate the welfare implications. We find that reports of anti-doctor violence lead to a decrease in medicine students' average and lowest percentile ranks in the college entrance examinations.

This paper contributes to several strands of literature. First, we add to the literature that examines how beliefs affect students' college major choices. These beliefs are usually related to two factors: economic returns and self-assessment. Manski (1993) argued that educational choices are determined by students' perceived returns instead of actual returns. In particular, students in developing countries have strong misperceptions about the true returns of attending college (Jensen 2010; Kaufmann 2014). A more recent study by Choi et al. (2019) demonstrated that the salience of superstar firms drives students into these industries. In addition to future returns, students' educational choices are affected by their perceptions of non-economic factors, such as their own abilities and preferences.

Stinebrickner and Stinebrickner (2013) found that students' misperceptions about their ability to perform well academically in science drove them to choose a science degree upon entering college but not to graduate with one. Zafar (2011) reached a similar conclusion using subjective belief data, while Wiswall and Zafar (2014) found that earnings expectations and ability perceptions both play an important role. Fricke et al. (2018) found that exposure to a field of study also affected students' choices of majors. Delavande and Zafar (2019) further extended the analysis to non-pecuniary factors enjoyed at school.

To the best of our knowledge, this paper is among the first to examine the salience of occupational risk and students' college major choices. In particular, we study students' perceptions of a novel and important factor, future career risk. Although we do not explicitly provide new information for students regarding their future income, the incidents of violence that we focus on can be interpreted as information about the risk-adjusted income for medicine-related majors.

We also extend the focus of the media persuasion literature to choices made in an educational context. The literature mainly focuses on how media affect the decisions of voters and politicians (DellaVigna and Kaplan 2007; Enikolopov et al. 2011; Lim et al. 2015; Qin 2013),[1] and occurrences of wars and civil conflicts (Yanagizawa-Drott 2014; Jetter 2017; Durante and Zhuravskaya 2018). Outside of the political sphere, fewer studies have examined human choices and decisions in response to media shocks in other areas (DellaVigna and La Ferrara 2015). Our paper further explores the role of media in education, and provides evidence that media attention affects the accumulation of human capital through its influence on students' college major choices.

The remainder of this paper is organized as follows. Section 2 describes the institutional background. Section 3 describes the data. In Section 4, we use two empirical methods to investigate whether violence-related news discourages students from choosing medicine majors. Section 5 further explores the consequences of discouraging students. Section 6 concludes the paper.

2 Institutional Background

In this section, we provide a synopsis of the institutional environment of the health-care system, college admissions procedures, and newspaper publishing in China.

[1] See Strömberg (2015) for a recent survey.

2.1 Health-care System and Violence Against Doctors in China

In this subsection, we describe how the health-care system in China functions and discuss some of the features that have contributed to the escalating tension between patients and doctors.

Medical Services. The health-care system in China is structured into three-tiers. Primary-level clinics, such as community health centers, provide basic medical care including giving flu shots, checking blood pressure, and performing electrocardiography. Secondary and tertiary hospitals provide specialized medical services. Patients are free to visit any public hospital without referral from a primary-care provider, and receive diagnoses, medical treatment, and drug prescriptions.

Despite the considerable variation in the quality of care provided, hospitals in different tiers charge similar fees for outpatient and inpatient visits. As a result, top-tier hospitals are often over-crowded, forcing physicians to reduce the length of their consultations and often leave patients dissatisfied.

Economic Incentives. Public hospitals in China face two sources of financial difficulty. First, state funding of medical services has been in continuous decline since 1979, with the proportion of national health-care spending decreasing from 32% to 15% over the ensuing 20 years (Liu et al. 2000). Second, the state has implemented strict controls over how much hospitals can charge for routine visits and services to ensure that everyone has access to basic health-care (Wu et al. 2014).

As a result, Chinese hospitals need to find alternative sources of revenue to remain operational. Hospitals are permitted to earn profits from prescribing new drugs and providing high-tech tests, with a markup of 15% (Hesketh and Zhu 1997). In response, hospitals encourage physicians to prescribe drugs and issue tests by tying their income to the profits they can bring to the hospital. This policy has resulted in the over-prescription of unnecessary and expensive pharmaceuticals and high-tech services such as imaging (Currie et al. 2014), leading patients to believe that doctors and hospitals conspire to increase the charges (Lancet 2010).

Violence Against Doctors. Although incidents of anti-doctor violence have their own idiosyncrasies, the common triggers of violence fall into several broad categories, such as patient dissatisfaction, mistrust of physicians and nurses, and disputes over the outcomes of medical treatments. Patients and/or their relatives resort to various forms of physical violence as a means to express their anger. Group protests are also organized by relatives to force hospitals to pay compensation for what they believe to be medical malpractice.

For more context, we provide examples of each type of incident below.

1. In 2006, a man visited a hospital to receive a diagnosis for his stomach problems, and engaged in a dispute with his doctor over medical charges. The next day, he returned to the hospital and assaulted the doctor with a knife, cutting off her right hand and rupturing her left eyeball.[2]

2. In 2007, after a woman gave birth to a stillborn baby, angry relatives congregated in the front hall of the hospital. The relatives lit candles, burnt incense, placed funeral flowers, and set off firecrackers, forcing the hospital to shut down its operations and send its patients elsewhere. The relatives left a few days later after the hospital agreed to pay RMB200,000 in compensation.[3]

As these examples suggest, violence against doctors is closely related to unique features of the Chinese health-care sector, and has resulted in large disruptions to the daily operations of hospitals and severe physical and psychological injuries to the affected doctors and/or nurses. In other words, such violence poses a serious threat to the safety and wellbeing of medical personnel.

2.2 College Admissions System

In China, the college admissions process involves two steps. First, in June each year, all college-aspiring high school graduates sit the National College Entrance Examination (NCEE). The graduates then submit their school and major preferences and they are matched to their colleges through a centralized admissions process.

Examinations. Entrance examinations and college admissions are organized at the province level, and each provincial educational authority sets and grades papers independently. Similarly, the four directly-administered municipalities (Beijing, Shanghai, Tianjin, and Chongqing) set their own papers and have independent admissions procedures.

High school students self-select into one of two academic tracks, science or humanities, in Year 11, and take the NCEE in June of Year 12.[4] Each academic track has its own examination papers, admission quotas, and matching procedures. Therefore, only scores

[2]Sohu, Inc. "Yiyuan Duoshou Shijian Zhong De Nan Zhujue." SOHU.com. http://news.sohu.com/20070103/n247397673.shtml (accessed September 29, 2019).

[3]Sohu, Inc. "Fujian Sheng Yige Yue Nei Lianxu Fasheng Siqi Yihuan Jiufen Shijian Fansi." SOHU.com. http://news.sohu.com/20070821/n251690410.shtml (accessed September 29, 2019).

[4]All students must study Chinese, mathematics, and English. Students in the science track study biology, chemistry, and physics, while students in the humanities major study geography, history, and politics.

in the same province-year-track are comparable, as we emphasize in our empirical analysis. Furthermore, students in the science track face no restrictions in choosing their college majors, whereas students in the humanities track cannot enroll in science and engineering disciplines. Thus, undergraduates in medicine majors are almost exclusively from the science track.

Preferences and Admission. Either before or immediately after taking the NCEE, high school graduates submit a ranked list of preferences for college-and-major to their provincial educational authorities. The deadline is usually before July, thus making the academic year from July of the previous year to June of the current year. This timing means that when making their decisions, many students lack important information about the content of the specific majors, and the potential career paths and expected incomes for different disciplines. Thus, their choices of college and major are influenced by peers, family preferences, and external information sources such as news reports. As most Chinese colleges are public institutions, tuition fees are unlikely to be a concern in students' college preferences because the fees are generally low and vary minimally. Moreover, financially disadvantaged students can apply for aid.

In accordance with the examination and admissions systems, colleges allocate quotas for each major at the province-track level and select students with the highest test scores from the entire pool of applicants. In other words, a student's admission outcome is determined by her NCEE score, her submitted preference, and the relevant admission quota. As college degrees are highly valued in the Chinese labor market, nearly all of the programs have minimal trouble filling the assigned quotas. Popular majors and programs are usually associated with higher admission scores.

Medicine Majors. Unlike in the United States, medicine-related studies in China begin at the undergraduate level. Students who wish to become doctors or pursue graduate studies in medicine must complete a bachelor's degree in medicine.

Thus, college admissions in medicine-related majors largely determine the future supply of doctors for two reasons. First, students cannot transfer between colleges, and only a very small number of students, normally less than 5%, are allowed to transfer to a different major if they meet certain requirements set by their college and department. Second, in China, a bachelor's degree in medicine is the minimum prerequisite for obtaining physician certification. Therefore, the number of college medicine admissions provides a close estimate of the number of college medicine students, and can serve as an upper bound on the number of future eligible doctors.

2.3 Newspapers

All Chinese newspapers are entirely or primarily owned by the state. As a result, the ownership structure of newspapers parallels the four-level hierarchy of China's political administration: state, province, city, and county. According to the *China Newspaper Industry Statistical Yearbook (2016)*, city-level newspapers constitute the largest category, and are published by more than 800 different outlets. About 1,000 newspapers are published at the provincial level and the state level, and only 20 at the county level.

Corresponding to the four-level administrative hierarchy, newspapers are circulated strictly at the level of their administrative affiliation. For instance, newspapers published at the city-level are circulated within their cities of publication, whereas national newspapers are circulated nationwide. Moreover, city-level newspapers are also circulated in counties that are the cities' lower-level administrative subdivisions.

Our empirical analysis exploits this circulation pattern, and makes use of articles on anti-doctor violence that are printed in newspapers circulated at the city-level. The institutional arrangements detailed above allow us to make the assumption that these articles only potentially affect students who are residents of their cities of publication or the cities' administrative subdivisions. We exclude articles published in the provincial and national newspapers, because students in each province have the same access to these stories, and the impact of the stories can be accounted for by province-by-year fixed effects in a regression. In this way, our identification strategy utilizes variations in the number of violence-related articles in each city and academic year, and compares the admission outcomes of students in the same province-year-track.

The ownership structure of China's newspapers has implications for their content. While newspapers disseminate the Communist Party's ideology, promote government policies, and suppress negative news, they need to attract commercial advertisers to earn profits (Zhao 1998; Qin et al. 2018). Driven by the profit-seeking motive, Chinese newspapers also publish stories that cater to public demand. Violence against doctors is a topic that has attracted public attention and stimulated discussions, as many perceive healthcare issues to be relevant to their daily life. Therefore, local newspapers tend to publish stories about incidents of anti-doctor violence that occur both within and outside their region of circulation to attract readership.

2.4 Political Turnover

The Chinese bureaucratic hierarchy consists of five levels of local government: province, city (prefecture), county, township, and village. Two leaders are responsible for local

economic and political affairs. The party secretary is the head of local branch of the Communist Party, while the mayor is the head of the local government. The two political leaders are appointed by the higher-level government (Li et al. 2018; Burns 1994; Landry 2008). The term of office is five years, and each leader can only serve a maximum of two terms. In practice, however, the tenure served is often less than five years and the rate of turnover is rather high (Xu 2011; Yao and Zhang 2015). Between 1983 and 2012, the average tenure of party secretaries and mayors was 3.7 and 3.3 years, respectively (Li et al. 2018). Studies have demonstrated that economic performance and factionalism are two key determinants of political promotion in China (Chen et al. 2005; Li and Zhou 2005; Shih et al. 2012; Jia et al. 2015; Meyer et al. 2016; Fisman et al. 2019) .

