

NBER WORKING PAPER SERIES

MALPRACTICE REFORM AND THE SORTING OF
NEW PHYSICIANS BY MEDICAL HUMAN CAPITAL

Pinka Chatterji
Siyang Li
Gerald R. Marschke

Working Paper 24401
<http://www.nber.org/papers/w24401>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2018, Revised March 2020

The authors sincerely thank Dr. David Armstrong at the University at Albany Center for Health Workforce Studies for providing access to the Survey of Residents Completing Training in New York State. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Center for Health Workforce Studies or the National Bureau of Economic Research. All errors are the responsibility of the authors

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Pinka Chatterji, Siyang Li, and Gerald R. Marschke. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Malpractice Reform and the Sorting of New Physicians by Medical Human Capital
Pinka Chatterji, Siyang Li, and Gerald R. Marschke
NBER Working Paper No. 24401
March 2018, Revised March 2020
JEL No. I1

ABSTRACT

We test whether state malpractice reforms differentially attract physicians whose human capital attributes may predispose them towards higher-than-average malpractice risk and lower quality patient care. Using an exit survey of physicians completing residencies between 1998 and 2017, we estimate willingness-to-pay to locate their first practice in a malpractice-reformed state. We find physicians are willing to forego on average about \$11 in hourly wages to locate in a reform state. Training in a high vs. low-risk specialty, graduating from a less vs. more selective medical school, and training at a low vs. higher-ranked teaching hospital increases willingness-to-pay to locate in a reform state by \$18 to \$24 per hour. We argue that the generally strong human capital-bias in physician sorting responses to litigation reform may play a role in the geographic variation in patient care documented in the health literature.

Pinka Chatterji
State University of New York at Albany
Economics Department
1400 Washington Avenue
Albany, NY 12222
and NBER
pchatterji@albany.edu

Gerald R. Marschke
State University of New York at Albany
Economics Department
1400 Washington Avenue
Albany, NY 12222
and NBER
gerald.marschke@gmail.com

Siyang Li
State University of New York at Albany
Economics Department
1400 Washington Avenue
Albany, NY 12222
sli23@albany.edu

1. Introduction and Background

Since the 1970's, almost all states in the US have passed tort reforms intended to reduce unnecessary litigation and excessive settlements related to medical malpractice. Some of the most common and widely studied tort reforms are state caps on damage payments, which limit the dollar amount of damages paid to successful plaintiffs in malpractice suits (Currie & MacLeod, 2008). Damage caps can be applied to total, non-economic, or punitive damages, and they appear to have reduced the “malpractice pressure” that physicians face. Prior research indicates that these laws are associated with declines in the number of lawsuits,¹ reductions in the value of awards made, and decreases in physicians' malpractice premiums (Danzon, Epstein, and Johnson, 2004; Kilgore, Morrissey, and Nelson, 2006; Avraham, 2007; Sloan & Chepke, 2008; Born, Viscusi, and Baker, 2009; Seabury, Helland, and Jena 2014). The concept of “malpractice pressure,” however, includes not only malpractice premiums but also psychic costs, time costs, and reputational harm (Currie & MacLeod, 2008). Thus, if tort reforms reduce the likelihood of lawsuits against physicians, these reforms may affect physician behaviors immediately, since changes in malpractice premiums may not be the only mechanism.

If state-level malpractice reforms are successful in reducing malpractice pressure on physicians, there are likely to be three kinds of effects on physicians. First, physicians may respond by taking less care in treating patients or using riskier treatments (moral hazard), or by accepting higher risk patients, since their malpractice liability is reduced (Kessler & McClellan, 1996; Currie & MacLeod, 2008). They may practice less *defensively*, prescribing fewer tests or treatments that have little medical value to the patient but protect the physician from a medical malpractice tort action. Second, enactment of these reforms may affect physician supply to a state, if physicians are more likely to locate in states with reformed malpractice laws. Third, enactment of malpractice laws may lead to a re-sorting of physicians

¹ Patients may be less likely to file malpractice claims in states that have passed reforms, since these claims are less likely to be successful and yield worthwhile payouts.

across states by physician type. In this paper, we test whether malpractice laws affect physician supply, but we focus on this third possible effect – the effect of malpractice reforms on physicians’ sorting across U.S. states along aspects of physician human capital that may be related to the degree of malpractice pressure and may predict patient care quality.

There is mixed evidence that state malpractice reforms affect the aggregate supply of physicians in a state. Most studies in this area utilize state-level and county-level yearly counts of physicians from the American Medical Association’s Physician Masterfiles. Recent years of these data are available in the US DHHS Area Health Resource File (AHRF). Kessler et al. (2005) using state-level data from 1985 to 2001 find that “direct” malpractice reforms (defined as passage of any law capping damage awards, removing punitive damages or mandatory pre-judgment interest, or reforming the collateral source rule)² are associated with greater growth in the aggregate supply of physicians. They report stronger effects on physicians more likely to face malpractice pressure, including physicians not working in group practices and physicians in specialties with high malpractice premiums. Malani & Reif (2015) use state-year data from 1980-2001 and find that state punitive damage caps increase physician supply by 2-6 percent, depending on their specification. They also find evidence that physician supply increases prior to punitive damage caps becoming effective, suggesting that anticipation effects may be important.

Other studies find that the effects of malpractice reforms on physician labor supply tend to be concentrated among physicians in specialties that face the highest malpractice risk. Klick & Stratmann (2007), for example, using state-level data from 1980-2001, find that caps on damages are associated with increases in the supply of physicians working in specialties facing the highest average medical malpractice award per doctor. Matsa (2007), using state-level data from 1970-2000, reports that caps on damages increase the supply of specialist physicians working in rural areas, arguing that this is because

²Pre-judgment interest compensates the plaintiff for the delay in receiving the award between the time of injury and time of judgment. Appendix Table 1 shows definitions of some other commonly studied malpractice reforms (including the collateral source rule) in the economics literature.

rural doctors face greater uninsured litigation costs, as well as more elastic demand. Chou & Lo Sasso (2009), using data on graduating residents from New York State from 1998-2003 (the same survey used in the present study), find that new surgeons tend to locate in states with laws that cap malpractice damages, but this is not true for primary care physicians and OB/GYNs. Pesko et al. (2017) use state-level counts of physicians from 2000-2011 and an event-study design, using states that previously adopted damage caps as the control group.³ They find that non-economic damage caps enacted between 2003 and 2006 increased the supply of physicians younger than 35 years old working in high-risk specialties. Helland and Seabury (2015) contain a recent review of the literature on reform and the supply of high-risk specialties and provide additional evidence of a positive effect. On the other hand, two papers since the Helland and Seabury review have found no added effect of reform on high-risk specialties. Paik et al. (2016) use state-level and county-level data from 1992-2011 and, across a variety of control groups, find that damage cap laws that were passed between 2002 and 2005 did not affect physician supply overall and of the high-risk specialties (plastic surgeons were an exception). Hyman et al. (2015) report the same null finding specifically for the damages cap law that went into effect in Texas in 2004.

Lieber (2014) focuses on the idea that state malpractice reforms may affect not just physician supply but also geographic sorting of physicians by individual characteristics. Using county-level data, Lieber finds that when a neighboring state enacts a law placing caps on non-economic damages there is both a 4 percent fall in the supply of physicians and a 4 percent decline in the state's malpractice rate. By focusing on effects of a neighboring state's malpractice reforms, one can isolate the effect on the malpractice rate caused by physician sorting from the effect caused by a change in physician behavior

³ If damage cap laws induce physicians to move from states without laws to states with laws, traditional difference-in-difference methods may "double-count" the physicians who move in response to the policy change. This is one advantage of using states that had already passed laws as the comparison group (Pesko et al., 2017).

(moral hazard). These findings support an adverse selection story, in which physicians who are likely to commit malpractice are induced to move to reformed malpractice environments.

In this study, we build on Lieber's intriguing result and test whether malpractice laws induce sorting of physicians by medical human capital: specifically, the medical specialty of training, years of training, the selectivity of the medical school attended, and quality of the teaching hospital during residency. As we discuss later in the paper, we focus on these measures of human capital because they are plausibly associated with physicians' vulnerability to malpractice pressure and with patient treatment outcomes. We use pooled data from the 1998-2017 New York State (NYS) Residents' Exit Survey, which include all exiting medical residents from hospitals in NYS, along with zip codes of the locations of their first practices, and their individual characteristics. The analysis focuses on caps on non-economic damages, which limit damage awards for pain, suffering, and other nonpecuniary injuries, because: (1) these reforms have been shown to reduce medical liability costs (CBO, 2004); (2) we have a stronger theoretical underpinning for predicting the direction of effects on physician supply for non-economic damage caps compared to other tort reforms; and (3) we have ample within-state variation in these laws during the analysis period, 1998-2017.

Our paper contributes to the literature in the following ways. First, we focus on how tort reforms affect how physicians distribute themselves across states. We consider sorting by specialty and by the aforementioned measures of medical human capital that may be associated with susceptibility to malpractice risk and patient treatment outcomes, and which have been neglected in supply studies. Second, our data extend until 2017, giving us sufficient post-policy data to study the recent wave of tort reforms in the 2000's. Third, unlike prior work (Chou & Lo Sasso is an exception), we focus on physicians who are just graduating from residency programs and entering their first jobs. Tort reform laws are more likely to affect location choice among new physicians due to the high fixed costs of establishing or joining a practice and building a patient base. Finally, most prior studies use aggregated data (e.g., AHRF), but we use microdata and a methodological approach (described later), that allows us

to test for and gain information on heterogeneous preferences across physicians regarding tort reform. Specifically, our approach allows us to calculate physicians' willingness to pay (WTP) to locate in a state with tort reform, and how this WTP varies along physician human capital characteristics that potentially are related to malpractice risk.

Our findings suggest that caps on non-economic damages are associated with an increased probability of new physicians locating in the state that passed the reform, and the magnitudes of these effects are economically meaningful. On average, a new physician is willing to give up about 13 percent of his/her hourly wage to locate in a state that has passed a non-economic damages cap. When we focus on the effects of non-economic damage cap laws on physician sorting, an intuitive pattern of findings emerges. We hypothesize that physicians whose specialty and other human capital characteristics may predispose them to higher malpractice risk will respond more to tort reforms, and the results support this idea. Effects of the laws are stronger for physicians in specialties that tend to face the highest risk of malpractice claims compared to non-high-risk physicians. Physicians trained in lower-ranked and -performing hospitals, and graduated from less selective medical schools, respond more to damage cap laws than physicians trained in higher-ranked and -performing hospitals and more selective medical schools. We find, however, no differences in responsiveness to tort reform by physician's years of training.

2. Physician Human Capital

Physicians must be covered by malpractice insurance. Such insurance is often only weakly experience-rated (Sloan, 1990) and some physicians are largely insulated from changes in premia due to working in group practices. Hospitals, however, often self-insure so do face financial incentives to reduce malpractice exposure which they may translate into pressures on hospitalists. But whether or not a physician is financially insulated from malpractice, a malpractice suit can cause the physician

emotional stress and consume her time, and, importantly if it leads to an adverse judgment, can damage her reputation, reducing her future earnings and labor market opportunities.

Expected malpractice costs should vary with the quality of a physician's patient care, and this should vary with the quality of her training. However, according to an influential 1990 study malpractice judgments have a strong random component and their likelihood is imperfectly correlated with patient care. Based on an examination of medical case files, this study's authors conclude that actual medical negligence rarely leads to a malpractice claim and most judgments that result from claims are not justified based on the medicine practiced (Harvard Medical Practice Study, 1990). But White (1994) has argued that conditional on a claim involving negligence, the likelihood of a payout is much higher than was found in the HPMS study. Moreover, based on her review of the literature and her own work with Farber (Farber and White, 1994) she concludes that the average ex ante cost to a provider of an incident of negligent care is \$4500 (inclusive of legal fees) and thus substantial.

Reform and the number of physicians

Theoretically, malpractice reform, by increasing the difficulty in initiating a malpractice suit and reducing the potential payout, should reduce malpractice insurance premiums, uninsured costs, and the variance (risk) in physician utility, and thus shift out the physician supply curve in the reforming state. Malpractice reform may, however, have countervailing effects on consumer demand for health care services. From the perspective of the consumer, malpractice reform increases her difficulty filing a suit, and reduces the odds and size of an award. Consumers may also see malpractice reform reducing physician accountability, and thus the quality of physician care. This would tend to shift in the demand for physicians' services. On the other hand, evidence suggests less than half of the award from a judgment accrues to the patient-plaintiff.⁴ There is also evidence that malpractice concerns motivate physicians to practice defensive medicine (Kessler and McClellan, 1996, 2002a, and 2002b). A

⁴ See the citations in Kessler, 2011, including Studdert et al, 2006.

perception that physicians are likely to prescribe unnecessary treatment in an un-reformed state may lead consumers to view the services they receive more favorably after reform. Whether the forces shifting physician supply out dominate the forces shifting demand in, and thus whether reforms attract physicians on net, including new physicians, to the state, is an empirical question.

Physician human capital and the supply effect

A physician's education and training background should influence her attraction to reform in two ways. First, the appeal of legal protections from malpractice risk will vary with the *kind* of medical expertise (e.g., family practice vs. surgery), as detailed above. Second, the quality and length of a physician's education and training should be predictive of her ability and thus her patients' outcomes, and consequently her exposure to malpractice risk. A growing literature on physician performance is establishing that within specialties, patients do better under the care of some doctors compared to others. Jha and Epstein (2006) show that for surgeons who perform coronary artery bypasses, past (risk-adjusted) performance predicts future mortality rates. Currie, MacLeod, and Van Parys (2016) report evidence that physicians systematically vary in their treatment of heart attack patients, with some physicians tending to treat patients aggressively and others to tailor their treatment strategies more closely to the characteristics of the patient. They find that cardiologists who are more aggressive treaters and whose treatment is more tailored produce better outcomes.⁵ Interestingly for our study, they also find that physicians trained in the top 20 medical schools are more likely to be both aggressive and responsive, suggesting a training component to these physician fixed effects.⁶ Currie and MacLeod (2017), using administrative data to identify physician skill in a population of obstetricians, show skill variation among obstetricians that also has measurable health consequences for their patients' outcomes.

⁵ An issue in measuring physician quality is identifying the physician's effect on a patient's outcome when patient-physician matching is nonrandom. The authors argue that this treatment variation is likely entirely physician driven as the data are from emergency department visits.

⁶ They also find that physicians trained in U.S. medical schools are more aggressive and less-responsive than physicians who had been trained abroad.

While the relation between patient outcome-related physician human capital and malpractice filings and judgments is not well-studied, the available evidence, described in this section below, suggest one. We use pre-practice measures of education and training: the selectivity of the medical school, the quality of the teaching hospital in which the residency takes place, and years of residency. These human capital measures capture malpractice risk if performance and selectivity are correlated with innate, pre-training ability or with the training's value-added. But even if school selectivity or residency and training hospital quality measures indicate no differences in ability to treat patients, they still would be informative of malpractice risk if the reputation of a physician's training institution affects patients' likelihood to sue or juries' verdicts.⁷

Medical school selectivity: There are surprisingly few studies of the relationship between medical school selectivity and physician performance. An exception is Doyle et al. 2010. They use evidence from a large natural experiment that randomized patients to physician teams, one set of teams comprising residents and attending physicians from a highly ranked medical school and teaching hospital, and the other set comprising residents and attending physicians from a much lower ranked medical school and teaching hospital. The teams from the more prestigious medical school dominated the teams from the less prestigious medical school on medical school characteristics (student MCAT scores and NIH funding ranking), affiliated hospital (US New and World Report ranking), residents' medical schools (US New and World Report rankings, NIH funding rankings, percent which are foreign medical schools), and residents' board certification rates. They find that patients assigned to the highly ranked teams experienced substantially lower costs (by as much as 25 percent)—the physicians in the lower-ranked program ordered more tests and took longer to order them—but no differences in health outcomes. Cuddy et al. (2017), using a large sample of MDs graduating from medical school between 1994 and 2006, report that a one standard deviation increase in the Step 2 Clinical Knowledge score on the United States Medical

⁷ See Hartz et al. (1999) who show that physicians responding to surveys of coronary artery bypass surgeon quality are influenced by the reputation of the surgeons' institutions of training.

Licensing Exam (USMLE)⁸ is associated with about a 25 percent decrease in the likelihood of having a disciplinary action from a U.S. state medical board, after controlling for gender and the number of years since medical school graduation. While the evidence on medical education quality and postgraduate performance is scant, as Gardner and Vishwasrao (2010, p. 63) note, “[t]here is certainly a presumption that the quality of physicians is important to patients even if the effects on the outcome are not measured. Hospitals and physician groups often advertise where their new hires were trained, and many insurance companies attempt to provide information about physician attributes to prospective patients.”

Our first measure of medical school selectivity is whether the physician trained at a medical school outside the US and Canada rather than at a US medical school. U.S. citizen graduates of international medical schools have lower first-time pass rates on Step 1 and Step 2 of the USMLE compared to US citizen graduates of US medical schools (Boulet et al. 2009). Eckhert and Van Zanten (2015) also note their much lower success rates in residency matches compared to US medical school graduates.⁹ Most US citizens attending medical schools outside the US and Canada do so in the Caribbean at medical schools that recruit international students who intend to practice in the U.S.¹⁰ Caribbean medical schools have higher acceptance rates and lower average MCAT scores than US medical schools (Busnaina, 2011).

Our second measure is whether the physician received her medical education at an allopathic or osteopathic medical school. In 2016, nine percent of actively licensed physicians held Doctor of Osteopathic Medicine (DO) degrees, with most of the rest holding Doctor of Medicine (MD) degrees (Young et al. 2017). Whereas the coursework in schools of allopathic and osteopathic medicine is similar and graduates of both types of schools have to pass the same state boards, admission to

⁸ Passing the three-step USMLE is required to practice medicine in the US.

⁹ The success rates in 2014 were 53% for U.S. international medical school graduates (IMG) compared to 94% for seniors in U.S. allopathic medical schools, and 78% for students and graduates of U.S. osteopathic medical schools, according to data from the National Resident Matching Program.

¹⁰ In 2016, there were 953,695 actively licensed physicians, of which 21,519 were U.S. citizen Caribbean medical graduates (Young et al. 2017).

allopathic medical schools is more competitive, with allopathic schools admitting more academically accomplished students.¹¹ In Appendix A, using medical malpractice and medical adverse action data from the National Practitioner Data Bank and physician population data from the AHRF, we show that the (unconditional) rates of malpractice judgments and non-malpractice adverse actions (e.g., loss of clinical privileges, licensure and censure and certification suspensions) are higher for osteopaths compared to allopaths (see Appendix A).

Quality of teaching hospitals: Residencies, where recent medical school graduates learn practical skills working on hospital wards supervised by experienced physicians, are seen as important in the transition from student to independent physician (Bard, 2011), and yet there are few studies of their impact on physician patient care. Doyle et al.’s evidence can be interpreted as showing the importance of resident quality on treatment efficiency. Asch et al. (2009) report systematically lower maternal complications in deliveries performed by obstetricians who trained at some residency programs compared to others. (Epstein and Nicholson, 2009, however find that residency programs explain only a small percentage of variation in obstetrician “treatment styles.”). As one measure of residency quality, following Doyle et al. and Gardner and Vishwasrao, we use the US News and World Report (USNWR) hospital ranking in which the residency training occurred, as well as a composite of the hospital’s USNWR individual performance metrics.

Subspecialty training (years of training): After completing residency, physicians in many specialties can choose to complete additional years of training in a sub-specialty. More years of training may mean greater skill, translating to better patient outcomes. It may also connote higher ability, as more able physicians are more likely to engage in extra training, or it may connote more “learning by doing”; for many procedures practice is important for proficiency (for such evidence in coronary artery

¹¹ MCAT scores and undergraduate GPAs are higher among matriculants at allopathic schools (<https://www.aamc.org/download/321494/data/factstablea16.pdf>) compared to osteopathic schools (<https://www.aacom.org/become-a-doctor/applying/general-admission-requirements>). Graduates of allopathic schools also have higher residency program acceptance rates (see footnote 8).

bypass graft surgery see Wen et al., 2006, and Agency for Healthcare Research and Quality, 2004).

Whether additional years of training indicate greater learning by doing or more skills or greater innate ability at the start of the physician's career, we hypothesize that subspecialty training may lead to better patient outcomes and less susceptibility to malpractice risk.

3. Methods

The mixed logit model

Holding other factors constant, reform should attract physicians, but especially physicians who are particularly vulnerable to malpractice risk. We posit that a physician's willingness to pay for reform is positive and increasing in the human capital quality measures that we use to proxy malpractice risk. To measure these willingness-to-pay predictions, we estimate a mixed logit model of location choice for physicians just beginning their careers.¹² The mixed logit model allows one to estimate the association between graduating physicians' first practice location choices and state characteristics, including malpractice reforms, while allowing for flexibility in substitution patterns and variation across physicians in their tastes for different locations (Train, 2009). Each graduating physician faces the same choice set – 50 states and the District of Columbia (we ignore the possibility of working abroad). The attributes of states within the choice set, however, vary across physicians since they are graduating in different years, and state characteristics including policies change within states over time. Moreover, as we motivated above, physicians who differ in their vulnerability to malpractice risk may respond differently to state tort reform when making decisions about where to locate.

To be concrete, we let U_{nj} be the utility that physician n receives from choosing to locate in state j . In Equation 1 below, the physician's utility is based on x_{nj} , the observed characteristics of the location choice related to the physician and the state, such as the state's malpractice laws and the

¹² Chou & LoSasso (2009) study determinants of physician location choice using earlier waves of the NYS resident survey, but they use a conditional logit model instead of the mixed logit model, and a more limited sample.

average state-level physician wage in the physician's specialty in the year the physician graduates, β_n , a set of random coefficients corresponding to these observed characteristics, and ε_{nj} , and a random error term that is assumed to be distributed iid type 1 extreme value. An alternative-specific constant, ζ_j , captures the effect of all unchanging attributes of the state not measured in x_{nj} on the physician's utility. We assume the β coefficients vary across physicians according to a density $f(\beta)$, unlike in the standard logit specification in which the β coefficients are assumed to be fixed (Train, 2009, page 137). In this manner, the model captures physicians' varying tastes over different locations.

$$U_{nj} = \beta'_n x_{nj} + \zeta_j + \varepsilon_{nj} \quad (1)$$

This formulation represents the additive random utility model, and the individual physician chooses the alternative that provides the highest utility (see e.g. Cameron & Trivedi, 2005). The physician knows her own β_n , ζ_j , and ε_{nj} for each alternative (as well as her own x_{nj} , of course), but the researcher can only observe each physician's x_{nj} . If ε_{nj} were the only source of randomness the the probability that physician n chooses state i , would assume the standard logit formulation shown in Equation 2 below.

$$L_{ni} = \frac{e^{\beta'_n x_{ni} + \zeta_i}}{\sum_j e^{\beta'_n x_{nj} + \zeta_j}} \quad (2)$$

However, because β_n is random one must integrate over the possible values of β_n , as shown below in Equation 3, to obtain the unconditional choice probabilities. Typically, $f(\beta)$ is assumed to be a normal or lognormal distribution, with the latter being used when the coefficient is thought to have the same sign for all decision makers. The mixed logit model is estimated using maximum simulated likelihood.

$$P_{ni} = \int \left(\frac{e^{\beta'_n x_{ni} + \zeta_i}}{\sum_j e^{\beta'_n x_{nj} + \zeta_j}} \right) f(\beta|\theta) d\beta \quad (3)$$

In our context, the mixed logit has advantages over standard approaches such as the logit and conditional logit. First, this model provides information about the underlying taste parameters, and it allows heterogeneity in preferences across physicians. This flexibility has not been incorporated in prior work on the effects of tort reform on physician decision making. Second, the standard logit and conditional logit models have the independence from irrelevant alternatives (IIA) property, implying that the ratio of the probabilities of choosing two alternatives is not affected by the existence or attributes of any other alternatives. IIA also implies that an improvement in an attribute of one alternative reduces the probabilities of choosing all other alternatives by the same percentages. For example, IIA means that the odds of a physician choosing Florida vs. New Jersey is not affected by whether staying in New York is an option. It also implies that if New York were to enact tort reform, and if tort reform increased the likelihood that physicians choose New York, the reform would draw physicians away from all other states by the same percentages. The mixed logit model does not have the IIA property which leads to these restrictive substitution patterns. This is especially important in our context, since substitution patterns between different states are likely to vary.