When political turnovers occur, local newspapers, as the mouthpieces of the local governments, tend to publish more propaganda covering the background and past track records of the newly appointed leaders, and the achievements and legacies of the departing leaders. We exploit these features of political leaders' turnover at the city level, and newspapers' pattern of publishing propaganda to obtain exogenous variations in media coverage of anti-doctor violence. After controlling for local economic development, political turnovers generate exogenous shocks to the amount of space that local newspapers can allocate to other stories, including violence against doctors. Therefore, we can use this variation to construct an instrument for our main explanatory variable, the number of news articles about violence against doctors. We discuss the details of the instrumental variable strategy in section 4.2.

3 Data Description

In this section, we describe the datasets used in this paper. Our main analysis combines data from three sources: the WiserNews database, which contains newspaper articles published in China after 2000, a unique administrative dataset on college enrollment, which contains information on every college admission in China from 2005 to 2011, and the official statistical yearbooks.

3.1 News Dataset

As there are no publicly available statistics on media coverage of violence against doctors, we use the WiserNews archives, a Hong Kong-based database that contains every newspaper article published in China since 2000. Qin et al. (2018) uses this database to study media bias in China. We extract relevant newspaper articles from WiserNews by au-

tomating a keyword search for news articles published between 2000 and 2011. First, we perform a generic keyword search for *yinao* (violence against doctors)[5] and extract a set of newspaper articles reporting incidents of violence against doctors across China from 2000 to 2011. From these articles, we obtain a list of events in which anti-doctor violence transpired, and construct unique keywords for each event. We include this step because some relevant articles do not explicitly use the generic keywords, but instead contain keywords referencing the actual events. For an event of violence, the unique keywords usually include the city and the name of the hospital at which the event happened.[6] We also include alternate names for the hospitals such as abbreviations and aliases. Lastly, we repeat the automated search for each incident of violence using these keywords.

We find a total of 1,984 unique articles for 233 incidents of violence against doctors, 757 of which are published in city-level newspapers. On average, 0.2 articles appear in a city each year. For each article, we have data on its full textual content, date of publication, page on which it was published, name of the newspaper in which it was published, and the city in which the newspaper was circulated. Figure A.1 in the Online Appendix provides an overview of the spatial distribution of these articles. It is clear that media coverage of anti-doctor violence is not restricted to a few cities or provinces, but is widespread across China. Panel A in Table 1 presents a more detailed summary of these articles over the sample period.

This dataset enables us to construct our key measure of media coverage of violence towards doctors. We define media coverage, $NewsArticle_{ijt}$, as the sum of articles reporting incidents of anti-doctor violence that are published in newspapers circulating in city i of province j in academic year t. Note that an academic year is from July of the previous year to June of the current year because students take the NCEE and submit their college-and-major preferences in June. An implicit assumption underlying this measure in that individuals have a short memory, and only recent reports affect students' choices of college major. We relax this assumption in Section 4.1.6, where we include newspaper reports from the previous years at a discounted rate.

Moreover, the rich set of information on news articles contained in our dataset allows us to examine additional dimensions of media coverage that are driven by other exogenous factors. We extract an article's sentiment from its title using a machine-learning

[5]The generic keywords that we use are *yinao* and its several synonyms and variants, all of which mean violence against doctors. They include: *baoli shang yi* (violently injure doctors), *exing shang yi* (maliciously injure doctors) *yihuan jiufen* (doctor-patient disputes), *yihuan chongtu*, and *yihuan maodun* (doctor-patient conflicts)

[6]In some cases, details such as the name and identity (e.g., patient, relative) of the aggressor(s), the weapon(s) used (e.g., knife), the aggressors' actions during the assault (e.g., stab, strike, detain), and the outcome of the event (e.g., death or injury of medical personnel), are also used as keywords.

approach.[7] The algorithm assigns a value between 0 and 1 to each article, with a value closer to 1 indicating a more negative tone, and a value closer to 0 indicating a less negative tone, in our context. The mean of the sentiment for articles in the dataset is 0.66, suggesting a consistently negative tone towards violent incidents. It should be noted that there are considerable variations in the sentiment of articles. For instance, the algorithm assigns a value of 0.99 to an article titled "Man Suspects Negligence from Doctors, Stabs Four Doctors," suggesting a high degree of negativity. A report titled "Angels in Steel Helmets" is given a value of 0.50, representing a neutral tone. The algorithm assigns a value of 0.01 to a report titled "True Philanthropy from Angels in White," indicating a more positive message. Overall, the standard deviation of the sentiment is 0.26, providing us with sufficient variation to work with.

An apparent disadvantage of this strategy is that we only observe media coverage through newspapers. Any discussions of violence against doctors in other media outlets are not captured in our dataset. Although the inclusion of news reports from other types of media outlets would be beneficial for our analysis, there are no reliable data sources that contain historical records of these reports. Nevertheless, we proxy for violence-related reports from other types of media outlets using information from household surveys in Section 4.1.6.

3.2 Administrative Dataset on College Admissions

We use a unique individual-level administrative dataset on college admissions in China. This dataset covers every student who took the NCEE and was admitted to a college between 2005 and 2011. For each student, we observe his or her academic track in high school, NCEE examination ID number (which can then be used to extract the location at which she was enrolled to take the examination), NCEE examination score, the college to which she was admitted, and a six-digit code representing her major.[8] This dataset allows us to observe the changes in the quantity and quality (as measured by the NCEE score) of students enrolled in medicine-related majors. More details about this dataset can be found in Bo et al. (2019).

Panel B in Table 1 presents the relevant summary statistics. The dataset covers all 30 provinces in mainland China except for Tibet.[9] Due to the expansion of tertiary education,

[7] The machine-learning engine is provided by Boson, a firm providing commercial natural language processing platforms.

[8] We use this code to determine the type of a student's college major. The first two digits of medicine majors are 10. For example, the first four digits of a major in clinical medicine are 1003, and the first four digits of a major in nursing are 1007.

[9] Due to unknown errors in the data collection process, there are no data for Jiangsu province in 2009

the number of college admissions steadily increased from 2 million to 2.7 million during the sample period. The total number of students studying medicine-related majors increased by approximately 30% for the same reason. Column 4 suggests that the ratio of students choosing medicine majors among all college students remained approximately constant during the sample period. This pattern arises from the quota system and the consent option discussed in the institutional background section.

3.3 Other Data

Most of the variables describing the local economy are obtained from the city statistical yearbooks and the regional economy statistical yearbooks, officially published by China's National Bureau of Statistics (NBS). The yearbooks contain information on various features of the local economy, covering all cities and counties in China. We extract data on GDP, the urban and rural population, fiscal revenue and expenditure, the number of hospitals, doctors and hospital beds, and the total number of industrial enterprises and their total output. Additional data on local demographics such as the percentages of illiterate people and elderly people are computed from the 2005 mini population census conducted by the NBS.

We calculate the penetration rate of televisions, computers, and mobile phones from the Urban Household Survey (UHS), a nationally representative survey conducted by the NBS annually since 1988. The design of the survey is similar to that of the Current Population Surveys in the United States, covering the basic socioeconomic conditions of urban households in China (Ge and Yang 2014). The survey has been widely used by scholars to study China's labor market and the urban economy (Meng 2012; Feng et al. 2017). We were able to access the 2005 UHS sample for 12 provinces, namely Shanxi, Liaoning, Heilongjiang, Jiangsu, Anhui, Jiangxi, Shandong, Henan, Hubei, Guangdong, Sichuan, and Gansu.

Lastly, we complement our dataset with information about local political leaders, in particular their turnover. This information was extracted from Chen (2015), who digitized the resumes of city-level party secretaries and mayors since 2000, covering all of the 333 prefecture-level cities in the 27 provinces in our dataset.

and Zhejiang province from 2011.

4 Media Discourages Students from Studying Medicine

In this section, we quantify media persuasion effects. We use two empirical strategies to causally identify this effect. Our first strategy uses a generalized DID approach that exploits variations in media reports across cities and over time. We extend this analysis to utilize variations in the sentiment and position of news articles. We perform a series of tests to check that the identification assumption of the DID approach is not violated. Our second strategy uses an instrumental variable approach to address inherent limits of the DID specification. This strategy exploits plausibly exogenous variations in news pressure in each city and year.

4.1 A Generalized Difference-in-differences Approach

4.1.1 Empirical Strategy

We first use a generalized DID framework to identify the causal effect of media attention on students' college major choices. Dahl and DellaVigna (2009) and Jacobsen (2011) used similar strategies to identify the impact of films on violent crimes and decisions to purchase carbon offsets, respectively. Following their approach, the baseline DID specification is displayed below:

$$y_{ijt} = \alpha + \beta NewsArticle_{ijt} + \gamma' X_{ijt} + \delta_i + \theta_{jt} + \epsilon_{ijt} \tag{1}$$

where y_{ijt} is the outcome variable of interest for city i of province j in year t; $NewsArticle_{ijt}$ is the measure of media coverage in city i of province j in academic year t; and X_{ijt} is a vector of control variables including the logarithm of local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total number of industrial enterprises and their total output, and GDP growth rate. We also include the logarithm of the total number of college admissions meeting the same requirements as the outcome variable;[10] δ_i denotes city fixed effects; θ_{jt} represents province-by-year fixed effects; and ϵ_{ijt} is the error term.

Our identification strategy exploits regional and temporal variations in $NewsArticle_{ijt}$, and the treatment effect is captured by coefficient β. We exclude articles published in higher level newspapers, as their impact on students' decisions is accounted for by the province-by-year fixed effects. We exclude articles published in county-level newspapers, as there are only 20 such papers.

[10]For example, if a regression uses the number of medicine admissions at elite colleges as the dependent variable, the control variable is also the number of enrollments at elite colleges.

4.1.2 Main Results

Our analysis begins with testing the primary hypothesis: students who live in cities that face higher levels of media coverage of anti-doctor violence are less likely to pursue careers related to health-care. We examine this hypothesis by estimating Equation 1, with the dependent variable being the log number of students enrolled in health-care majors in city i of province j in year t. We also control for the log number of the total college enrollments in that city.

The first two columns of Table 2 report the baseline results. Column 1 displays the regression estimates of the generalized DID model with province-by-year fixed effects but without the additional control variables. Column 2 adds the control variables including the logarithm of local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total number of industrial enterprises and their total output, and the GDP growth rate. The regression model in column 2 is our preferred specification and acts as the baseline for comparisons with alternative specifications. All of the estimated standard errors in this paper are two-way clustered at the province and year levels to account for possible heteroskedasticity and serial correlation of the residuals in the estimation of standard errors (Cameron et al. 2011).