Finally, the ratio of coefficients in the mixed logit yields useful information about physicians' willingness to pay (WTP) for tort reform. In our model (data details discussed below), the covariates include an indicator of whether or not the state had a non-economic damages cap, and the average hourly wage rate in the physician's specialty in the state, both measured in the year the physician graduated from residency. The ratio of the estimated coefficients on these two covariates captures, on average, how much a physician is willing to pay (in terms of reduced wages) to locate in a state with a

non-economic damages cap law in place (Train, 2009, page 39).¹³ The models also include the density of physicians in the physician's specialty in the state, the state unemployment rate, and the log of the distance from the training hospital to the state (details on variable construction provided below).

Initially, we estimate the mixed logit model using the whole sample. Based on the model estimates, we calculate physicians' average WTP, as well as the distribution of physicians' WTP, for damage cap laws. In our specification, we assume all reform parameters follow a normal distribution, while all other parameters, including the hourly wage and the distance variables, are assumed to be fixed. Models are estimated using maximum simulated likelihood with 100 Halton draws (Train, 1999). All models include state fixed effects, with New York State (NYS) set as the baseline. Although our main results are based on a model with includes indicators for current tort reforms, we also consider lead and lag specifications. Models with lags address a possible fading or accelerating effect and leads capture any anticipation effects. In models that include lead and lag indicators of tort reform, we limit the data to 2009 and prior years to reduce interpretation problems because after 2009 some states "turn off" rather than "turn on" laws. If we limit the sample to 2009 and earlier years, all states are turning on laws during our study period except Oregon, which turns off a law in 2000.

Interactions between tort reform and physician human capital

¹³ There are two ways to obtain the distribution of our WTP estimate. Consider the following specification for simplicity.

$$U_{nj} = \beta_n \text{reform}_j + \alpha_n \text{wage}_j + \varepsilon_{nj} \quad (\text{a})$$

The conventional method is to estimate the model first and then compute β_n/α_n . Train and Weeks (2005) call (a) a model *in preference space*. But the limitation of this method is one must make assumptions about the distributions of parameters to ensure that the distribution of WTP is well defined. For example, one cannot assume that both the distribution on cap laws and the distribution of wages are normally distributed because then the ratio of these coefficients would not have defined moments (Hole, 2008). To avoid this problem, Train and Weeks (2005) estimate the model *in WTP space* such that WTP can follow any random distribution. We divide Equation (a) by α_n , so it can be re-written as

$$V_{nj} = \beta_n/\alpha_n \cdot \text{reform}_j + \text{wage}_j + \varepsilon_{nj}/\alpha_n \quad (\text{b})$$

From equation (b), we can see that estimated coefficients include WTP itself, which indicates that we can directly make any assumption about the distribution of WTP. The log-likelihood of our main model estimated in WTP space was considerably lower than the model estimated in preference space. Thus, our approach in the paper is to estimate the model in preference space (the conventional method), and then compute WTP.

Next, we test whether WTP for tort reform varies by physician human capital characteristics associated with malpractice risk. To do so, we estimate a set of mixed logit models which include an interaction term between each human capital characteristic and the indicator for cap on noneconomic damages (we include the main effect of the human capital characteristic in each model as well). Then, we test for differences in WTP between physicians with and without the human capital characteristic. For example, we estimate the mixed logit model with an interaction between non-economic damages cap law and international medical graduate; compute the difference in WTP between international and US medical graduates (this is just the estimated coefficient on the interaction term divided by the estimated coefficient on the wage); and use the delta method to compute the standard error of this difference. We estimate separate mixed logit models for each opposing attribute pair (e.g., high-risk vs. non-high-risk specialty, international vs. US medical school graduate).

To make sound comparisons in WTP between sub-samples of physicians, we need to control for potentially confounding physician-level characteristics such as demographic characteristics and physician specialty that may vary systematically across sub-samples. For example, when comparing WTP for tort reform between osteopaths and allopaths, it is important to adjust for differences in specialty between these two groups, since osteopaths are less likely than allopaths to be surgeons, and non-surgeons face lower malpractice risk. Physician-level (“case-specific”) variables such as specialty can be included as covariates in the mixed logit model by interacting each case-specific variable with each alternative in the choice set. In practice, however, given the large sizes of our choice set and sample, including case-specific covariates increases computational time considerably and causes model convergence issues.

To address this issue, we use propensity score weighting to balance sub-sample pairs along observable characteristics. To generate the propensity weights, we use a machine learning approach to balance the full, joint distribution of covariates across each pair of sub-samples rather than the marginal distribution of each variable individually (Haviland et al., 2016). For example, we estimate a set of

propensity weights that balances the joint distribution of covariates across the osteopath and allopath sub-samples. In particular, we weight the osteopath and allopath physicians to estimate a population average effect of the osteopathy degree on the desire to locate in a reform state.¹⁴ To construct the weights, we use generalized boosted regression, a regression tree approach which essentially “...adds together many simple functions to estimate a smooth function of a large number of covariates (McCaffrey et al., 2004, page 407).” This approach iteratively searches over regression tree models, each fit to the residuals from the prior iteration, until the degree of balance in the covariates between the two sub-samples is maximized (see details in McCaffrey et al. 2004).¹⁵ Following Haviland et al. (2016), the algorithm uses the Kolmogorov-Smirnov (KS) statistic to gauge the degree of balance in covariates across sub-samples. This approach is based on the difference across the samples in the full distribution of a variable rather than just the differences in the means.

Using propensity scores, we balance the following set of covariates across each pair of sub-samples: (1) whether the physician is an international medical graduate; (2) whether the physician went to medical school in NYS; (3) whether the physician is female; (4) eight indicators of medical specialty (with one left out as the baseline) based on the Standard Occupational Classification System (SOCC) – anesthesiologists, family and general medicine, internal medicine, obstetrics/gynecology, general pediatrics, psychiatrists, surgeons, and all other specialties; (5) age; (6) an indicator for solo practitioner; and (7) an indicator for solo practitioner is missing.¹⁶ This method led to weights that achieved excellent balance in covariates across each pair of sub-samples (discussed later in the paper). We applied these propensity weights in all regressions which include interaction terms between physician

¹⁴ The population average “treatment effect” (or ATE) of the osteopathic degree on the reform preference is produced by weighting each osteopath by $1/p$ and each allopath by $1/(1-p)$, where p is the propensity score, which is here the probability that the physician takes the osteopath attribute.

¹⁵ We implemented this approach using the TWANG module in R. We used the following tuning parameters: maximum regression tree level of 2, shrinkage=0.01 and bagging fraction=0.5.

¹⁶ Note that if the sub-sample is based on a particular covariate, then this covariate is not included in the propensity score model. For example, if propensity scores are being generated for the international medical graduate vs. US medical school pair of sub-samples, the propensity score model cannot include the indicator for whether or not the physician is an international medical graduate.

human capital characteristics and cap on noneconomic damages. Intuitively, this approach allows us to compare WTP across two sub-samples, while adjusting in a highly flexible way for differences in observable characteristics across the two sub-samples. Propensity score weighting does not address possible confounding by unmeasured characteristics; this issue remains a limitation of this analysis.

4. Data

Data for this study come from the Survey of Residents Completing Training in New York (Exit Survey), an annual survey of all physicians completing a graduate medical education (GME) (either a residency or fellowship) program in NYS. The survey, conducted by the University at Albany Center for Health Workforce Studies (CHWS), provides the medical education community in NYS with detailed information on the outcomes of residency and fellowship training programs. The CHWS fields the survey each spring through GME administrators at all teaching hospitals in NYS. The survey includes questions about basic demographics, the name of the NYS residency/fellowship program, type and length of post-graduate training, the name of the medical school attended (if that school is located in NYS), whether the medical school is located in the US if not, whether the medical school is osteopathic or allopathic, specialty, educational debt, job market experience, and plans for practice after graduation (Armstrong, Chung & Forte, 2015).

In this study, we pool data from 17 years which include the 1998-2002, 2005, 2007, and 2008-2017 surveys. The response rate over the time period during which the survey has been conducted is about 60 percent (Armstrong, Chung & Forte, 2015). The analysis sample is limited to residents/fellows who were planning to enter positions that primarily are in patient care, had accepted a job offer at the time of the survey, and provided the state in which the new position was located. To ensure that physician location choices were not constrained by immigration considerations or Health Professional Shortage Area (HPSA) location requirements, we further limit the sample to current US citizens (including native-born citizens, naturalized citizens, and Green Card holders) without a HPSA obligation.

We match state medical malpractice reform laws to physicians based on the graduation year.

The data on medical malpractice laws come from an extensive database on state tort law reforms that has been compiled and made publicly available by Ronen Avraham (Avraham, 2019). The database contains the following state malpractice reform laws: caps on non-economic damages; caps on punitive damages; caps on total damages; split recovery reform; collateral source reform; punitive evidence reform; periodic payments reform; contingency fee reform; patient compensation fund reform; and comparative fault reform.

As shown in Table 1, during our study period (1998-2017), 15 states either enacted or rescinded laws that place caps on non-economic damages. The laws are coded as belonging to the subsequent year if they went into effect on or after July 1 of that year. Based on this database, we create a set of five tort reform indicators (for the most-widely studied reforms in the economics literature), with each indicator set equal to one if the state had that law in place in the year the physician graduated and set equal to zero otherwise. The indicators are: (1) non-economic damages cap (our focus); (2) punitive damages cap; (3) split recovery reform; (4) collateral source reform; and (5) joint and several liability reform. For the non-economic damages cap, the database also contains the dollar amount of the cap. The dollar amount of the cap changes within some states during our study period mainly because the cap is inflation-adjusted in some states. About 84% of the noneconomic damages caps (by state-year) range between \$250,000 to \$500,000. We follow much of the literature (e.g., Lieber and Currie and MacLeod, 2008) in that we do not construct a continuous measure of the reform but instead adopt an indicator variable categorizing whether a cap exists.

In addition to state tort reforms, we merge into the data state-level, time-varying data on the density of physicians in the physician's specialty; mean hourly wages in the physician's specialty; and the unemployment rate. These data come from the 1998-2017 Occupational Employment Statistics Survey conducted by the Bureau of Labor Statistics, and the AHRF, which includes unemployment data from the American Community Survey. Mean hourly wages for physicians are based on a work year of

2,080 hours; are calculated for physicians who are not owners/partners; do not include bonuses but do include on-call pay; and are inflation-adjusted with 2010 as the base year. Physician density is the number of physicians in a specific state, year, and specialty per 10,000 state residents (population estimates come from Census projections). To measure distance, if the physician's county of training is available, we calculate the distance in miles from the county in which the physician completed training to the center of each of the 50 states and DC. If the physician's county of training is missing, we compute the distance from the center of NYS to the center of each of the 50 states and DC. In the models, we use the natural log of distance as a covariate to allow for a non-linear effect of distance on physician location decisions.

Physician human capital characteristics capturing malpractice risk

First, we create an indicator of high-risk specialty that is set equal to one if the physician specializes in neurosurgery, neurology, internal medicine, pulmonary medicine, general surgery, pathology, pediatrics, obstetrics and gynecology, thoracic–cardiovascular surgery, cardiology, or family general practice. Physicians who specialize in any other fields are considered “not high-risk specialists.” The categorization above is based a recent paper by Jena et al. (2011) in which the authors use claims data from 1991 to 2005 from a large, liability insurer with clients nationwide to classify physician specialties by the proportion of physicians in each specialty who have a malpractice claim in a year; the size of the claim; and the cumulative career malpractice risk by specialty. We consider a physician to be in a high-risk specialty if the physician's specialty is one that has higher than median size of the claim based on the classification shown in Jena et al. (2011).