The regression estimates suggest that one additional violence-related article leads to a 0.5% to 0.6% decrease in the number of students studying medicine majors. In our preferred specification (column 2), one additional article leads to a 0.63% decrease in the number of students studying health-care related majors each year. This coefficient estimate is statistically significant at the 1% level. Moreover, the average number of medicine enrollments in each city is 491, which means that, on average, one additional violence-related article leads to 3.09 fewer medicine admissions in each city. Since approximately five news articles correspond to each event of violence, this coefficient translates into 15.45 ($= 491 \times 0.0063 \times 5$) fewer medical students in a city in response to one event of violence. Given that on average, 0.545 articles appear in a city each year for our sample, the magnitude of our estimated coefficient is not huge, but is still non-trivial.

The results above focus on the intensive margin of media reports on students' choices of college major. Another way to understand the media persuasion effect is by examining the extensive margin. We do this in two ways. First, we create a dummy variable $I(\#$ of News Articles$_{ijt} > 0)$, which equals 1 if year t. Second, we create a dummy variable $I(\text{Post 1st News Article}_{ijt})$, which equals 1 if year t comes after the publication year of the first report in city i, and 0 otherwise. Thus, this specification is a standard DID regression comparing changes in the number of medical students before and after the first violent news across cities. We re-estimate Equation 1 using these two variables as the

main explanatory variables. Column 3 and 4 in Table 2 reports the extensive margin results. We find that the appearance of any violence-related news reduces the number of medical students by an average of 2.5% and 3.16% respectively. These estimates suggest that the extensive margin of violence-related news also effectively discourages students from choosing medicine majors.

Violence in medical settings mostly directly affects physicians and nurses, whose working environment involves close interactions with patients and their relatives. As physicians and nurses usually hold degrees in clinical medicine and nursing, we hypothesize that the persuasion effect is more pronounced for these two majors. We re-estimate Equation 1 using the number of students enrolled in clinical medicine and nursing as the dependent variable. Columns 5 and 6 of Table 2 report the estimation results. Compared with baseline results, the size of the coefficient estimate almost doubles to nearly 1%, thus providing support for our hypothesis. Here, one additional article leads to a decrease of 2.16 admits in clinical medicine and nursing, accounting for 70% ($= 2.16/3.09$) of the decrease in all medical students.

4.1.3 Identification

In this subsection, we perform a series of checks to ensure the baseline results can adopt a causal interpretation.

Parallel Trend A common pre-trend condition is necessary for a DID estimator to be consistent. To verify this assumption, we use an event-study design to examine the treatment effects before and after the news events. The identification assumption is questionable if the treatment and control cities appear to be following different time trends before the first news article on anti-doctor violence is published. We test this identification assumption by estimating a set of yearly treatment effects starting at three years prior to the publication of the first local news article. The specification is displayed as follows:

$$y_{ijt} = \alpha + \sum_{\tau=-3}^{2} \beta_\tau I(YearSinceFirstNews_{ijt}) + \gamma * X_{ijt} + \delta_i + \theta_{jt} + \epsilon_{ijt} \tag{2}$$

where $I()$ is an indicator function, $YearSinceFirstNews_{ijt}$ counts the number years between t and the year of publication of the first report on violence in city i, and y_{ijt}, the log number of medicine admits, is our main outcome variable. Then, $YearSinceFirstNews_{ijt}$ takes negative values for the years prior to the publication of the first article on anti-doctor violence, and positive for the years thereafter.

By allowing the effect to vary by year in relation to the first newspaper article, this specification takes a more flexible form than the standard DID model in column 4 of Table 2. In particular, this specification enables us to check the pre-trends in the yearly treatment effects to verify the validity of the identification assumption. If the parallel trend assumption holds prior to the first violent event, we should see $\beta_\tau = 0$ when $\tau < 0$. Figure 1 presents coefficient plots for 3 and more years, 2 years, and 1 year prior to, and for 1 year, 2 years, and 3 or more years after the publication of the first article on anti-doctor violence. The omitted group is the year before the first incident, so all of the coefficients should be interpreted as relative effects to the effect in the event year.

Figure 1 shows that the treatment and control cities followed very similar trends before the first media coverage. There are no statistically significant differences in the pre-existing trends, and the coefficients before the first incident of violence are not statistically significantly different from zero. Thus, cities exposed to news shocks followed a similar trend in the number of students admitted to medicine majors compared to cities that had no such media coverage. In other words, the absence of evidence on differential pre-trends suggests that the incidence of news reports on violence is unrelated to changes in the number of students who gain admission to medicine majors in college. Therefore, the negative effects found in the baseline regressions can adopt a causal interpretation.

Nevertheless, two further concerns arise beyond this check. First, it only exploits the extensive margin as in column 4 of Table 2, but not the intensive margin as in our preferred specification in column 2 of Table 2. To address this issue, we conduct a falsification test using the preferred specification and use an instrumental variable approach to provide stronger evidence for causality. Second, as shown in Table 1, reports of anti-doctor violence had already been published in over 20 cities by 2005, the year when the college admission dataset begins. Therefore, these cities do not contribute to the event study that counts the first incident of violence beginning from 2005. We perform a check in the robustness section to address this concern.

Falsification Test. In addition to checking the parallel trend assumption, we perform two falsification tests. First, we examine the impact of media coverage of anti-doctor violence on the number of students enrolled in arts majors. Since the risks and returns of careers in arts are unrelated to anti-doctor violence, media coverage of these events should not affect arts students. We also examine the impact of next year's media coverage of anti-doctor violence on the number of medicine enrollments in the current year. As students' decisions are based on current information, we should expect there to be no statistically significant effect. If we find otherwise, our identification strategy may be problematic.

We perform the first test by estimating Equation 1 using the log number of students enrolled in arts as the dependent variable. Results are displayed in Panel A of Table A.1 in the Online Appendix. These estimates are not statistically different from zero, suggesting that media coverage of anti-doctor violence does not affect students' decisions to study arts.

To implement the second test, we reproduce columns 2 and 6 from Table 2 using $NewsArticle_{ijt+1}$ as the key explanatory variable. The results are displayed in Panel B of Table A.1. None of the estimates are statistically different from zero, indicating that future information does not affect current decisions.

4.1.4 Perception of Newspaper Articles

In the baseline results, we find that the number of newspaper reports related to anti-doctor violence is associated with a reduction in the number of medical students. As the number of newspaper reports of anti-doctor violence is obviously correlated with the actual events, it is possible that students are discouraged by the events themselves rather than media coverage. Therefore, our estimates could overstate the media persuasion effect. In this subsection, we separate media persuasion effects from the actual incidents.

A straightforward way to address this issue is to focus on city-year observations with zero anti-doctor events. The first column of Table 3 presents estimation results for this subsample. In column 2, we further restrict the sample to include only cities with zero anti-doctor events in the entire sample period, as people may remember past events. Both of the estimated coefficients on the number of violence-related articles are negative and statistically significant. By construction, this estimate precludes the effect of violent incidents, and represents pure media persuasion effects.

To explicitly distinguish the two effects, we include the number of violent events in the baseline regression using the full sample. We account for violence in the past by computing the discounted sum of violent events in each city:

$$SumViolentEvent_{ijt} = \sum_{k>=0} \phi^k \times ViolentEvent_{ijt-k}, \tag{3}$$

where ϕ is the discount rate, and $ViolentEvent_{ijt-k}$ is the number of violent events in city i and year $t-k$. We allow individuals to discount historical events at a positive rate, $\phi \in [0,1]$, which can be interpreted as the probability that an individual recalls a past event. In column 3 of Table 3, we control for the total sum of local violent incidents (or, use a discount rate of $\phi = 1$). In column 4, we use a discount rate of $\phi = 0.5$. In both columns, coefficient estimates remain statistically significant, while the effect of the sum

of past local violent incidents is close to zero and statistically insignificant. These results suggest that our baseline results are mainly driven by students' exposure to violence-related news, and not by the effects of violent incidents themselves. For simplicity, we do not present the results for every possible discount rate, although they are actually highly similar. Since it is important to account for the direct effects of the violent incidents, in the following empirical analysis, we control for the accumulated number of past local violent events with a discount rate of 0.5 in every regression, without loss of generality.

In the remaining columns of Table 3, we provide further evidence that the effects on students' major choices do come from their perception of the violence-related news. The effectiveness of media persuasion depends not only on the number of articles in circulation, but also on the sentiments conveyed in them. For example, articles that discuss patients' hostility towards doctors at length or express pessimism about medical careers are more likely to change students' beliefs about the benefits of a career in medicine. Therefore, the persuasion effect will be greater if the articles adopt a more negative tone. Although few studies have explored variations in this dimension, our dataset contains information on the titles of the newspaper articles, which gives us a great opportunity to investigate how the sentiment of the articles affects students' choice of college major.[11]

To do so, we construct a measure for the sentiment of an article's title using a machine-learning approach. Section 3.1 provides more details on the construction of this variable. The sentiment score ranges from 0 to 1. A score close to 1 indicates greater negativity, and a score close to 0 indicates non-negativity, in our context. We then divide all articles into two groups: negative and non-negative, based on whether the article has a sentiment score that is more negative than the sample average of 0.66. We re-estimate our baseline regression using the numbers of negative and non-negative articles as our key explanatory variables. From column 5 in Table 3, we can see that articles with negative sentiment drive the media persuasion effect, while articles with non-negative sentiment produce little impact. Moreover, negative articles usually report about physical harm and details of violent actions, while non-negative articles are more likely to reflect a decrease in the reputation and social image of doctors. Thus, these results provide suggestive evidence that students are primarily discouraged by career risks rather than reputational concerns.

Another factor that can help us test media persuasion effects is the position of an article in a newspaper, which can affect how many readers it can reach. Cover stories tend to have more readers and generate larger persuasion effects, whereas articles printed on

[11] In principle, it is possible to analyze the full texts of newspaper articles for their sentiment. However, because the full text of many articles is longer than our on-line engine for sentiment analysis allows, we have to step back and analyze the titles instead.

page 20 are unlikely to attract much attention. Our news dataset documents the page on which an article appears, and the page number is 14.35. We define an article as front-page if it appears no later than page 14, and non-front-page otherwise. We then aggregate the number of articles in each group, and estimate the baseline model using the number of articles from the two groups as key explanatory variables. Results are presented in the last column of Table 3. Consistent with our intuition, only front-page articles affect students' major choices, while non-front-page articles produce little impact.

4.1.5 Heterogeneity

Thus far, our analyses have focused on the persuasive effects of newspaper articles. In this subsection, we examine different sub-populations' response to media coverage. More specifically, we divide our sample in three different ways that all share one common intuition, namely, that cities with a higher proportion of people who are more likely to read newspapers are more likely to see greater media persuasion.

First, we divide our sample into cities with high and low literacy rates. A city is in the high-literacy group if its literacy rate in 2005 is higher than the sample average, and vice versa. We expect cities with a more literate population to be more strongly affected by newspaper articles. In the first two columns in Table 4, we re-estimate the baseline regressions on these two groups. As expected, media persuasion effect only exists in high-literacy cities.

Following the same intuition, we divide sample cities based on their urbanization rates. As transportation infrastructure is under-developed in areas with a lower rate of urbanization, the circulation of newspapers will be more limited. Thus, we expect that the media persuasion effect will be larger in cities with a higher rate of urbanization. Columns 3 and 4 in Table 4 confirm this hypothesis.

Finally, we investigate whether cities with a more senior population respond more strongly to newspaper articles. The rationale is that senior citizens are more likely to rely on newspapers as their main source of information, and therefore more likely to read about the risks associated with medical careers. As educational choice is a collective household decision, we expect media persuasion effect to be stronger in cities with a higher percentage of senior people. We divide the sample based on a city's percentage of citizens aged over 60, the official retirement age. Results are reported in columns 5 and 6 in Table 4, and are consistent with intuition.

4.1.6 Robustness

In this section, we check whether our results are robust to a larger set of control variables, and to the use of alternative definitions of media attention.

In the baseline specification, our control variables are comprised of the logarithm of local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total number of industrial enterprises and their total output, the GDP growth rate, and the logarithm of the total number of students. These variables help to address the OVB problem. For example, the output of industrial enterprises proxies students' alternative career options in the manufacturing sector. In this subsection, we add a larger set of variables to control for more unobserved confounders.

The first set of extended controls includes the number of hospitals and doctors for each city. We add these controls to the baseline specification and present results in column 1 of Table A.2 in the Online Appendix. We did not include these controls in the baseline regression due to the large number of missing values for these two variables in the dataset (which make up around 20% of the sample). Nevertheless, the estimated coefficient has very similar magnitude and statistical significance as the baseline regression. In fact, all regressions in this paper are robust to controlling for these two variables.

The second set aims to control for media coverage by alternative sources. Apart from newspapers, television and the internet are important sources of information. However, there are no reliable data on the exact numbers of violence-related stories reported on television and the Internet. To partially address this concern, we use information on the possession of household appliances from the 2005 Urban Household Survey to compute the percentage of households with televisions, computers, and cell phones at the city level. We then multiply the 2005 penetration rates by the numbers of provincial anti-doctor violent events in each year, to proxy for the probability that citizens learn about these events through television and internet. We include these interaction terms in the regression and report estimation results in column 2 of Table A.2. The sample size decreases by about a half because our access to the UHS is restricted to 12 provinces, as discussed in the data section 3.3. Nonetheless, the estimated coefficient has a similar magnitude as the baseline regression, and remains statistically significant at the 1% level.

As suggested by Table 1, violent incidents had already been reported in over 20 cities by 2005, the first year of the college admission dataset. Therefore, these cities do not contribute to the event study in which we count the incidents of violence beginning from 2005. In the third column of Table A.2, we re-estimate the baseline regression without these 20 cities. We can see that the estimated coefficient is even larger than the baseline results, suggesting that this concern will not affect our conclusion.

In the fourth column of Table A.2, we extend our baseline regression to include a quadratic term of the number of the violent reports. This specification helps us to capture potential desensitization effect. If desensitization exists, the coefficient of the quadratic term should be negative. As column 4 shows, the estimated coefficient of the quadratic term is close to zero and statistically insignificant. Thus, we do not find evidence for desensitization.

The measure of media coverage is key to our analysis. In Section 4.1.2, we defined media coverage as the total number of violence-related articles published in the academic year in which students took the NCEE. This definition carries the implicit assumption that students only recall and respond to current year newspapers. However, historical coverage of anti-doctor violence could also affect students' choices of college major. Here, we account for news coverage in the past and define media attention as the discounted sum of newspaper articles on anti-doctor violence (this has a similar flavor to the cumulative sum of violent events in Section 4.1.4):

$$SumNewsArticle_{ijt} = \sum_{k>=0} \phi^k \times NewsArticle_{ijt-k}, \tag{4}$$

where ϕ is the discount factor ranging from 0 to 1. We reproduce the baseline regression using $SumNewsArticle_{ijt}$ as the key explanatory variable. Results with discount factor of 0.5 are shown in column 4 of Table A.2. We find that the coefficient of interest remains negative and statistically significant. The estimation is robust to a series of discount factors ranging from 0 to 1, which are not displayed for simplicity.

4.2 An Instrumental Variable Approach

A substantial part of our analysis using the generalized DID method aimed at demonstrating the causal effects of media persuasion. However, despite these efforts, endogeneity concerns may still arise. For example, there might be time-varying unobservables correlated with news articles and students' major choices, which could either over- or under-estimate the coefficient of interest. Moreover, potential measurement errors might exist and bias the estimates towards zero. In this section, we use an instrumental variable approach to further validate that our findings reflect causality.

4.2.1 Construction of the Instrumental Variable

Our instrumental variable approach borrows from a common strategy used in studies on media persuasion that exploits plausibly exogenous variations in media pressure gener-

ated by the presence of other newsworthy events (Eisensee and Strömberg 2007). The intuition is that the presence of newsworthy events will crowd out media coverage for other events, which in our case includes anti-doctor violence. As a result, newsworthy events will produce plausibly exogenous variations in the local newspapers' likelihood of reporting violence against doctors.

Studies have used the Olympic Games and natural disasters as newsworthy events (Eisensee and Strömberg 2007; Jetter 2017). We use the turnover of local political leaders, namely, city-level party secretaries and mayors. The logic is that during the years in which political turnovers occur, newspapers tend to increase their coverage of the newly appointed leaders, and report about their background, track record, and new policies. The newspapers also tend to publish articles about the achievements the departing leaders made in office. Thus, the turnover of local political leaders should be negatively correlated with the number of articles on violence against doctors in the local newspapers.

Building on this intuition, we follow Nunn and Qian (2014) and Sequeira et al. (2019) to construct a DID style instrument. The reason is that our dataset includes a sizable number of cities that have zero violent incidents, but still experience political turnover. This will weaken the explanatory power of the news pressure from political turnovers in our setting. Therefore, we add an additional dimension of variation, namely the pool of violent incidents that a city's newspapers can report, which we measure as the number of violent incidents in the city's province. Using the two sources of variation together, we interact the dummy variable indicating whether a city experiences a political turnover in a given year with the total number of violent events in the same province and year. The first-stage equation of our 2SLS framework can thus be written as:

$$NewsArticle_{ijt} = \alpha + \beta_1 Turnover_{ijt} \times Events_{jt} + \beta_2 Turnover_{ijt} + \gamma' X_{ijt} + \delta_i + \theta_{jt} + \epsilon_{ijt} \quad (5)$$

where $Turnover_{ijt}$ is a dummy variable that is equal to 1 if there is a turnover of mayor or party secretary in city i of province j in year t, and $Events_{jt}$ is the total number of violent events in province j in year t.

Note that both of the variables that comprise the interaction term appear in the first-stage equation, though the $Events_{jt}$ term is absorbed by the province-by-year fixed effect. β_2 captures the effect of turnover on news articles when there are no in-province violent events, and β_1 captures the differential effect of turnover when there is a positive number of violent events. If our identification strategy is valid, we should expect β_1 to be negative and statistically significant, and β_2 to be statistically indifferent from zero.

The second stage regression is equivalent to regressing the outcome variable on the predicted value of the main independent variable $NewsArticle_{ijt}$ in the first stage regression:

$$y_{ijt} = \alpha + \beta \widehat{NewsArticle}_{ijt} + \gamma' X_{ijt} + \delta_i + \theta_{jt} + \epsilon_{ijt}. \tag{6}$$

Our identification strategy relies on the interaction term being exogenous conditional on the baseline controls. There are a few potential concerns regarding the exclusion restrictions.

First, political turnover in a city may be correlated with economic conditions in the city, which can influence students' choices of major through channels other than news reports. A large body of literature has demonstrated that political turnovers in China depend on local economic outcomes and the personal connections with higher-level leaders (Li and Zhou 2005; Jia et al. 2015; Chen and Kung 2016; Landry et al. 2018). After controlling for a large set of variables related to the local economy, including the logarithm of local GDP, fiscal revenue and expenditure, the total number of industrial enterprises and their total output, and the GDP growth rate, we are left with variations in local politicians' connections to higher-level leaders. We argue that these variations are unlikely to be associated with factors influencing students' college enrollment decisions. Thus, conditional on our baseline controls, the turnovers of city mayors and party secretaries should be exogenous.

Second, it is possible that the number of incidents of anti-doctor violence in the whole province is not exogenous. To address this, we explicitly control for the number of local violent incidents. Thus, the number of events in all other cities within the same province should be uncorrelated with local unobserved conditions, given that the local events and province-by-year fixed effects are properly controlled for.

Third, though not a direct threat to the validity of the instrument, it is unclear how much variation in the instrument is accounted for by province-level incidents of violence and political turnover respectively. To answer this question, we describe the joint distribution of the province-level incidences of violence and the political turnovers. Among the 804 observations that experienced political turnover, 595 witnessed zero province-level event and 209 observations had one or more event. For the 1257 observations with no political turnover, 931 had zero violent events, and 326 had one or more event. Overall, both variables provide sizable variation for identification. The joint distribution table is omitted for brevity.

4.2.2 IV Results and Interpretation

The IV regression results are reported in Table 5. The first column shows a strong first stage relationship between the instrument and the number of news articles. As expected, the coefficient of the interaction term is negative and statistically significant at the 1% level. This suggests that, conditional on the pool of violent incidents that a city's newspapers can report, the presence of political turnover will reduce the number of published reports. Moreover, the excluded F statistic is 23.663 and larger than the commonly used 5% critical value (Stock and Yogo 2005), suggesting there is no weak instrument issue. We can also see that the coefficient of the uninteracted turnover dummy is not statistically different from zero. This is reassuring because political turnover should produce no effect on violence-related articles when the pool of potential reports is empty.

The remaining columns report the two-stage least squares (2SLS) results using the instrument specified above. Results using the baseline independent variable, the number of news articles related to anti-doctor violence, are shown in column 2. 2SLS estimates indicate that one more related article leads to a 1.99% decrease in the number of medical students. Estimates are statistically significant at the 1% level.

The 2SLS estimator is slightly larger than the OLS in terms of absolute value, suggesting measurement errors or downward OVB may exist. For example, China's 2009 health-care reform, which was rolled out at county and city levels, reduced the income from drug sales and the job satisfaction of Chinese doctors but increased patient satisfaction (Li et al. 2017). Therefore, the roll-out of the health-care reform may be negatively correlated with reports of anti-doctor violence and the attraction of medicine majors. This unobservable effect would not have been captured by province-by-year or city fixed effects in the baseline regression. This can cause the OLS estimator to be biased towards zero and the 2SLS estimator to be larger in absolute value, as we see from column 2. Nevertheless, given the similar estimation results produced by OLS and 2SLS, we perform the endogeneity test using a heteroskedasticity-robust version of the Durbin-Wu-Hausman test. The F statistic is 2.124 and fails to reject the null hypothesis that the number of articles is exogenous. This suggests that endogeneity is not a serious issue here and our baseline estimations also produce credible causal results.