In sensitivity checks, we test two alternate definitions of high-risk specialty. Based on Jena et al. (2011), we consider an alternative classification that is sensitive to both the size of claim and likelihood of a claim: we assign a specialty to the high-risk class if physicians in the specialty have both a higher than median likelihood of having a claim, and a higher than median average claim. As a second alternative definition, we also use the definition of high-risk specialty proposed by Klick & Stratmann

(2007, page S125 Footnote 5), who use data on average medical malpractice awards per doctor from the Florida Closed Claims Medical Malpractice data set. These authors categorize “high risk” specialists as those with the top 10 highest average awards, and “low risk” specialists as those with the lowest 10 average awards. Note that we are implicitly assuming that specialty is exogenous whereas survey evidence suggests that malpractice concerns may be affecting specialty choice (Kiker & Zeh, 1998).

Next, we create an indicator for whether the physician is an international medical graduate versus a US medical school graduate (note that our sample is limited to US citizens, so international graduates are US citizens who graduated from international medical schools), and second indicator for whether the physician is an osteopath (D.O.) rather than an allopath (M.D.). Like medical specialty, malpractice concerns may affect the choice to attend international or osteopathic medical schools, but, to our knowledge, this topic has not been studied.

We consider the overall quality of the teaching hospital during residency based on the current (2019) USNWR ranking of NYS hospitals. To measure “lower than median quality,” we create an indicator of whether the physician’s training hospital was ranked either 16th (the median ranking in our sample) or worse in NYS. This indicator is set to one for un-ranked training hospitals as well. Using the hospital’s USNWR ranking, we also generate a quality index based on a factor analysis of the USNWR’s quality measures. The USNWR ranking system, as well as the factor analysis, are described in Appendix B.

Our final measure is whether the physician’s number of training years was at the median (in the physician’s specialty, based on our sample) or below the median. This indicator captures physicians who likely have not completed any sub-specialty training, as well as (possibly) physicians who have left residency programs without completing them – for example, a physician in a surgical residency may complete the first internship year only, and then seek a non-surgical job without completing the remaining years of the surgical residency.

5. Results

Table 2 describes the sample we analyze. (Recall we are analyzing only the US citizens among NYS residents, thus non-citizens are omitted from Table 2.) The average resident in our sample is young, 33.4 years old, and about 45 percent are female. Interestingly, about one third of NYS residents are graduates of foreign medical schools. About half the rest went to medical school in NYS. About 10 percent are osteopaths and 48 percent are in high risk specialties. About 34 percent of our sample attended high school in NYS, suggesting that that many NYS physicians have ties to NYS. About 20 percent of the sample is starting a job in a state with a cap on non-economic damages, and 31 percent is starting a job in a state with a cap on punitive damages. In the sample, the most common states chosen by physicians are: New York (60%); Florida (5%); California (5%); Pennsylvania (3%); and Texas (3%), and at least one physician chose each of the 51 alternative state options (see US map, Figure 1).

Effects of non-economic damages caps on physician location

Table 3 shows findings from mixed logit models in which we assume that the malpractice reform coefficients are normally distributed and the coefficients on all other covariates are fixed. Choosing the normal distribution over the lognormal allows us to be empirically agnostic regarding the signs of the effects across the distribution of preferences. Column 1-5 show results for models that include indicators for cap on non-economic damages, cap on punitive damages, split recovery reform, collateral source reform and joint and several liability reform, respectively, one at a time. (Note that each of these reforms should either reduce the likelihood of litigation or the expected penalty if an adverse judgement or both. See Appendix Table 1.) Column 6 shows findings from a model that includes all five reforms at the same time. For the malpractice reform coefficients, we show the estimated mean and the estimated standard deviation. The findings in Table 3 indicate that caps on non-economic damages and collateral source reform both increase the likelihood that a new physician will choose to locate in that state. Only the finding for non-economic damages caps, however, persists when all the reforms are entered into the model at the same time (column 6, Table 3). No other tort reforms have a statistically significant

association with physician location. Except for collateral source reform, the estimates of the standard deviations of the reforms are large but imprecisely estimated.

From Table 1, we see that our findings for collateral source reform are driven by within-state variation from just a few states. In addition, a likelihood ratio (LR) test indicates that the goodness of fit does not improve when we include all malpractice reforms at the same time (col. 1 vs. col 6) (LR test results shown in last row of Table 3). We move forward in the rest of the analyses with the specification shown in column (1) of Table 3 that includes only the cap on non-economic damages, which has ample within-state variation during our time period. To interpret the magnitude of the coefficient on the non-economic damages cap, we compute the willingness to pay (WTP), the marginal rate of substitution between reform and the real hourly wage. Since we assume the coefficients of reforms have a normal distribution and the coefficient of wage is fixed, the WTP (which is mathematically the ratio of these two coefficients) is also normally distributed. We find that an average new physician is willing to sacrifice \$10.83/hour to work in a state with cap on non-economic damages, which is approximately a 13% decrease at the mean wage (Column 1, Table 3).

Figures 2.1-2.2 show the distributions of the estimated coefficient on non-economic damages cap (Figure 2.1), and of estimated physicians' WTP to locate in a state with a non-economic damages cap (Figure 2.2). We construct these figures based on the point estimates of the utility parameter and its standard deviation shown in Column 1 of Table 3 and with the caveat that we cannot rule out a degenerate distribution. Taking the estimates at face value, the figures imply that about 72 percent of residents have positive WTP to locate in a state with a non-economic damages cap law (or, equivalently, the probability that the estimated coefficient on non-economic damages cap is positive is about 0.72). Most physicians, therefore, value having a non-economic damages cap law in place in the location of their first job. But the figures also suggest that some prefer high (malpractice) pressure states, which is consistent with a contracting view of malpractice (Courty and Marschke, 2008) in which physician ability is difficult to observe and malpractice activity signals it to patients, insurers, and employers.

According to this view, malpractice reform strips the malpractice record of some of its information content, increasing the relative attractiveness of non-reform states to high ability physicians.¹⁷

All specifications in Table 3 include state physician density (number of physicians per 10,000 state population), state unemployment rate, and distance from each state to the physician's training hospital in New York. In all specifications, these parameter estimates are fixed, not random. In all models estimated, higher physician density and closer proximity increase the likelihood of the physician locating in the state. By construction, distance matters more to physicians when they live closer to training hospital, which is intuitive. The coefficient estimate on the state unemployment rate is generally statistically insignificant.

Robustness checks of main specification

In Appendix Table 2, we show results from a diagnostic check of the model. Note that $f(\beta|\theta)$ in equation (3) is the unconditional distribution of β in the entire population. To test whether our model is correctly specified and consistently estimated, following Train (2009), we compare the unconditional distribution with the conditional distribution aggregated over all physicians. Using Bayes' rule, we can readily derive the conditional distribution.

$$h(\beta|S_n = i, \theta) = \frac{P(S_n = i|\beta) \times f(\beta|\theta)}{P(S_n = i|\theta)} \quad (4)$$

where S_n is the state physician n chooses and $h(\beta|S_n = i, \theta)$ is the conditional distribution of β among physicians who choose state i . From (4), we can obtain the conditional expectation of β_n .

$$E[\beta_n|S_n = i, \theta] = \int \beta \times \frac{P(S_n = i|\beta) \times f(\beta|\theta)}{P(S_n = i|\theta)} d\beta$$

¹⁷ Thus, high ability and incentive-backed performance measures in medicine are complements. See also Lazear, 2000, who reports evidence of complementarity between high-stakes performance measurement and high ability in car windshield installers. Malpractice reform should also generate compensating wage differentials. The higher wages in non-reform states are likely to attract higher ability physicians but be insufficient to compensate lower ability physicians for their higher costs of exposure. Ability-biased sorting responses are the focus of Table 4 and the surrounding text.

$$= \frac{\int \beta \times P(S_n = i|\beta) \times f(\beta|\theta) d\beta}{\int P(S_n = i|\beta) \times f(\beta|\theta) d\beta} \quad (5)$$

Since (5) does not have a closed form, we use simulations and present conditional distributions of random parameters of malpractice. If the model is properly specified, the mean of the conditional distribution should be similar to that of the unconditional distribution. As shown in Appendix Table 2, the mean of the expectations of the distributions of β conditional on individual choices are almost identical to our main results from the mixed logit models, which provides support for the model's validity.

In Appendix Table 3, we consider specifications with leads and lags, based on the model shown in column 1 of Table 3. In the models shown in Appendix Table 3, we limit the sample to years prior to 2009 to (mostly) avoid the interpretation issues involved in evaluating lead and lag effects of laws being repealed, which occurs more frequently starting after 2009. Between 1998 and 2009, all states that made changes to non-economic damage cap laws turned the laws on, apart from Oregon which repealed the law in 2000. Column 1 in Appendix Table 3 shows our main specification from column 1 in Table 3 estimated on data from 1998-2009 (our main results are based on a sample that spans 1998-2017). The magnitude of the estimated coefficient on non-economic damages cap is almost the same, although it is no longer statistically significant when we limit the years to 1998-2009. In column 2 of Appendix Table 2, we add three lagged terms to this model, which capture whether the state had a non-economic damages cap in place three years, two years, and one year prior to the physician's year of graduation. The estimated coefficients on these three lagged terms are all negative in sign but they are not statistically different from zero. Column 2 of Appendix Table 3, therefore, suggests that conditional on the current law, the effect of the cap on non-economic damages neither fades nor accelerates with time.

In Column 3 of Appendix Table 3, we add three lead terms to the specification – whether the state had a law in effect one, two, or three years after the physician's year of graduation. The estimated one-year lead coefficient is large in magnitude and statistically significant, although the two- and three-

year leads are not statistically different from zero and are negative in sign. These findings suggest that physicians may make location decisions in anticipation that a law will turn on in the next year. This is consistent with Malani & Reif, and with the fact that non-economic damages cap laws often are widely publicized before enactment. This finding may also reflect the fact that our laws are coded as “turning on” if the law is in effect before July 1 in the year the physician graduates. For example, if a law became effective in August of 1998 and the physician graduated in 1998, then the law would be coded as “turning on” in 1999 for that physician. This is appropriate because physicians generally complete residencies prior to July 1 of the year in which they graduate – but it affects the interpretation of our specification that includes lead terms since this kind of effect captures very short-run anticipation. In sum, we do not find that the effect of a state’s non-economic damage cap fades or strengthens over time, but we do find evidence that physicians anticipate changes to the cap law in the near future.

In Appendix Table 4, we test the robustness of our main findings in column 1 of Table 3 to alternate approaches. First, we estimate the model using a conditional logit, a model estimated elsewhere (Chou and Lo Sasso) instead of a mixed logit (column 1, Appendix Table 4). Second, we consider a mixed logit with the reform parameter assumed to be lognormally distributed instead of normally distributed, which is our main approach (column 2, Appendix Table 4). Assuming a lognormal distribution for the reform parameter is a natural assumption if we were confident that reform always entered the physician’s utility positively. Third, we estimate a mixed logit model assuming that all coefficients are normally distributed (a more general specification than our main specification in which we assume fixed coefficients for all coefficients but the malpractice laws) (column 3, Appendix Table 4); and, finally, we estimate a generalized multinomial logit model (G-MNL) which allows for possible scale heterogeneity (column 4, Appendix Table 4).

In the simple conditional logit model, the variance of the unobserved error term is $\sigma^2 \frac{\pi^2}{6}$, where σ is a scaling parameter. To understand the role σ plays in the model, we rewrite the utility function in an explicit way:

$$U_{nj} = \beta' x_{nj} + \varepsilon_{nj} / \sigma \quad (6)$$

Multiplying equation (6) by σ , we obtain¹⁸:

$$\tilde{U}_{nj} = \sigma \beta' x_{nj} + \varepsilon_{nj} \quad (7)$$

In practice, to identify β one usually normalizes σ to 1. A natural way to relax the assumption of homogeneity of variance of the error term is to allow the scale parameter to vary across individuals.