Remaining columns repeat IV estimations for the other main OLS regressions. Results confirm that the 2SLS estimation conveys very similar information as the OLS. In column 3, we control for the log number of doctors and hospitals as in Section 4.1.6. In columns 4 to 6, we use the sub-sample consisting of high-literacy cities, highly-urbanized cities, and cities with more senior citizens, respectively. As shown in the heterogeneity Section 4.1.5, these cities contribute most to the effects. In column 7, we use the alternative mea-

sure of the independent variable, namely the discounted accumulated number of news articles. In column 8, we perform a falsification test using the number of arts students as the outcome variable, as in Section 4.1.3. As expected, all of these results share similar magnitude and significance as their OLS counterparts.

5 The Consequences of Discouraging Students

In the previous section, we showed that media persuasion exists, and that exposure to news articles about anti-doctor violence discourages students from studying medicine. In this section, we explore possible welfare implications by answering the following questions: which students are discouraged from studying medicine; which disciplines do deterred students choose instead; and how does this discouragement affect the quality of medical students?

5.1 Who Are Discouraged from Studying Medicine?

Baseline results from Section 4.1.2 capture the average responses of the student population to newspaper articles. To shed light on possible welfare implications of this discouragement, it is necessary to examine how media coverage affects different subgroups of students. In this subsection, we ask the following question: do students with different levels of ability respond differently to media coverage of anti-doctor violence? For example, if more capable students respond more strongly to media coverage, this will result in a larger welfare loss stemming from the reduction in the quality of the workforce in the health-care sector compared to if all students respond equally strongly.

We proxy for a student's ability in two ways: whether she is admitted to an elite college, and her NCEE score. First, we classify a college as elite if it is included in Project 211, and non-elite otherwise. This is a standard way of classifying elite colleges in the Chinese higher education system (Jia and Li 2017).[12]

We estimate the baseline equation for students enrolled in medicine-related majors at elite and non-elite colleges separately, and control for the log numbers of total admits at the respective colleges. Results are reported in Table 6, with columns 1 and 2 showing estimates for non-elite colleges, and columns 3 and 4 showing estimates for elite colleges.

[12]The Project 211 initiative was implemented by the Chinese Ministry of Education in 1995 with the aim of improving the quality of education and scientific research at the 116 colleges that met certain quality standards. These colleges have since received substantial support from the central government, invested heavily in research and student training, and as a result, become leading tertiary institutions in China. In other words, students who study at Project 211 colleges are among the brightest and most hard-working.

We can see that the estimated coefficient of interest is statistically significant across all regression specifications at both the elite and non-elite colleges. This suggests that both types of students respond to media coverage. However, the magnitudes of the estimates suggest that the persuasion effect is stronger for students at the elite colleges.

Second, we group students into four quartiles based on their NCEE scores. Note that the comparison is between students within the same province-year-track. Here, a higher quartile corresponds to higher NCEE scores, with students in quartile 1 having the lowest scores, and students in quartile 4 having the highest scores. We focus on students in the science track because biology and chemistry are prerequisites for studying medicine, and colleges normally allocate almost all of their admission quotas to students in the science track. We control for the log number of the total college enrollments in the science track in each quartile.

We estimate the baseline equation for students in each quartile separately and report the results in Table A.3 in the Online Appendix. We find that media persuasion effects are negative and statistically significant for students in the third and fourth quartiles, and are close to zero and statistically insignificant for students in the first and second quartiles. This finding confirms our hypothesis that more capable students respond more strongly to newspaper coverage of anti-doctor violence. Specifically, this finding suggests that media coverage of violence deters students with above-median scores from studying medicine-related majors.

This result renders two possible explanations: students with above-median scores have better alternative options, or they have a greater propensity to read newspapers and therefore learn about reports of violence against doctors. This result is also consistent with our findings in the elite versus non-elite decomposition. Moreover, since total enrollments at Project 211 colleges are considerably less than 25% of all college admits, the decrease in medicine enrollment at Project 211 colleges is entirely driven by the top students in quartile 4. The decrease in medicine enrollment at non-Project 211 colleges is driven by the remaining students in quartile 4 and the students in quartile 3.

5.2 Which Majors Do Discouraged Students Choose Instead?

We demonstrated that media coverage of violent events reduces student enrollment in medicine-related subjects, and that this decrease is driven by the responses of the relatively more capable students. A question that naturally follows is: which majors do these students choose instead of medicine? In this subsection, we answer this question by examining enrollment in other majors. We divide majors into five categories: pharmacy,

biology/chemistry, engineering, science (including math, physics, and other natural science majors), and economics/business. For simplicity, we only consider these common majors in the science track and skip humanities majors because most students who are eligible to choose medicine majors are from the science track. In light of the findings in the last subsection, we perform regressions separately for students at elite and non-elite colleges, due to heterogeneity in the responsiveness of these two groups of students.

Results presented in Table A.4 in the Online Appendix show that exposure to media coverage of anti-doctor violence induces students to choose majors that are somewhat related to medicine. We see an increase in the enrollment for pharmacy, biology/chemistry, engineering and science majors, but not for economics or business, which generates higher future returns but unrelated to medicine. These effects only exist for elite colleges.

5.3 Does the Quality of Medical Students Decrease?

In the above discussion, we demonstrated that media coverage of violence against doctors leads to a reduction in the number of students admitted to medicine majors at the city level, and that this decrease is driven by the responses of students with above-median test scores. However, this persuasion effect should be interpreted with care. Recall from Section 2.2 that Chinese colleges and the Ministry of Education collectively determine the allocation of enrollment quotas for each major and province, and that the quotas are almost always filled. This means that a spot that one student gives up will be taken by another student, leaving the total number of enrollments unchanged.[13] This leaves two caveats to interpreting the findings.

First, our estimated effects on the quantity of medicine students will be contaminated by spillovers across cities. Ideally, using students' preferences instead of admission outcomes can fix this issue, but unfortunately this information is not contained in our dataset. The second concern is that media coverage of anti-doctor violence simply shifts the composition of students enrolled in medicine majors in that a student with less exposure to media coverage can substitute for a student who becomes discouraged by more media coverage. As a result, welfare consequences will not be clear until we determine whether this substitution effect will induce an overall decrease in the quality of students enrolled in medicine majors.

To formally address these two concerns, we demonstrate that media coverage of anti-doctor violence leads to a decrease in the quality of medicine admits, even though the total number of medicine enrollment at the province level is determined by quotas. More

[13] As shown in Table 1, the total number and ratio of medicine students remain stable over time.

specifically, we test the following hypothesis: compared with students enrolled in non-medicine majors, medicine enrollment from provinces in which more violence-related articles are published will have lower admission rankings on average. We propose the following generalized triple-difference model to test this hypothesis:

$$y_{mjt} = \alpha + \beta Medicine_m \times NewsArticle_{jt} + \theta_{jt} + \eta_{mt} + \mu_{mj} + \epsilon_{mjt} \tag{7}$$

where y_{mjt} is a measure of the quality of students who took the NCEE in province j in academic year t, and are enrolled in major m in college; $NewsArticle_{jt}$ is the measure of media coverage in province j in academic year t; $Medicine_m$ indicates whether major m is a medicine major; and θ_{jt}, η_{mt}, and μ_{mj} capture the province-by-year, major-by-year, and province-by-major fixed effects, respectively. The uninteracted terms for medicine major $Medicine_m$ and the number of articles $NewsArticle_{jt}$ are absorbed by the province-by-year and the major-by-year fixed effects θ_{jt} and η_{mt} and do not explicitly appear in the model. This province-level regression is in line with the classic strategy proposed by Miguel and Kremer (2004) to address the concern of spillovers: to perform analysis at a higher level than the treatment level.

In this setup, our outcome and key explanatory variables are constructed as follows. For a given major, we use the average and the lowest NCEE rank percentiles for all enrollment from province j as proxies for the quality of enrollment from province j. The rank percentile is used because NCEE scores in different province-years are not directly comparable, and therefore must be normalized. In addition, $NewsArticle_{jt}$ is computed as the sum of all of the violence-related articles published in every city-level and province-level newspaper in province j in year t.

Similar to the DID framework, this triple-differences model also uses regional and temporal variation in $NewsArticle_{jt}$ to identify causality. The intuition is rather straightforward: a greater number of violence-related articles should produce a larger effect on the quality outcomes of medicine students (the "treatment group"), compared to non-medicine students (the "control group").

The construction of the triple-difference specification is as follows. We first examine the difference in enrollment rankings of medicine majors between higher-media-coverage provinces and lower-media-coverage provinces, which forms a standard generalized DID setting. We then examine this difference for non-medicine majors between those provinces. Finally, we take the difference between these differences for medicine and non-medicine majors, resulting in a triple-difference design. The key explanatory variable is the interaction term of the number of local and provincial newspaper articles on anti-doctor violence, and the dummy for medicine majors. Thus, the coefficient of interest here is β.

Results are presented in Table 7. For simplicity, we only examine the quality effect for the science track because most medicine majors only enroll students from the science track. We examine students at elite and non-elite colleges separately due to heterogeneity in the responsiveness between these two groups of students, as reported in Section 5.1. The dependent variable in the odd-numbered columns is the average percentile rank (within province and track) of the enrolled students for each major in every province; and the dependent variable in the even-numbered columns is the lowest percentile rank (within province and track) of the enrolled students for each major in every province. The latter can be interpreted as the cut-off rank.

Relative to non-medicine majors, a one-standard-deviation increase in the provincial-level reports of violent event (about 11.88) reduces the average enrollment ranking of medicine majors by 0.47%, and reduces the lowest enrollment scores by 0.34%. Both estimates are statistically significant. Multiplied by the average number of college enrollments from the science track in each province (around 70,000), this effect translates into a decrease of over 300 places in absolute rank within provinces.

In addition, the effects are not statistically significant for the non-elite colleges, as shown in columns 3 and 4. A natural explanation for the difference between the elite and non-elite colleges lies in their heterogeneous response to the newspaper reports on anti-doctor violence, as revealed in Table 6. All results in Table 6 are robust to controlling for province-major specific time trends, adding to our confidence that the estimates are not driven by confounders affecting both reporting as well as the desirability of medical careers.

6 Conclusion

In this paper, we examine how media coverage affects students' choices of college major. In particular, we focus on how newspaper coverage of anti-doctor violence in China affects students' decisions to study medicine in college. We first use a generalized DID strategy that exploits regional and temporal variation in media reports of these events. We find that each additional article leads to an approximately 0.6% decrease in the number of students studying health-care related majors. This effect is concentrated in the sub-disciplines in which physicians and nurses are trained. We perform several checks to ensure that the results are driven by media persuasion stemming from students' perception of violence-related articles, but not from the violent incidents themselves. For a more convincing identification strategy, we use an instrumental variable approach that exploits plausibly exogenous variation arising from other newsworthy events and province-wide

violent events. Using the interaction between local political turnover and the number of in-province events as the instrumental variable, we find similar results to the baseline DID model.

We further investigate the welfare implications of discouraging students from enrolling in medicine majors. We find that students attending elite colleges or have above-median test scores are more responsive to media coverage, and substitute towards related disciplines such as pharmacy, biology/chemistry, and engineering. In addition to the quantity effect, we find that exposure to newspaper articles on violent events reduces the average percentile rank and the cut-off percentile rank of medicine majors, which suggests that media coverage affects the quality of medicine majors. It should be noted that in China, an individual's educational path is not entirely a personal choice, and is jointly decided by the whole family. Therefore, the findings in this paper suggest that media exposure can change student families' perception of future career risks and affect the youth's educational choices.