Thus, the utility function becomes:

$$\tilde{U}_{nj} = \sigma_n \beta' x_{nj} + \varepsilon_{nj} \quad (8)$$

Equation (8) is known as the scale heterogeneity logit model (S-MNL). Comparing equation (1) with (8), we can see that both mixed logit and S-MNL introduce individual-level heterogeneity into the model, but in different ways. In the mixed logit model, β_n follows a certain distribution, where $\beta_n = \beta + \eta_n$ and η_n is the deviation from the mean β ; in the S-MNL model, $\beta_n = \sigma_n \beta$ with heterogeneity acting multiplicatively through the scale parameter. Fiebig et al (2010) proposed an even more flexible model, the generalized multinomial logit (G-MNL), that nests the mixed logit and S-MNL (Keane & Wasi, 2012). In the G-MNL model,

$$\beta_n = \sigma_n \beta + [\gamma + \sigma_n(1 - \gamma)]\eta_n \quad (9)$$

where η_n is the individual-specific deviation from the mean, σ_n is the individual-specific scale of the error term that has a log-normal distribution with standard deviation τ and mean $\bar{\sigma}$ —specifically, $\sigma_n =$

¹⁸ In the logit model, what matters is the relative and not the absolute value of utility. Multiplying by or adding a constant to the utility function will not affect the ranking across alternatives, and thus does not affect estimated coefficients.

$\exp(\bar{\sigma} + \tau v_n)$, with v_n distributed standard normal—and γ is a parameter between 0 and 1 that controls how the taste heterogeneity varies with the scale parameter.¹⁹

In the first three columns of Appendix Table 4, we see that estimating a conditional logit, or re-estimating the mixed logit assuming the reform parameters to be lognormally distributed or allowing for all coefficients except the coefficient on wage to be normally distributed, does not significantly change our estimates of the willingness-to-pay for non-economic damages caps.²⁰ Column 4 reports the results from our estimation of the G-MNL model. The insignificant τ and small γ suggest that there is no strong evidence for the existence of scale heterogeneity. The estimated standard deviation of the reform taste parameter remains small and statistically insignificant. The willingness to pay for a cap on non-economic damages computed from the G-MNL model is \$11.37 per hour and thus like the other estimates in the table is similar to but slightly larger than the implied willingness to pay from model 1 in Table 3.

Sorting along physician human capital characteristics

In the remaining discussion of results, we focus on findings related to physician sorting along human capital characteristics. In Table 4, we consider the effects of non-economic damages caps on high-risk vs. not high-risk specialties (column 1); international vs. US medical graduates (column 2); and osteopaths vs. allopaths (column 3). Note that observations are weighted using the propensity score weights (discussed previously) designed to make observable characteristics similar across each sub-sample pair.²¹

¹⁹ When $\sigma_n = 1$, G-MNL is reduced to the mixed logit model. When $Var(\eta_n) = 0$, G-MNL becomes the S-MNL.

²⁰ We conducted a likelihood ratio test between the model shown in column 3 of Appendix Table 4 (all coefficients assumed to be normally distributed) and the model shown in column 1 of Table 3 (the malpractice reform coefficient assumed to be normally distributed and all other coefficients fixed). By the Bayesian Information Criterion, the specification assuming all coefficients are normally distributed is not superior to the main specification constraining non-reform parameters to be fixed. The BIC values are 69668.71 and 69654.93 for columns 3 and 1, respectively.

²¹ In Appendix C, we gauge in detail the success of the weighting approach in balancing the observed characteristics across each sub-sample pair. Tables C.1, C.2 and C.3 show characteristics of each sub-sample before and after propensity score weighting for the high risk vs. not high risk sub-samples (Table C.1), the international vs. US medical graduate sub-samples (Table C.2), and the osteopath vs. allopath sub-samples (Table C.3). These tables demonstrate that propensity score weighting nearly eliminates measurable differences in covariates across each sub-sample pair. Figures C.1, C.2 and C.3

The propensity score weighted findings in column 1 of Table 4 offer modest evidence that non-economic damages caps are a greater lure for high-risk specialists. The estimated coefficient on the damages cap law is positive, but small and imprecisely estimated. The estimated coefficient on the damages cap law interacted with the high-risk specialty indicator is 0.17 and marginally statistically different from zero. The implied estimate of the mean willingness-to-pay for reform among physicians without training in high-risk specialties is about \$7 ($.062/.009$) and for high-risk specialists it is about \$26 per hour ($.062+.168/.009$), which is a difference of \$19 per hour and this difference is marginally statistically significant (see last row of Table 4).

The findings in Table 4 are based on the definition of “high-risk” of Jena et al. (2011) that includes physicians whose specialties have higher-than-median claim size if a claim is awarded. We estimate the same models using two alternate definitions of high-risk; these results are shown in Appendix Table 5. In Appendix Table 5, the columns and their corresponding definitions of high-risk specialties are: (1) having both higher than median probability and higher than median award size in the past year (column 1); and (3) Klick & Stratmann’s definition of high-risk, which includes only the specialties having the top ten highest level of awards based on the Closed Claims Medical Malpractice data from Florida (column 2). These models are propensity-weighted in the same manner as our main results for high-risk vs. not high-risk specialists.

Appendix Table 5 shows that the estimated difference in WTP between high-risk and not high-risk specialists while positive in sign is not statistically different from zero when using a criterion that includes likelihood of a claim (column 1) or Klick & Stratmann’s definition.²² Thus, while our main

show that after propensity score weighting, the standardized differences in the means of each covariate are no longer statistically significant across each sub-sample pair, with only a few exceptions.

²² We have also estimated this equation without covariate balancing, to test whether our controlling for human capital quality explains why we do not find that high-risk specialists are more sensitive to reform, while Klick and Stratmann who do not include such controls do. While the WTP difference between high- and low-risk physicians becomes larger in an unweighted regression, it remains statistically insignificant. The WTP difference under the definition of high-risk employed in Table 4 becomes larger when we do not use our propensity score weights (compare columns 1 in Appendix Table 6 and Table 4). This suggests that some of the sensitivity of risk to reform reported in the literature is due to unmeasured differences in medical human capital quality.

results suggest that high-risk physicians respond more to non-economic damages cap laws than not high-risk physicians, this finding is sensitive to how one defines high-risk. Note that the definition of high-risk specialty employed in Table 4 includes pediatrics and pathology, two large specialties omitted from the definitions in Appendix Table 6. In other results not reported here, we have found pediatricians more responsive to cap laws than other physicians. Even though the frequency of a malpractice claims among pediatricians is relatively low, median award is relatively high and mean award is the highest among all specialties (Jena et al., figure 3) suggesting a higher likelihood compared to other specialties of very large awards.

Column 2 in Table 4 shows propensity score weighted findings for international vs. US medical graduates. Note that our analysis sample is limited to US citizens, thus international medical graduates are US citizens who attended medical school abroad. The estimated coefficient on non-economic damages cap law is small and not statistically different from zero. The coefficient estimated on the interaction term is large and statistically significant, so that the implied mean WTP for reform is about \$19 per hour higher for international medical graduates than for US medical school graduates, which is statistically different from zero by conventional standards of significance.

The propensity score weighted results for the osteopaths vs. allopaths are shown in the column 3 of Table 4. The estimated coefficient on non-economic damages cap is .09 and statistically insignificant. The estimated coefficient on the interaction term is large, about .56, and statistically significant implying a higher mean WTP for reform for osteopaths compared to allopaths of about \$41 per hour. This difference is fairly precisely estimated (the standard error of the mean difference is about 8.6). The estimated mean WTP to locate in a reformed state is more than three times larger for osteopaths compared to that of allopaths. The magnitude of the mean WTP for osteopaths would appear unreasonably high. Based on the propensity weighted data, these findings appear to be driven by osteopaths being more likely than allopaths to move to Florida and Texas, two states that passed non-economic damages caps during our analysis period (results not shown).

As a whole, then, columns 1 through 3 of Table 4 support the general idea that physicians' specialty risk and type of training (aspects of human capital) affect their responsiveness to non-economic damages cap laws – high-risk specialists, international medical graduates and osteopaths respond more to non-economic damages caps than physicians not trained in high-risk specialties, US medical graduates and allopaths.

In columns 4-6, we test whether responsiveness to non-economic damages caps varies by another aspect of medical human capital – the quality of the teaching hospital in which the residency was conducted, and the number of years of graduate medical education training. Column 4 in Table 4 singles out for comparison physicians whose hospital is at or below the median USNWR ranking, while column 5 shows results for sub-samples based on whether the hospital ranking factor score is at or below the median. Column 6 compares physicians who do not have and have sub-specialty training.²³

In both residency hospital quality regressions, estimates of the mean of the taste parameter for non-economic damages cap are negative, though not statistically significant. The estimates of the mean of the interaction effect are both positive and statistically significant. The implied mean of the WTP for physicians who trained in lower quality hospitals is between \$22 and \$24 per hour greater than it is for physicians who did not train in lower quality hospitals. These differences are marginally statistically significant.

In column 6 of Table 4, the estimate of the mean of the reform preference term for physicians with specialty training is negative but small in magnitude and statistically insignificant. The estimate of the mean interaction effect is positive but imprecisely estimated, implying a \$19 higher WTP for physicians with no sub-specialty training but the difference is not statistically different from zero. In

²³ Appendix tables C.4, C.5 and C.6 show the propensity score weighting nearly eliminates differences in covariates across the comparison samples used in Table 4 columns 4-6. Figures C.4, C.5, and C.6 show that after propensity score weighting, the standardized difference in the means of each covariate are no longer statistically significant across these subsamples, with only a few exceptions.

sum, our results indicate that where a physician trained affects her responsiveness to tort reforms – specifically, physicians who have received their training in high-risk specialties (depending on the definition), international medical schools, osteopathic medical schools, and in residencies in lower-ranked teaching hospitals—are more attracted to tort reform than their otherwise similar peers. While we find that the residency hospital’s ranking matters, sub-specialty training does not.

We also estimate the models of Table 4 without weights. These results are presented in Appendix Table 6. The qualitative pattern of findings is similar to the results with propensity score weighting, but the differences are instructive. The WTP premium demonstrated for high-risk specialty compared to non-high-risk specialty physicians is smaller and less precisely estimated in Table 4 compared to Appendix Table 6, at least in part because propensity score weighting controls for the fact that physicians in high-risk specialties are also more likely to be graduates of international medical schools. Conversely, the higher WTP difference observed for osteopaths compared to allopaths in Table 4 compared to Appendix Table 6, may be explained again by the weighting which controls for the fact that osteopaths are far less likely to be graduates of international medical schools. The much lower graduate rates of osteopaths from international medical schools is partly masking the greater preference for reform among osteopaths in Appendix Table 6.

6. Discussion and Conclusions

Spatial inequality in access to physicians and quality of care is a long-standing public health problem in the US (Chandra and Skinner, 2003). Physician groups argue that tort reform may be a useful policy tool to remediate this disparity, and that physicians’ location decisions can be influenced by changing the medical liability environment in which they practice medicine (AMA, 2019). Our findings are consistent with this idea -- the results suggest that new physicians entering their first clinical positions prefer to locate in reformed malpractice environments. We find that new physicians in high-risk specialties are more responsive to changes in tort laws than new physicians in lower-risk specialties.

More broadly, we find that physician human capital characteristics in general – beyond just medical specialty – influence responsiveness to state tort reforms. As revealed by their location choices, new physicians whose human capital characteristics may predispose them to more malpractice risk value non-economic damages caps more than their lower-risk peers. Because some of these human capital markers of lower quality training have been associated in the literature with worse patient outcomes or greater treatment expenses, this human capital effect may partly offset any improvement in access brought about by reform.²⁴

One limitation of this paper is we cannot rule out the possibility that state tort reforms are endogenous. States may pass tort reforms, for example, to address a problem with physicians leaving the state due to malpractice pressure, which would bias our willingness-to-pay estimates downward. Alternatively, passage of tort reform may be a proxy for physician political power in the state, or strong demand for physician services (Klick & Stratmann, 2007). This story implies a positive bias of our estimates of willingness-to-pay for reform in our base model, and in our estimate for willingness-to-pay for high-risk specialists vs. low-risk specialists, if the pressure for reform is coming from the surging political power of or demand for high-risk specialists. A reverse causation story explaining our finding that physicians with lower quality human capital have a higher willingness-to-pay is less straightforward. In addition, our measures of human capital only imperfectly capture vulnerability to malpractice risk. Ideally, we also would like to have measures of malpractice claims at the physician level, or more direct measures of physician ability, such as test scores, grades, or clinical outcomes.

²⁴ A well-documented aspect of the current opioid crisis is the geographical disparity in rates of overdoses and deaths. Public health researchers who study the crisis have begun to note that physicians vary in their opioid prescribing behavior (e.g. Schnell, 2018), and that there may be medical human capital antecedents of prescribing behavior (Schnell and Currie, 2018, report a strong relationship between physician opioid prescribing behavior and medical school rank which they argue is at least partly causal). Intriguingly, the cross-state correlation between an indicator variable indicating the presence of any reform and the opioid prescription rate is high. The simple correlation is .4 on 2007 and .6 in 2017 (based on CDC prescription data found here, <https://www.cdc.gov/drugoverdose/maps/rxrate-maps.html>, accessed June 24, 2019). An analysis carefully teasing out the various mechanisms behind these correlations may find that physician human capital sorting in response to reforms explains some of geographical heterogeneity in the opioid crisis.

With the passage of the Affordable Care Act in 2010, federal reform of malpractice laws is increasingly being viewed as part of a broader package of reforms to reduce waste and increase the cost-effectiveness of the health care system (Sage & Hyman, 2015). Findings from this paper will be helpful in informing this debate by providing recent information regarding how states' malpractice laws affect the sorting of new physicians across states, a perhaps unanticipated consequence of these laws.

References

- American Medical Association (AMA). 2019. Medial Liability Reform Now. <https://www.ama-assn.org/system/files/2019-03/mlr-now.pdf>, Accessed 6/19.
- Armstrong D, Chung R, Forte GJ. 2014 New York residency training outcomes: a summary of responses to the 2014 New York resident exit survey. Rensselaer, NY: Center for Health Workforce Studies, School of Public Health, SUNY Albany; 2015.
- Asch DA, Nicholson S, Srinivas S, Herrin J, Epstein AJ. Evaluating obstetrical residency programs using patient outcomes. *JAMA* 2009;302:1277-83.
- Avraham, Ronen. Database of state tort law reforms. Available at: <https://law.utexas.edu/faculty/ra22397/links/>. Accessed 6/7/19.
- Avraham, Ronen. 2007. "An empirical study of the impact of tort reforms on medical malpractice settlement payments." *Journal of Legal Studies* 36: S183-S229.
- Bard, J.S. 2011. "‘Practicing Medicine and Studying Law’: How Medical Schools Used to Have the Same Problems We Do and What We Can Learn from Their Efforts to Solve Them," 10 *Seattle Journal for Social Justice*, 135.
- Born, Patricia, Viscusi, W. Kip, and Baker, Tom. 2009. "The Effects of Tort Reform on Medical Malpractice Insurers’ Ultimate Losses." *The Journal of Risk and Insurance*, 76:1, pp. 197-219.
- Boulet, John R.; Cooper, Richard A.; Seeling, Stephen S.; Norcini, John J.; and McKinley, Danette W. 2009. "U.S. Citizens Who Obtain Their Medical Degrees Abroad: An Overview, 1992–2006." *Health Affairs*, 28:1, pp. 226-233.
- Busnaina, Ibrahim. 2001. "Pros and Cons of Applying to Foreign Medical Schools," US News & World Report, <https://www.usnews.com/education/blogs/medical-school-admissions-doctor/2011/08/01/pros-and-cons-of-applying-to-foreign-medical-schools>, accessed August 27, 2019.
- Cameron, A. Colin and Trivedi, Pravin K. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press, New York.
- Chandra, A. & Staiger, D. 2007. "Productivity spillovers in health care: Evidence from the treatment of heart attacks." *Journal of Political Economy* 115;1: 103-140.
- Chou, Chiu-Fang and Lo Sasso, Anthony T. 2009. "Practice location choice by new physicians: The importance of malpractice premiums, damage caps, and health professional shortage area designation." *Health Services Research*, 44: 1271-1289.
- Congressional Budget Office (CBO), July 2004. "The effects of tort reform: Evidence from the states." The Congress of the United States.

Courty, Pascal and Marschke, Gerald. "On the Sorting of Physicians across Medical Specialties: Understanding Shortages and Growth in Specialization," NBER Working Paper Series (No. 14502), November 2008.

Cuddy, Monica, Young, Aaron, Gelman, Andrew, Swanson, David B., Johnson, D.A., Dillon, Gerard F. & Clauser, Brian E. 2017. Exploring the relationships between USMLE performance and disciplinary action in practice: A validity study of score inferences from a licensure examination. *Academic Medicine*, 92: 1780-85.

Currie, Janet, and W. Bentley MacLeod. 2008. "First Do No Harm? Tort Reform and Birth Outcomes." *Quarterly Journal of Economics* 123:795–830.

Currie, Janet and Macleod, W. Bentley. 2017. "Diagnosing Expertise: Human Capital, Decision Making, and Performance among Physicians." *Journal of Labor Economics*, 35, issue 1, p. 1 - 43.

Currie, Janet, MacLeod, W. Bentley, & Van Parys, Jessica. 2016. "Provider practice style and patient health outcomes: The case of heart attacks." *Journal of Health Economics*, 47, p. 64-80.

Danzon, P. M., Epstein, A.J., Johnson, S.J. 2004. "The Crisis in Malpractice Insurance," in *Brookings-Wharton Papers on Financial Services*, eds: Herring, R. and Litan, R., Washington, DC: Brookings Institution Press, pp. 55-95.

Doyle, Joseph J., Ewer, Steven M., & Wagner, Todd H. 2010. "Returns to physician human capital: Evidence from patients randomized to physician teams." *Journal of Health Economics*, 29, p. 866-882.

Eckhert, N.L. & van Zanten, M. 2015. "US-Citizen international medical graduates – a boon for the workforce?" *New England Journal of Medicine* 372; 18: 1686-1687.

Epstein, A.J. & Nicholson, S. 2009. The formation and evolution of physician treatment styles: An application to cesarean sections. *Journal of Health Economics* 28: 1126-1140.

Farber, H.S. and White, M.J. 1994. "A Comparison of Formal versus Informal Dispute Resolution in Medical Malpractice," *Journal of Legal Studies*, 23(2): 777-806

Fiebig, Denzil, Keane, Michael, Louviere, Jordan and Wasi, Nada, (2010), The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity, *Marketing Science*, 29, issue 3, p. 393-421.

Gardner, L & Vishwasrao, S. 2010. Physician quality and health care for the poor and uninsured. *Inquiry* 47: 62-80.

Hartz, Arthur J., Kuhn, Evelyn M. & Pulido, J. 1999. "Prestige of training programs and experience of bypass surgeons as factors in adjusted patient mortality rates." *Medical Care*, 37(1): p. 93-103.

Haviland, A.M., Eisenberg, M.D., Mehrotra, A. Huckfeldt, P.J. & Sood, N. 2016. "Do "consumer-directed" health plans bend the cost curve over time?" *Journal of Health Economics* 46: 33-51.

- Harvard Medical Practice Study. 1990. Patients, Doctors, and Lawyers: Medical Injury, Malpractice Litigation, and Patient Compensation in New York. A report of the Harvard Medical Practice Study to the State of New York. Cambridge, MA: The President and Fellows of Harvard College.
- Helland, E. & Seabury, S.A. 2015. Tort reform and physician labor supply: A review of the evidence. *International Review of Law & Economics* 42: 192-202.
- Hole, A.R. 2008. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. *Journal of Health Economics*, 27: 1078-1094.
- Hyman, D.A., Silver, D.A., Black, B., & Paik, M. (2015). Does tort reform affect physician supply? Evidence from Texas. *International Review of Law and Economics*, 42: 203-218.
- Jena, A.B., Seabury, S., Lakdawalla, D. & Chandra, A. 2011. Malpractice risk according to physician specialty. *The New England Journal of Medicine* 365: 629-36.
- Jha, A. K., and A. M. Epstein. 2006. The Predictive Accuracy of the New York State Coronary Artery Bypass Surgery Report-Card System. *Health Affairs* 25(3):844–855.
- Keane, M., & Wasi, N. (2013). Comparing Alternative Models of Heterogeneity in Consumer Choice Behavior. *Journal of Applied Econometrics*, 28(6), 1018–1045. <https://doi.org/10.1002/jae.2304>
- Kessler, Daniel P. "Evaluating the medical malpractice system and options for reform." (2011) *The Journal of Economic Perspectives* ,25(2): (2011): 93-110.
- Kessler, Daniel, and Mark McClellan. 1996. "Do Doctors Practice Defensive Medicine?" *Quarterly Journal of Economics* 111:353–90.
- Kessler, Daniel, and Mark McClellan. 2002a. "How Liability Law Effects Medical Productivity." *Journal of Health Economics* 21:931–55.
- Kessler, Daniel, and Mark McClellan. 2002b. "Malpractice Law and Health Care Reform: Optimal Liability Policy in an Era of Managed Care." *Journal of Public Economics* 84:175–97.
- Kessler, Daniel P., William M. Sage, and David J. Becker. "Impact of malpractice reforms on the supply of physician services." (2005) *Jama* 293(21): 2618-2625.
- Kiker, B. F. and Zeh, Michael. 1998. Relative Income Expectations, Expected Malpractice Premium Costs, and Other Determinants of Physician Specialty Choice, *Journal of Health and Social Behavior*, Vol. 39, No. 2m pp. 152-167.
- Kilgore, M. L., Morrissey, M.A., and Nelson, L.J. 2006. "Tort Law and Medical Malpractice Insurance Premiums," *Inquiry*, 43: 255-270.
- Klick, Jonathan and Stratmann, Thomas. 2007. "Medical malpractice reform and physicians in high-risk specialties." *Journal of Legal Studies* 36: S121-S142.
- Lazear, Edward P. 2000. "Performance Pay and Productivity." *The American Economic Review* 90 (5): 1346-1361.

Lieber, Ethan M.J. 2014. "Medical malpractice reform, the supply of physicians, and adverse selection." *Journal of Law and Economics* 57: 501-527.

McCaffrey, D.F., Ridgeway, G., Morral, A.R. 2004. Propensity score estimation with boosted regression for evaluating casual effects in observational studies. *Psychological Methods* 9; 4: 403-425.

Malani, A and Reif, J. 2015. Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform. *Journal of Public Economics* 124: 1-17.

Matsa, David A. 2007. "Does malpractice liability keep the doctor away? "Evidence from tort reform damage caps." *Journal of Legal Studies* 36: S143-S182.

National Practitioner Data Bank Public Use Data File, 2018, U.S. Department of Health and Human Services, Health Resources and Services Administration, Bureau of Health Workforce, Division of Practitioner Data Bank.

Paik, Myungho and Black, Bernard S. and Hyman, David A., "Damage Caps and the Labor Supply of Physicians: Evidence from the Third Reform Wave." (2016). *American Law and Economics Review* 18(2): 463-505.

Pesko, M.F., Cea, M., Mendelsohn, J. & Bishop, T.F. 2017. The effects of malpractice non-economic damage caps on the supply of physician labor: Heterogeneity by physician age and risk. *International Review of Law and Economics* 50, 7-14.

Picard, Robert & Stepner, Michael. 2015. "GEO2XY: Stata module to convert latitude and longitude to xy using map projections," Statistical Software Components S457990, Boston College Department of Economics, revised 08 Jun 2017.

Sage, William M. and Hyman, David A. 2014. "Let's make a deal: Trading malpractice reform for health reform." *Health Affairs* 33: 53-58.

Schnell, M. 2018. Physician behavior in the presence of a secondary market: The case of prescription opioids. Working paper.

Schnell, M. and Currie, J. 2018. "Addressing the Opioid Epidemic: Is There a Role for Physician Education?" *American Journal of Health Economics*, 2018, 4(3): 383–410.