This finding suggests that policies aimed at addressing anti-doctor violence and maintaining friendly working environments for medical staff are likely to produce positive effects in both the short and the long run. In contrast, subsidizing the hospital expenses in hiring security personnel and installing and maintaining surveillance cameras only yields immediate short-term results. Moreover, since a considerable fraction of anti-doctor violence results from medical disputes between patients and doctors, policies should be designed to improve the current liability and malpractice insurance for doctors and simplify the administrative process for dispute resolution.

References

BEAM, C. (2014): "Under the Knife - Why Chinese Patients Are Turning Against Their Doctors." *New Yorker*.

BO, S., J. LIU, J.-L. SHIU, Y. SONG, AND S. ZHOU (2019): "Admission mechanisms and the mismatch between colleges and students: Evidence from a large administrative dataset from China," *Economics of Education Review*, 68, 27–37.

BURNS, J. P. (1994): "Strengthening central CCP control of leadership selection: the 1990 nomenklatura," *The China Quarterly*, 138, 458–491.

CAMERON, A. C., J. B. GELBACH, AND D. L. MILLER (2011): "Robust Inference With Multiway Clustering," *Journal of Business and Economic Statistics*, 29, 238–249.

CHEN, S. (2015): "From Governance to Institutionalization: Political Selection from the Perspective of Central-local Relations in China–Past and Present (1368-2010)," *Department of Economics, Fudan University Working Paper*.

CHEN, T. AND J.-S. KUNG (2016): "Do land revenue windfalls create a political resource curse? Evidence from China," *Journal of Development Economics*, 123, 86–106.

CHEN, Y., H. LI, AND L.-A. ZHOU (2005): "Relative performance evaluation and the turnover of provincial leaders in China," *Economics Letters*, 88, 421–425.

CHOI, D., D. LOU, AND A. MUKHERJEE (2019): "The Effect of Superstar Firms on College Major Choice," *SSRN Working Paper*, 2845711.

CURRIE, J., W. LIN, AND J. MENG (2014): "Addressing Antibiotic Abuse in China: An Experimental Audit Study," *Journal of Development Economics*, 110, 39–51.

DAHL, G. B. AND S. DELLAVIGNA (2009): "Does Movie Violence Increase Violent Crime," *Quarterly Journal of Economics*, 124, 677–734.

DELAVANDE, A. AND B. ZAFAR (2019): "University choice: The role of expected earnings, non-pecuniary outcomes and financial constraints," *Journal of Political Economy*, forthcoming.

DELLAVIGNA, S. AND M. GENTZKOW (2010): "Persuasion: Empirical Evidence," *Annual Review of Economics*, 2, 643–669.

DELLAVIGNA, S. AND E. KAPLAN (2007): "The Fox News effect: Media Bias and Voting," *The Quarterly Journal of Economics*, 122, 1187–1234.

DELLAVIGNA, S. AND E. LA FERRARA (2015): "Economic and Social Impacts of the Media," in *Handbook of media economics*, Elsevier, vol. 1, 723–768.

DURANTE, R. AND E. ZHURAVSKAYA (2018): "Attack When the World Is Not Watching? US News and the Israeli-Palestinian Conflict," *Journal of Political Economy*, 126, 1085–1133.

EISENSEE, T. AND D. STRÖMBERG (2007): "News Droughts, News Floods, and U. S. Disaster Relief," *Quarterly Journal of Economics*, 122, 693–728.

ENIKOLOPOV, R., M. PETROVA, AND E. ZHURAVSKAYA (2011): "Media and Political Persuasion: Evidence from Russia," *The American Economic Review*, 101, 3253–3285.

FENG, S., Y. HU, AND R. MOFFITT (2017): "Long run trends in unemployment and labor force participation in urban China," *Journal of Comparative Economics*, 45, 304–324.

FISMAN, R., J. SHI, Y. WANG, AND W. WU (2019): "Social Ties and the Selection of China's Political Elite," Tech. rep., Boston University-Department of Economics.

FRICKE, H., J. GROGGER, AND A. STEINMAYR (2018): "Exposure to Academic Fields and College Major Choice," *Economics of Education Review*, 64, 199–213.

GE, S. AND D. T. YANG (2014): "Changes in China's wage structure," *Journal of the European Economic Association*, 12, 300–336.

GENTZKOW, M. AND J. M. SHAPIRO (2008): "Preschool television viewing and adolescent test scores: Historical evidence from the Coleman study," *The Quarterly Journal of Economics*, 123, 279–323.

HESKETH, T. AND W. X. ZHU (1997): "Health in China: The Healthcare Market," *BMJ*, 314, 1616.

JACOBSEN, G. D. (2011): "The Al Gore effect: An Inconvenient Truth and Voluntary Carbon Offsets," *Journal of Environmental Economics and Management*, 61, 67–78.

JENSEN, R. (2010): "The (perceived) returns to education and the demand for schooling," *The Quarterly Journal of Economics*, 125, 515–548.

JETTER, M. (2017): "The Effect of Media Attention on Terrorism," *Journal of Public Economics*, 153, 32–48.

JIA, R., M. KUDAMATSU, AND D. SEIM (2015): "Political selection in China: The complementary roles of connections and performance," *Journal of the European Economic Association*, 13, 631–668.

JIA, R. AND H. LI (2017): "Access to Elite Education, Wage Premium, and Social Mobility: The Truth and Illusion of China's College Entrance Exam," *Stanford Center for International Development, Working Paper*.

KAUFMANN, K. M. (2014): "Understanding the income gradient in college attendance in Mexico: The role of heterogeneity in expected returns," *Quantitative Economics*, 5, 583–630.

LANCET, T. (2010): "Chinese Doctors Are Under Threat," *The Lancet*, 376, 657.

LANDRY, P. F. (2008): *Decentralized Authoritarianism in China: the Communist Party's control of local elites in the post-Mao era*, Cambridge University Press New York.

LANDRY, P. F., X. LÜ, AND H. DUAN (2018): "Does performance matter? Evaluating political selection along the Chinese administrative ladder," *Comparative Political Studies*, 51, 1074–1105.

LI, H., K. LI, AND J. WU (2018): "Local Political Chief Turnover and Economic Growth: Evidence from China," *SSRN Working Paper*, 3223280.

LI, H. AND L.-A. ZHOU (2005): "Political turnover and economic performance: the incentive role of personnel control in China," *Journal of Public Economics*, 89, 1743–1762.

LI, X., J. LU, S. HU, K. CHENG, J. DE MAESENEER, Q. MENG, E. MOSSIALOS, D. R. XU, W. YIP, H. ZHANG, ET AL. (2017): "The primary health-care system in China," *The Lancet*, 390, 2584–2594.

LIM, C. S. H., J. M. SNYDER, AND D. STRÖMBERG (2015): "The Judge, the Politician, and the Press: Newspaper Coverage and Criminal Sentencing across Electoral Systems †," *American Economic Journal: Applied Economics*, 7, 103–135.

LIU, X., Y. LIU, AND N. CHEN (2000): "The Chinese Experience of Hospital Price Regulation." *Health Policy and Planning*, 15, 157–163.

MANSKI, C. F. (1993): "Adolescent econometricians: How do youth infer the returns to schooling?" in *Studies of Supply and Demand in Higher Education*, University of Chicago Press, 43–60.

MENG, X. (2012): "Labor market outcomes and reforms in China," *Journal of Economic Perspectives*, 26, 75–102.

MEYER, D., V. C. SHIH, AND J. LEE (2016): "Factions of different stripes: gauging the recruitment logics of factions in the reform period," *Journal of East Asian Studies*, 16, 43–60.

MIGUEL, E. AND M. KREMER (2004): "Worms: identifying impacts on education and health in the presence of treatment externalities," *Econometrica*, 72, 159–217.

NUNN, N. AND N. QIAN (2014): "US food aid and civil conflict," *American Economic Review*, 104, 1630–66.

QIN, B. (2013): "Chinese Microblogs and Drug Quality," *Institute for International Economic Studies working paper*.

QIN, B., D. STRÖMBERG, AND Y. WU (2018): "Media Bias in China," *American Economic Review*, 108, 2442–2476.

SEQUEIRA, S., N. NUNN, AND N. QIAN (2019): "Immigrants and the Making of America," *Review of Economic Studies*, forthcoming.

SHIH, V., C. ADOLPH, AND M. LIU (2012): "Getting ahead in the communist party: explaining the advancement of central committee members in China," *American Political Science Review*, 106, 166–187.

STINEBRICKNER, R. AND T. R. STINEBRICKNER (2013): "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout," *Review of Economic Studies*, 81, 426–472.

STOCK, J. H. AND M. YOGO (2005): "Testing for weak instruments in Linear IV regression," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge University Press, 80–108.

STRÖMBERG, D. (2004): "Radio's Impact on Public Spending," *Quarterly Journal of Economics*, 119, 189–221.

——— (2015): "Media and Politics," *Annual Review of Economics*, 7, 173–205.

WISWALL, M. AND B. ZAFAR (2014): "Determinants of College Major Choice: Identification using an Information Experiment," *Review of Economic Studies*, 82, 791–824.

WU, D., Y. WANG, K. F. LAM, AND T. HESKETH (2014): "Health System Reforms, Violence Against Doctors and Job Satisfaction in the Medical Profession: a Cross-Sectional Survey in Zhejiang Province, Eastern China," *BMJ Open*, 4.

XU, C. (2011): "The fundamental institutions of China's reforms and development," *Journal of economic literature*, 49, 1076–1151.

XU, W. AND S. ZHANG (2014): "Chinese Pediatricians Face a Crisis: Should They Stay or Leave?" *Pediatrics*, 134, 1045–1047.

YANAGIZAWA-DROTT, D. (2014): "Propaganda and Conflict: Evidence from the Rwandan Genocide," *The Quarterly Journal of Economics*, 129, 1947–1994.

YAO, Y. AND M. ZHANG (2015): "Subnational leaders and economic growth: evidence from Chinese cities," *Journal of Economic Growth*, 20, 405–436.

ZAFAR, B. (2011): "How do College Students Form Expectations?" *Journal of Labor Economics*, 29, 301–348.

ZHAO, Y. (1998): "Media, Market, and Democracy in China: Between the Party Line and the Bottom Line," *China Journal*, 72, 174.

Table 1: Summary Statistics for the News and College Admission Dataset

Panel A: News Dataset

Year	Number of Cities	Mean	Std. Dev.	Min	Max	Number of Cities with Zero Anti-doctor News
2001	337	0.0237	0.2548	0	4	333
2002	337	0.2582	1.3122	0	11	316
2003	337	0.0089	0.0941	0	1	334
2004	337	0.1276	0.6490	0	6	321
2005	337	0.0297	0.2871	0	4	332
2006	337	0.0445	0.3196	0	4	328
2007	337	0.2582	1.9430	0	23	325
2008	337	0.0474	0.3503	0	4	328
2009	337	0.3858	1.4329	0	15	290
2010	337	0.7418	3.0171	0	45	278
2011	337	0.3205	1.1642	0	15	314

Panel B: Administrative Dataset on College Admission

Year	Number of Provinces	Number of Students	Number of Students in Medicine-Related Majors	Ratio of Students in Medicine-Related Majors
2005	30	2,064,052	130,373	0.0632
2006	30	2,212,660	149,883	0.0677
2007	30	2,419,560	159,312	0.0658
2008	30	2,617,349	171,866	0.0657
2009	29	2,564,988	174,550	0.0681
2010	29	2,738,889	183,923	0.0672
2011	28	2,690,016	182,768	0.0679

Note: Panel A presents summary statistics on the number of newspaper articles on anti-doctor violence. Panel B provides basic description of the administrative dataset on college admission.

Table 2: Effect of Media Attention on Students' Choice of Medicine Majors

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Log # of Students in All Medicine Majors				Log # of Students in Clinical/Nursing	
# of News Articles	-0.0052***	-0.0063***			-0.0096***	-0.0090***
	(0.0016)	(0.0019)			(0.0019)	(0.0028)
I(# of News Articles>0)			-0.0250*			
			(0.0147)			
I(Post 1st News Article)				-0.0455***		
				(0.0136)		
Observations	2,249	2,061	2,061	2,061	2,249	2,061
R-squared	0.9848	0.9868	0.9868	0.9868	0.9738	0.9807
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Prov$\times$Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	No	Yes
Level Mean	491.0	491.0	491.0	491.0	240.3	240.3
Level SD	360.2	360.2	360.2	360.2	183.9	183.9

Note: *** denotes significance at 0.01, ** at 0.05, and * at 0.1. The outcome variables are the log number of students admitted to all medicine majors and clinical or nursing majors at city level. The main independent variable is the number of local newspaper reports on violent events against doctors. Baseline control variables include the logarithm of the total number of college admits, local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total output and the total number of industrial enterprises, and the GDP growth rate. Column 3 uses a dummy variable indicating whether there is any violent-related news articles as the main independent variable; column 4 uses a dummy variable indicating whether the period is after the first violent-related news article in the city; column 5 and 6 use the sub-disciplines of clinical or nursing as the outcome variables. Robust standard errors are two-way clustered at the province and year level.

Table 3: Perception of News Articles

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Log Number of Students in All Medicine Majors					
Sample	0-event obs	0-event cities	Full	Full	Full	Full
# of News Articles	-0.0094*	-0.0236**	-0.0044**	-0.0073***		
	(0.0052)	(0.0102)	(0.0020)	(0.0021)		
# of Negative-sentiment Articles					-0.0082**	
					(0.0032)	
# of Non-negative-sentiment Articles					-0.0049	
					(0.0076)	
# of Front-page Articles						-0.0078**
						(0.0037)
# of Non-front-page Articles						-0.0065
						(0.0045)
Cumulative # of Events			-0.0140**			
			(0.0059)			
Discounted Sum of Events: ϕ=0.5				0.0116	0.0118	0.0116
				(0.0100)	(0.0100)	(0.0100)
Observations	2,008	1,468	2,061	2,061	2,061	2,061
R-squared	0.9867	0.9858	0.9868	0.9868	0.9868	0.9868
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Prov$\times$Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Level Mean	482.0	414.5	491.0	491.0	491.0	491.0
Level SD	355.3	319.0	360.2	360.2	360.2	360.2

Note: *** denotes significance at 0.01, ** at 0.05, and * at 0.1. The outcome variables are the log number of students admitted to all medicine majors at city level. The main independent variable is the number of local newspaper reports on violent events against doctors. Baseline control variables include the logarithm of the total number of college admits, local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total output and the total number of industrial enterprises, and the GDP growth rate. Column 1 limits sample to city-year pairs with zero violent event occurring; column 2 limits sample to cities with zero violent event occurring in all years; column 3 controls for accumulated number of local violent events; column 4 controls for accumulated number of local violent events discounted at a rate of 0.5; column 5 counts the number of news articles by negative sentiment and non-negative sentiment; column 6 counts the number of news articles by front-page and non-front-page. City and province-by-year fixed effects are included in all regressions. Robust standard errors are two-way clustered at the province and year level.

Table 4: Heterogeneity in the Effect of Media Attention on Students' Choice of Medicine Majors

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Log Number of Students in All Medicine Majors					
Sample	High Literacy	Low Literacy	High Urbanized	Low Urbanized	High Aging	Low Aging
# of News Articles	-0.0059***	-0.0144	-0.0046*	0.0009	-0.0142**	-0.0034
	(0.0020)	(0.0102)	(0.0024)	(0.0096)	(0.0065)	(0.0028)
Observations	1,260	781	749	1,187	1,052	974
R-squared	0.9889	0.9863	0.9903	0.9880	0.9845	0.9880
City FE	Yes	Yes	Yes	Yes	Yes	Yes
Prov×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Local Events	Yes	Yes	Yes	Yes	Yes	Yes
Level Mean	533.6	424.6	504.0	483.0	554.5	427.9
Level SD	364.5	343.2	376.1	349.8	347.1	362.1

Note: *** denotes significance at 0.01, ** at 0.05, and * at 0.1. The outcome variables are the log number of students admitted to all medicine majors at city level. The main independent variable is the number of local newspaper reports on violent events against doctors. Column 1 and 2 divide the sample by high and low literacy rate; column 3 and 4 divide the sample by high and low urbanization rate; column 5 and 6 divide the sample by high and low aging rate. Baseline control variables include the logarithm of the total number of college admits, local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total output and the total number of industrial enterprises, and the GDP growth rate. City and province-by-year fixed effects and discounted accumulated number of local anti-doctor events are included in all regressions. Robust standard errors are two-way clustered at the province and year level.

Table 5: Instrumental Variable Estimation of the Effect of Media Attention

VARIABLES	(1) # of Articles	(2)	(3)	(4)	(5)	(6)	(7)	(8) Art
		Log Number of Students in All Medical Majors						
Models	1st Stage	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Sample	Full	Full	Full	High Literacy/Urbanized/Aging			Full	Full
Turnover×Provincial events	-0.1145***							
	(0.0234)							
Turnover	0.0233							
	(0.0412)							
# of News Articles		-0.0199**	-0.0182***	-0.0173**	-0.0152***	-0.0687*		0.0036
		(0.0088)	(0.0063)	(0.0086)	(0.0048)	(0.0350)		(0.0082)
Accumu. # of Articles: ϕ=0.5							-0.0164**	
							(0.0078)	
1st stage excluded F Stat	23.663							
Observations	2,061	2,061	1,681	1,260	749	1,052	2,061	2,061
R-squared	0.6427	0.9867	0.9845	0.9888	0.9902	0.9828	0.9868	0.9914
City FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prov×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
# of Doctors & Hospitals	No	No	Yes	No	No	No	No	No
Level Mean	0.2	491.0	491.0	533.6	504.0	554.5	491.0	843.9
Level SD	1.5	360.2	360.2	364.5	376.1	347.1	360.2	609.4

Note: *** denotes significance at 0.01, ** at 0.05, and * at 0.1. The outcome variables are the log number of students admitted to all medical majors at city level. The main independent variable is the number of local newspaper reports on violent events against doctors. The first column shows the 1st stage results. The rest columns are 2SLS results using IV: # of provincial anti-doctor events × whether a city-leader-transition year. Cragg-Donald Wald F statistics are reported. Baseline control variables include the logarithm of the total number of college admits, local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total output and the total number of industrial enterprises, and the GDP growth rate. City and province-by-year fixed effects and discounted accumulated number of local anti-doctor events are included in all regressions. Robust standard errors are two-way clustered at the province and year level.

Table 6: Effect of Media Attention on Students' Choice of Medicine Majors by College Type

	(1)	(2)	(3)	(4)
VARIABLES	Log Number of Students by College Type			
College Type	Non-elite Colleges		Elite Colleges	
# of News Articles	-0.0047***	-0.0072***	-0.0070**	-0.0075**
	(0.0016)	(0.0021)	(0.0027)	(0.0034)
Observations	2,249	2,061	2,249	2,061
R-squared	0.9836	0.9853	0.9238	0.9251
City FE	Yes	Yes	Yes	Yes
Prov×Year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
Local Events	Yes	Yes	Yes	Yes
Level Mean	445.2	445.2	45.9	45.9
Level SD	334.1	334.1	44.3	44.3

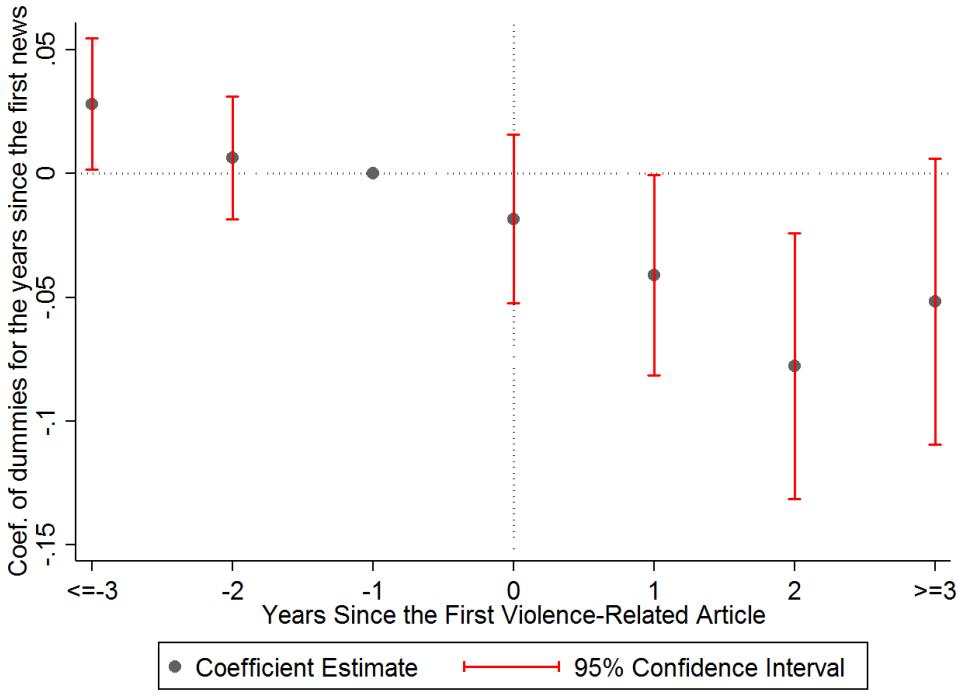
Note: *** denotes significance at 0.01, ** at 0.05, and * at 0.1. The outcome variables are the log number of students admitted to all medical majors at city level, grouped by elite colleges and non-elite colleges. The main independent variable is the number of local newspaper reports on violent events against doctors . Baseline control variables include the logarithm of the total number of college admits, local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total output and the total number of industrial enterprises, and the GDP growth rate. City and province-by-year fixed effects and discounted accumulated number of local anti-doctor events are included in all regressions. Robust standard errors are two-way clustered at the province and year level.