Seabury, Seth A., Helland, Eric, and Jena, Anupam B. 2014. "Medical Malpractice Reform: Noneconomic Damages Caps Reduced Payments 15 Percent, With Varied Effects by Specialty," *Health Affairs*, 33:11, pp. 2048–2056

Sloan, Frank A. 1990. "Experience Rating: Does It Make Sense for Medical Malpractice Insurance?" *American Economic Review*, 80(2): 128–33.

Sloan, Frank A. and Chepke, Lindsey, M. 2008. Medical Malpractice. The MIT Press, Cambridge, MA.

Studdert, David M. et al., May 11 2006. "Claims, errors, and compensation payments in medical malpractice litigation." *The New England Journal of Medicine*, 354: 2024-2033.

Train, Kenneth E., 1999. Halton sequences for mixed logit. Working paper.

Train, Kenneth E., 2009. *Discrete Choice Methods with Simulation*, Cambridge Books, Cambridge University Press.

Train, Kenneth & Weeks, Melvyn. 2005. “Discrete choice models in preference space and willingness-to-pay space.” Chapter 1 in Applications of simulation methods in environmental and resource economics, Alberini A. & Scarpa R. (Eds.), Kluwer Academic Publishers.

Wen,H.; C.Tang; H.Lin; C.Tsai; C.Chen; and C.Li. 2006. “Association between Surgeon and Hospital Volume in Coronary Artery Bypass Graft Surgery Outcomes: A Population-Based Study.” *Annals of Thoracic Surgery* 81:834–842.

White, Michelle. 1994. “The Value of Liability in Medical Malpractice.” *Health Affairs*, 13(4): 75-87.

Young, Aaron; Chaudhry, Humayun J.; Pei, Xiaomei; Arnhart, Katie; Dugan, Michael; and Snyder, Gregory B. 2017. “A Census of Actively Licensed Physicians in the United States.” *Journal of Medical Regulation*, 103(2): 7-21.

Table 1
Changes to states' tort reform laws, 1998-2017

State	(1) Cap on non-economic damages	(2) Cap on punitive damages	(3) Split recovery reform	(4) Collateral source reform	(5) Joint and several liability reform
AL		2000-		2001-	
AR		2003-2011			2003-
CA			2005-2006		
FL	2003-		1999- (turned off)		
GA	2005-2009				
ID		2004-			
IL	2006-2009				
ME	2000-	2000-			
MO	2013- (turned off)	2006-2014	2006-		
MS	2003-2012	2004-			
MT		2004-	2005-		
NC	2012-				
NV	2003-				
OH	2003-	2005-		2002-	2003-
OK	2004-			2004-	
OR	2000- (turned off)				
PA			2002-	2002-	2002-2005, 2011- 2006-
SC	2006-	2012-			
TN	2012-	2012-			
TX	2004-				
UT	2016- (turned off)				
WV				2003-	

Notes: The year indicated in column (1) is the year that the state “turned on” the non-economic damages cap with the exceptions of OR, UT and MO. These three states “turned off” (repealed) the law in the year indicated. The binary indicators for non-economic damages caps do not capture changes in the dollar amount of the cap – many states changed the amount of the cap during this time period or inflation-adjusted the caps while the law remained turned on during this time period. In the table, states that turned on and then turned off laws are indicated by showing the years in which the law was in effect.

Table 2
Descriptive Statistics

N=15,899	Mean	Std. Dev.
<i>Demographic Characteristics</i>		
Age (N=15,036)	33.374	4.288
Female	0.448	0.497
Went to medical school in NYS	0.335	0.472
Went to high school in NYS (N=15,677)	0.342	0.475
<i>Physician Human Capital Characteristics</i>		
Surgeon ¹	0.090	0.286
Emergency department physician	0.085	0.279
OB/GYN	0.068	0.252
High-risk specialist	0.480	0.500
International medical graduate	0.334	0.472
Osteopathic physician	0.102	0.303
No subspecialty training (N=13,513)	0.841	0.365
Number of graduate medical education training years (N=13,513)	4.535	1.235
Training hospital ranked ² below the top 32 (N=12,766)	0.354	0.478
<i>Tort Reforms</i> ³		
Cap on Non-economic Damages	0.204	0.403
Cap on Punitive Damages	0.307	0.461
Split Recovery Reform	0.043	0.202
Collateral Source Reform	0.866	0.341
Joint and Several Liability Reform	0.886	0.318
<i>State attributes</i>		
Physician real hourly wage ⁴	81.658	14.015
Physician density ⁵	4.418	6.300
Unemployment rate	6.020	1.824
Distance (in miles) ⁶	508.367	695.712

Notes:

1. Surgeon refers to physicians practicing in surgical specialties, including general surgery, cardio-thoracic surgery, neurological surgery, ophthalmology, orthopedic surgery, otolaryngology, plastic surgery, urology and other surgical subspecialties.
2. Ranking of physicians' training hospitals is based on 2019 US News and World report of NYS hospitals. A hospital in NY is either ranked (1 through 32) or unranked (implying a ranking of 33 or less).
3. Proportion of physicians that chose states with a certain type of malpractice reform is presented.
4. Physician state-level specialty-specific real hourly wage is computed, with 2010 set as the base year.
5. Physician density is defined as the number of physicians per 10,000 state population.
6. Distance refers to the distance between physician's training hospital and the center of the chosen state.

Table 3
Effects of tort reforms on physician location

	(1)	(2)	(3)	(4)	(5)	(6)
Cap on non-economic damages						
Mean	0.111** (0.055)					0.106* (0.059)
Standard deviation	0.191 (0.285)					0.252 (0.193)
Cap on punitive damages						
Mean		0.122 (0.092)				0.085 (0.106)
Standard deviation		0.110 (0.173)				0.031 (0.169)
Split recovery reform						
Mean			0.087 (0.083)			0.018 (0.165)
Standard deviation			0.030 (0.395)			0.464 (0.342)
Collateral source reform						
Mean				0.233** (0.118)		0.200 (0.141)
Standard deviation				0.548** (0.225)		0.485** (0.227)
Joint and several liability reform						
Mean					0.018 (0.087)	-0.155 (0.095)
Standard deviation					0.248 (0.245)	0.073 (0.243)
Real hourly wage	0.010*** (0.001)	0.010*** (0.001)	0.010*** (0.001)	0.010*** (0.001)	0.010*** (0.001)	0.010*** (0.001)
Physician density	0.007** (0.003)	0.007** (0.003)	0.007** (0.003)	0.007** (0.003)	0.007** (0.003)	0.008*** (0.003)
Unemployment rate	-0.003 (0.015)	-0.001 (0.015)	0.002 (0.015)	-0.001 (0.015)	0.0003 (0.015)	-0.001 (0.015)
Log (distance)	-0.233*** (0.036)	-0.233*** (0.037)	-0.233*** (0.041)	-0.236*** (0.041)	-0.233*** (0.041)	-0.236*** (0.041)
Log likelihood	-34557	-34559	-34559	-34556	-34559	-34552
WTP for cap on non-economic damages	10.831** (5.580)					
% reduction in wage	13.3%					
N physicians			15,899			
N observations			810,849			
LR test (col. 1 vs col. 6)			Chi sq stat: 10.05 p-value: 0.262			

Notes: Table shows results from mixed logit model. Each column represents a different model. Coefficients of the five medical malpractice reforms follow the normal distribution. Other coefficients are assumed to be fixed. The choice set for each physician is the 50 states and the District of Columbia. Each model includes alternative-specific constants (state fixed effects), with New York as the baseline category. * denotes statistically different from zero at the 0.10 level. ** denotes statistically different from zero at the 0.05 level. *** denotes statistically different from zero at the 0.01 level. Standard errors are in parentheses. Likelihood ratio test is based on model in col. 1 vs. model in col. 6.

Table 4

Interactions between medical human capital and tort reform

Model	(1) High-risk specialties	(2) International medical graduates	(3) Osteopath	(4) Low hospital ranking	(5) Low hospital quality index	(6) No sub- specialty training
Cap on NE Damages						
Mean	0.062 (0.067)	0.041 (0.071)	0.086 (0.064)	-0.048 (0.081)	-0.043 (0.078)	-0.054 (0.101)
Standard deviation	0.001 (0.295)	0.031 (0.177)	0.091 (0.112)	0.100 (0.348)	0.008 (0.380)	0.165 (0.777)
Cap on NE Damages \times Physician characteristics						
Mean	0.168* (0.093)	0.209** (0.094)	0.556*** (0.101)	0.208* (0.108)	0.193* (0.109)	0.154 (0.105)
Standard deviation	0.024 (0.291)	0.128 (0.172)	0.058 (0.116)	0.100 (0.578)	0.124 (0.546)	0.030 (0.407)
Real hourly wage	0.009*** (0.001)	0.011*** (0.001)	0.014*** (0.001)	0.009*** (0.001)	0.009*** (0.001)	0.008*** (0.001)
Physician density	0.008*** (0.003)	0.005* (0.003)	0.018*** (0.003)	0.007** (0.003)	0.006* (0.003)	0.011*** (0.003)
Unemployment rate	0.004 (0.015)	0.011 (0.015)	0.116*** (0.014)	-0.008 (0.016)	-0.009 (0.016)	-0.023 (0.017)
Log (distance)	-0.222*** (0.036)	-0.249*** (0.036)	-0.282*** (0.039)	-0.213*** (0.040)	-0.212*** (0.040)	-0.189*** (0.040)
Log-likelihood	-34265	-34149	-32777	-27774	-27757	-29361
N of physicians	15,899	15,899	15,899	12,766	12,757	13,513
Difference in WTP	19.082* (10.981)	18.517** (8.648)	40.844*** (8.454)	24.075* (13.126)	21.808* (12.881)	18.606 (12.995)

Notes: Each mixed logit model includes the individual-specific variable in the interaction term and uses weights from the generalized boosted model. The first two coefficients are random and follow the normal distribution. Other coefficients are assumed to be fixed. The choice set for each physician is the 50 states and the District of Columbia. Each model includes alternative-specific constants (state fixed effects), with New York as the baseline category. * denotes statistically different from zero at the 0.10 level. ** denotes

statistically different from zero at the 0.05 level. *** denotes statistically different from zero at the 0.01 level. Standard errors are in parentheses.

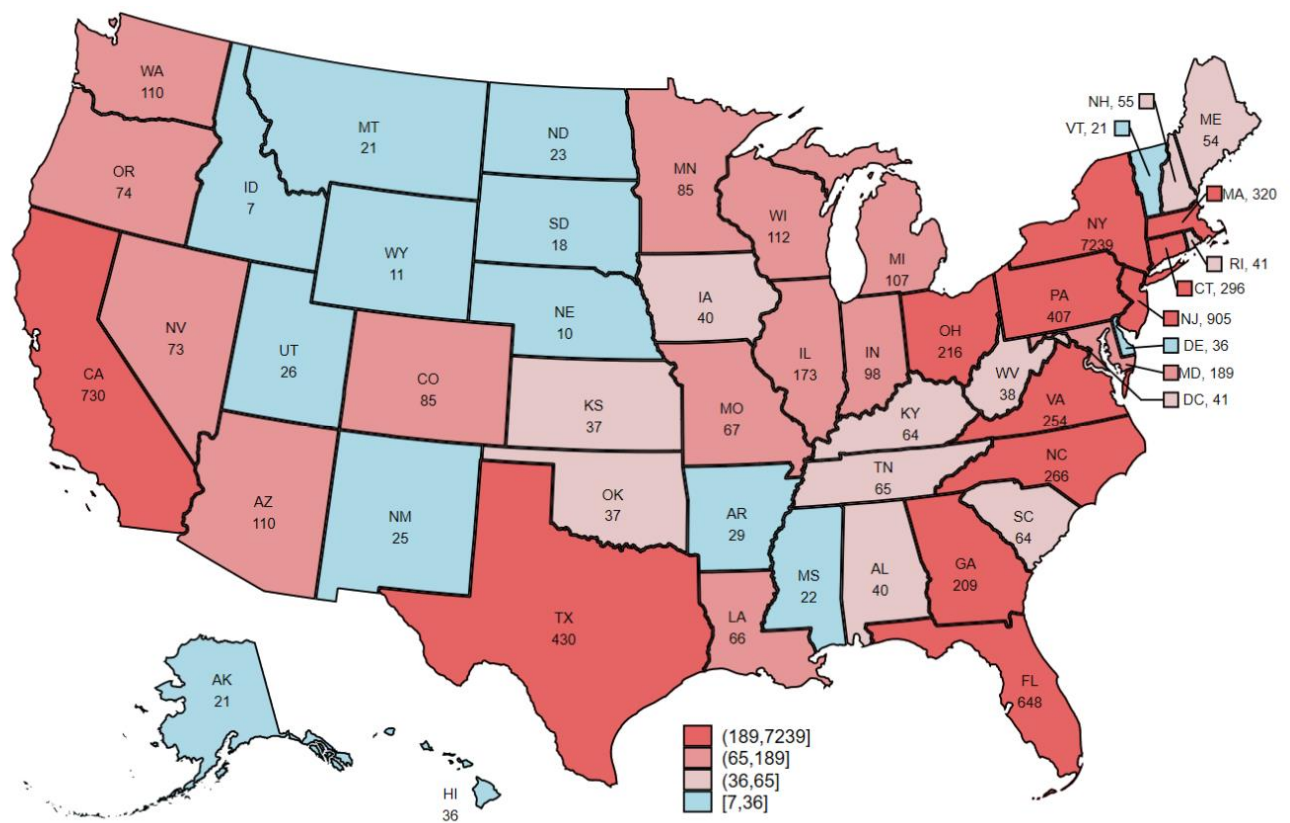


Figure 1. Location choices of new physicians trained in NYS. Each number shows how many physicians in our sample chose the corresponding state.