Table 7: Effect of Media Attention on Quality of Students Choosing Medicine Majors

	(1)	(2)	(3)	(4)
VARIABLES	mean rank (%)	min rank (%)	mean rank (%)	min rank (%)
Sample	Elite colleges		Non-elite colleges	
Medicine Major × # of Articles	-0.0396***	-0.0286***	0.0203	0.0250
	(0.0137)	(0.0097)	(0.0127)	(0.0182)
Observation	24,779	24,779	128,390	128,390
R-squared	0.1506	0.1320	0.0746	0.0960
Prov×Year FE	Yes	Yes	Yes	Yes
Major×Year FE	Yes	Yes	Yes	Yes
Prov×Major FE	Yes	Yes	Yes	Yes

Note: *** denotes significance at 0.01, ** at 0.05, and * at 0.1. The outcome variables are the mean and minimum percentile rankings of students admitted to a major in a province. The main independent variable is the interaction of the number of local newspaper reports on violent events against doctors in a province and the medicine major dummy. Province-by-year, major-by-year and province-by-major fixed effects are included in all regressions. Robust standard errors are two-way clustered at the province and year level.

Figure 1: Testing for Parallel Pre-Trend



Notes: This figure reports estimates of the dynamic effects of violence-related articles derived from a flexible DID specification, as shown in Equation 2, to test the parallel trend assumption. The omitted category is the year before the appearance of the first article.

Additional Figures and Tables

Table A.1: Falsification Test

VARIABLES	(1) Log # of Students in Art Majors	(2) Log # of Students in Art Majors	(3) All Medicine	(4) Clinical/Nursing
# of News Articles	-0.0007	0.0005		
	(0.0011)	(0.0013)		
# of News Articles, $t+1$			-0.0019	-0.0016
			(0.0027)	(0.0024)
Observations	2,249	2,061	1,841	1,841
R-squared	0.99151	0.99143	0.9879	0.9822
City FE	Yes	Yes	Yes	Yes
Prov$\times$Year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes
Level Mean	843.9	843.9	491.0	240.3
Level SD	609.4	609.4	360.2	183.9

Note: *** denotes significance at 0.01, ** at 0.05, and * at 0.1. In Columns 1 and 2, the outcome variable is the log number of students admitted to art majors at city level; and the main independent variable is the number of local newspaper reports on anti-doctor violence. In Columns 3 and 4, the outcome variables are the log number of all medicine admits and clinical medicine and nursing admits; and the main independent variable is the one-year lead number of local newspaper reports on anti-doctor violence. Baseline control variables include the logarithm of the total number of college admits, local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total output and the total number of industrial enterprises, and the GDP growth rate. City and province-by-year fixed effects are included in all regressions. Robust standard errors are two-way clustered at the province and year level.

Table A.2: Robustness: More Control Variables and Alternative Measures of Media Attention

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Log Number of Students in All Medical Majors				
# of News Articles	-0.0075***	-0.0081***	-0.0158**	-0.0088**	
	(0.0022)	(0.0024)	(0.0074)	(0.0041)	
(# of News Articles)^2				0.0000	
				(0.0001)	
Cumulative # of Articles: ϕ=0.5					-0.0077***
					(0.0019)
Observations	1,681	997	1,897	2,061	2,061
R-squared	0.9847	0.9834	0.9866	0.9868	0.9868
City FE	Yes	Yes	Yes	Yes	Yes
Prov$\times$Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Local Events	Yes	Yes	Yes	Yes	Yes
# of Doctors & Hospitals	Yes	No	No	No	No
Other Media $\times$ Provincial Events	No	Yes	No	No	No
Level Mean	564.4	563.2	465.4	491.0	491.0
Level SD	355.3	357.5	345.1	360.2	360.2

Note: *** denotes significance at 0.01, ** at 0.05, and * at 0.1. The outcome variables are the log number of students admitted to all medical majors at city level. The main independent variable is the number of local newspaper reports on violent events against doctors or the discounted stock of the reports. Baseline control variables include the logarithm of the total number of college admits, local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total output and the total number of industrial enterprises, and the GDP growth rate. Column 1 controls for the logarithm of the total number of doctors and the total number of hospitals; column 2 controls for the interaction between the total number of anti-doctor violent events in the same province and the penetration rate of television, computer and mobile phones respectively; column 3 drops all cities with any violent-related news articles before 2005; column 4 adds the square of the number of local newspaper reports; column 5 uses discounted accumulated number of past violent-related news articles at a rate of 0.5 as the main independent variable. City and province-by-year fixed effects are included in all regressions. Robust standard errors are two-way clustered at the province and year level.

Table A.3: Effect of Media Attention on Students' Choice of Medicine Majors by Score Quartile

VARIABLES	Log Number of Students by Score Quartile			
Quartiles	1	2	3	4
# of News Articles	0.0030	-0.0028	-0.0089***	-0.0075**
	(0.00352)	(0.00409)	(0.0023)	(0.0033)
Observations	2,036	2,036	2,036	2,036
R-squared	0.9508	0.9610	0.9629	0.9540
City FE	Yes	Yes	Yes	Yes
Prov×Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Level Mean	88.3	128.5	117.6	89.3
Level SD	81.4	104.2	97.1	79.6

Note: *** denotes significance at 0.01, ** at 0.05, and * at 0.1. The outcome variables are the log number of students admitted to all medical majors, divided by score quartiles. The main independent variable is the number of local newspaper reports on violent events against doctors. Baseline control variables include the logarithm of the total number of college admits, local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total output and the total number of industrial enterprises, and the GDP growth rate. City and province-by-year fixed effects and discounted accumulated number of local anti-doctor events are included in all regressions. Robust standard errors are two-way clustered at the province and year level.

Table A.4: Alternative Majors

Panel A: Log Number of Students in Alternative Majors (Elite Colleges)

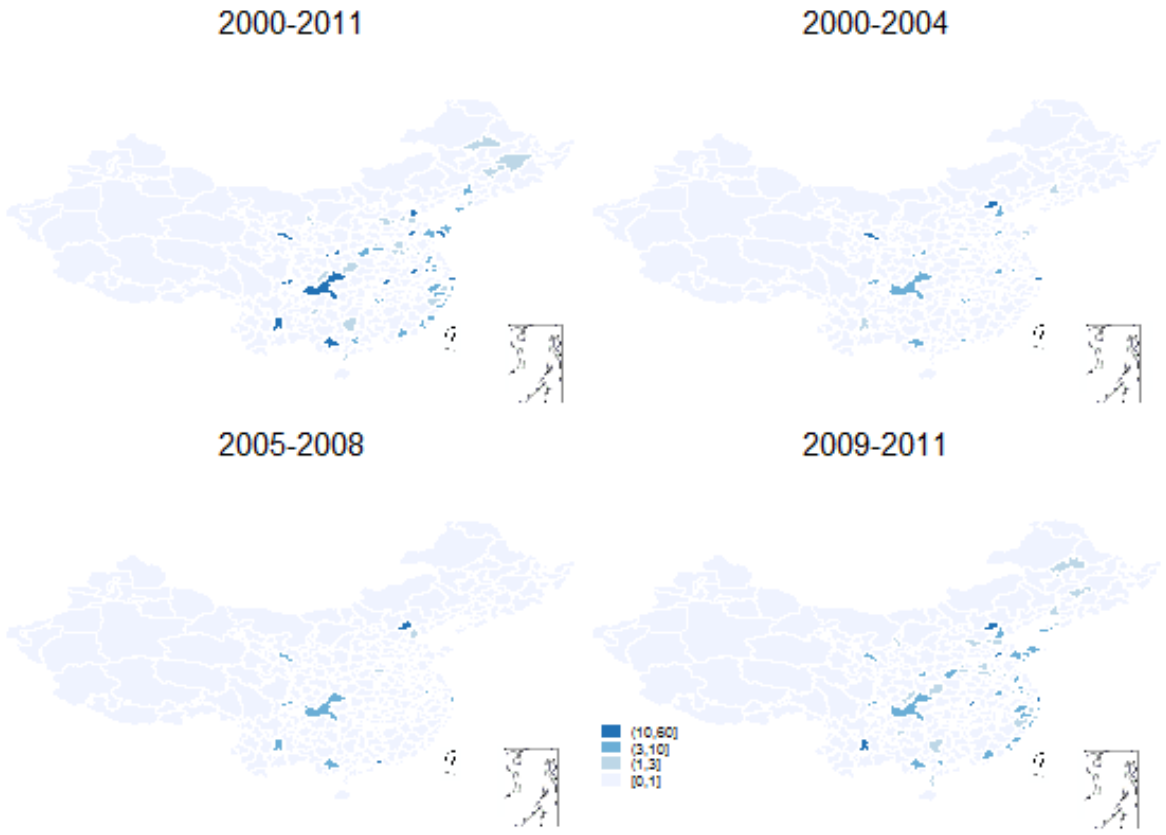
	(1)	(2)	(3)	(4)	(5)
Majors	Pharmacy	Bio/Chem	Engineering	Other Science	Econ/Business
# of News Articles	0.0076*	0.0092***	0.0084***	0.0071***	0.0070
	(0.0040)	(0.0030)	(0.0029)	(0.0026)	(0.0070)
Observations	2,061	2,061	2,061	2,061	2,061
R-squared	0.8988	0.9501	0.9788	0.9863	0.9366
City FE	Yes	Yes	Yes	Yes	Yes
Prov×Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Local Events	Yes	Yes	Yes	Yes	Yes
Level Mean	30.0	74.4	164.0	486.4	58.3
Level SD	25.2	60.3	135.6	412.4	57.8

Panel B: Log Number of Students in Alternative Majors (Non-elite Colleges)

	(1)	(2)	(3)	(4)	(5)
Majors	Pharmacy	Bio/Chem	Engineering	Other Science	Econ/Business
# of News Articles	-0.0019	-0.0008	-0.0004	-0.0006	0.0015
	(0.0037)	(0.0030)	(0.0023)	(0.0008)	(0.0020)
Observations	2,061	2,061	2,061	2,061	2,061
R-squared	0.9732	0.9887	0.9915	0.9975	0.9914
City FE	Yes	Yes	Yes	Yes	Yes
Prov×Year FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Local Events	Yes	Yes	Yes	Yes	Yes
Level Mean	162.6	357.0	710.4	2012.0	406.8
Level SD	137.7	281.1	525.3	1632.0	346

Note: *** denotes significance at 0.01, ** at 0.05, and * at 0.1. The outcome variables are the log number of students admitted to various alternative majors at city level. The main independent variable is the number of local newspaper reports on violent events against doctors. Baseline control variables include the logarithm of the total number of college admits, local GDP, total population, rural population, the number of hospital beds, fiscal revenue and expenditure, the total output and the total number of industrial enterprises, and the GDP growth rate. City and province-by-year fixed effects and discounted accumulated number of local anti-doctor events are included in all regressions. Robust standard errors are two-way clustered at the province and year level.

Figure A.1: Geographical Distribution of Newspaper Articles on Violence against Doctors



Notes: This figure displays the geographical distribution of the number of newspaper articles on anti-doctor violence. The map on the top left-hand-corner corresponds to all years from 2000 to 2011. The map on the top right-hand-corner corresponds to years 2000 to 2004, which is prior to the college admissions data. The two maps on the bottom correspond to the years 2005 to 2008, and 2009 to 2011 respectively.