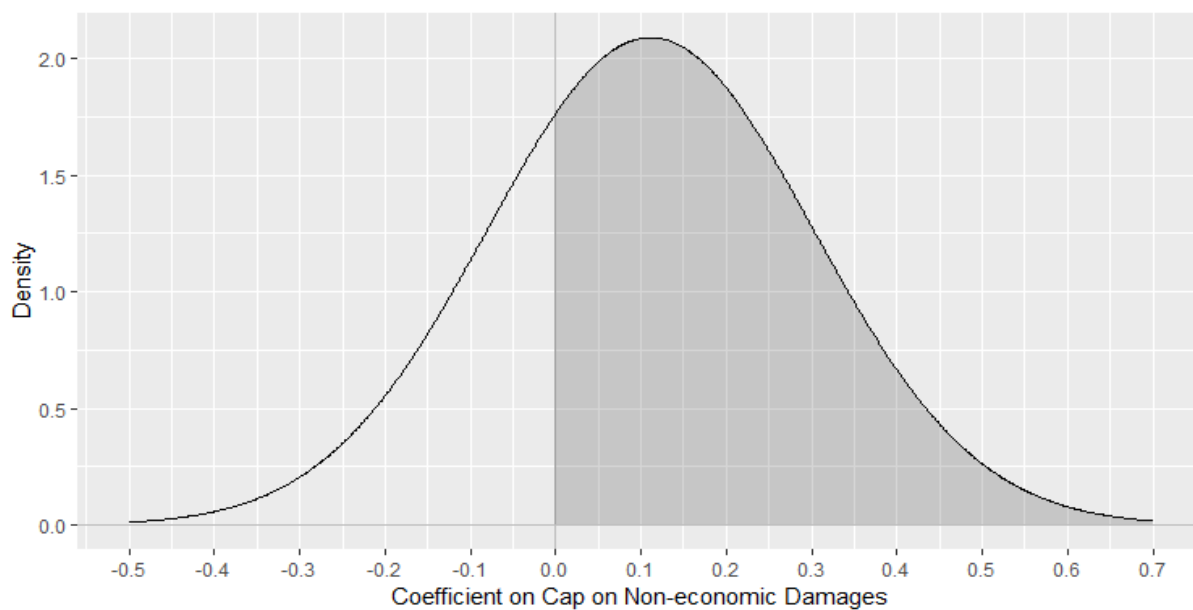


Figure 2.1. Distribution of coefficient on non-economic damages cap in full sample (from Table 3, column 1).

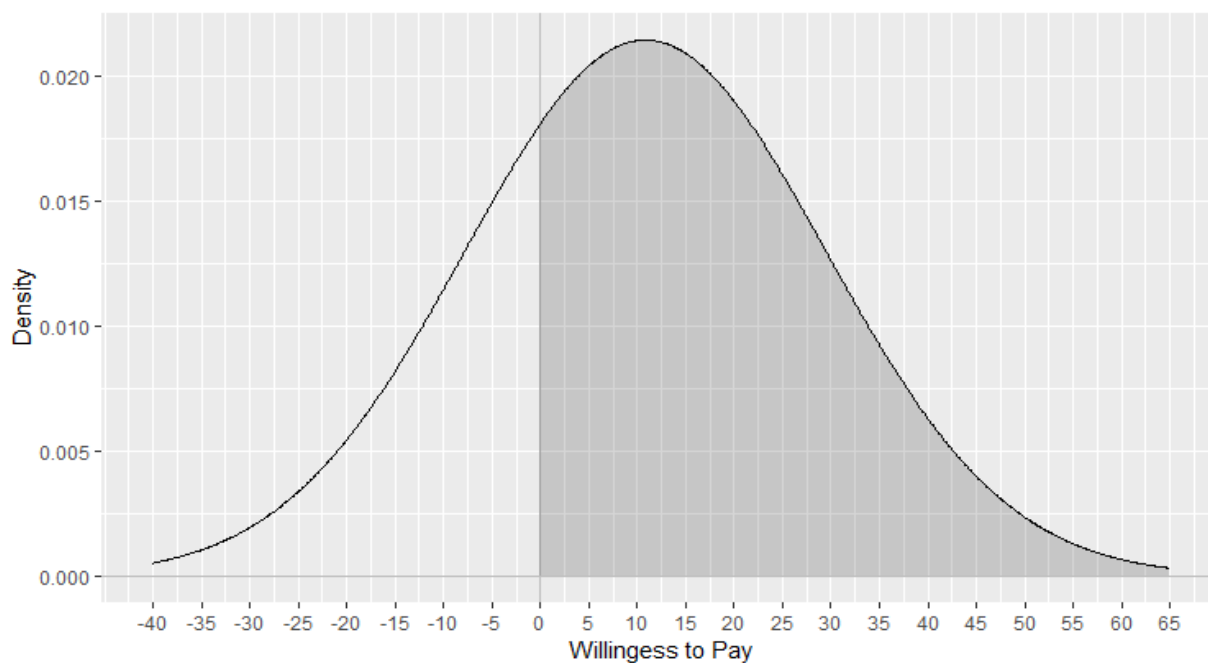


Figure 2.2. Distribution of willingness to pay for non-economic damages cap in full sample (from Table 3, column 1).

Table of Contents for Appendices

Table name	Content
<i>Appendix Tables Referenced in Text</i>	
Appendix Table 1	Common state medical malpractice reforms
Appendix Table 2	Diagnostic check results: conditional distributions of random parameters of malpractice reforms
Appendix Table 3	Models with leads and lags, 1998-2009
Appendix Table 4	Alternative specifications of main models
Appendix Table 5	Alternate definitions of high-risk specialty
Appendix Table 6	Unweighted results - physician human capital interactions analysis
<i>Self-contained Appendices</i>	
Appendix A	Malpractice and adverse outcomes: Osteopaths vs. Allopaths
Appendix B	Factor analysis of US News hospital rankings
Appendix C	Balance tables before and after propensity score weighting

Appendix Table 1
Common state medical malpractice reforms

Cap on non-economic damages	Statutory limit on non-economic damages awarded, which typically range from \$250,000 to \$500,000 and may be inflation-adjusted.
Cap on punitive damages	Statutory limit on punitive damages awarded.
Split recovery reform	Punitive damages awarded are to be split between the state and the plaintiff based on statutorily set percentages.
Collateral source reform	Collateral source rule permits plaintiff to recover damages from the defendant even if the plaintiff is also receiving damages from another party, such as an insurance company. Collateral source reform allows some or all of an award to be offset by payments from other parties.
Joint and several liability reform	Common law states that if more than one defendant is found guilty for a plaintiff's injuries or losses, then each defendant may be held 100% accountable. Joint and several liability reform modifies this law in various ways that apportion damages by each defendant's percentage of fault.

Sources: National Council of State Legislators, "Health Cost Containment and Efficiencies – Medical Malpractice Awards." NCSL Briefs for State Legislators, October 2011. Yu, H., Greenberg, M. & Haviland, A. (2011). "The impact of state medical malpractice reform on individual-level health care expenditures." Health Services Research 52(6): 2018-2034.

Appendix Table 2

Diagnostic check: conditional distributions of random parameters of malpractice reforms

	(1)	(2)	(3)	(4)	(5)	(6)
Cap on non-economic damages						
Mean	0.111					0.106
Standard deviation	0.191					0.251
Cap on punitive damages						
Mean		0.122				0.085
Standard deviation		0.110				0.031
Split recovery reform						
Mean			0.087			0.018
Standard deviation			0.030			0.462
Collateral source reform						
Mean				0.233		0.200
Standard deviation				0.539		0.478
Joint and several liability reform						
Mean					0.018	-0.155
Standard deviation					0.248	0.073

Notes: Estimates of parameters of conditional distributions are computed based on results from the mixed logit model (in Table 3). The mean presented here refers to the average of expectations of individual-specific parameters conditional on the physician's observed choice of location.

Appendix Table 3
Models with leads and lags, 1998-2009

	(1)	(2)	(3)
3 years before graduation			
Mean		-0.013 (0.110)	-0.001 (0.115)
Standard deviation		0.049 (0.491)	0.096 (0.489)
2 years before graduation			
Mean		-0.105 (0.146)	-0.110 (0.151)
Standard deviation		0.036 (0.468)	0.109 (0.475)
1 year before graduation			
Mean		-0.192 (0.245)	-0.210 (0.233)
Standard deviation		0.336 (0.726)	0.411 (0.527)
Cap on NE Damages			
Mean	0.113 (0.134)	0.297 (0.250)	0.179 (0.230)
Standard deviation	0.099 (1.928)	0.474 (0.615)	0.243 (0.702)
1 year after graduation			
Mean			0.305* (0.165)
Standard deviation			0.307 (0.488)
2 years after graduation			
Mean			-0.164 (0.136)
Standard deviation			0.010 (0.380)
3 years after graduation			
Mean			-0.078 (0.160)
Standard deviation			0.507 (0.418)
Real hourly wage	0.009*** (0.002)	0.010*** (0.002)	0.010*** (0.002)
Physician density	0.008* (0.004)	0.008* (0.004)	0.008* (0.004)
Unemployment rate	0.017 (0.022)	0.018 (0.023)	0.022 (0.023)
Log(distance)	-0.235*** (0.044)	-0.240*** (0.045)	-0.243*** (0.045)
Log-likelihood	-17607	-17607	-17603
N physicians		8,441	

Notes: The sample is limited to 2009 and earlier. The specification is the same as the one used in column 1 of Table 3. * denotes statistically different from zero at the 0.10 level. ** denotes statistically different from zero at the 0.05 level. *** denotes statistically different from zero at the 0.01 level. Standard errors are in parentheses.

Appendix Table 4
Alternative specifications of main models

Model	(1) Conditional logit model	(2) Mixed logit model with lognormal distribution	(3) Mixed logit model with 4 normally distributed parameters	(4) Generalized multinomial logit model
Cap on NE Damages				
Mean	0.121*** (0.009)	-2.111*** (0.388)	0.115** (0.052)	0.115** (0.048)
Standard deviation		0.015 (4.159)	0.035 (0.861)	0.020 (0.049)
Real hourly wage				
Mean	0.010*** (0.001)	-4.587*** (0.154)	0.010*** (0.001)	0.010*** (0.001)
Standard deviation		0.092 (0.829)		
Physician density				
Mean	0.007** (0.003)	0.007** (0.003)	0.007** (0.003)	0.007** (0.003)
Standard deviation			0.012 (0.011)	0.015** (0.007)
Unemployment rate				
Mean	-0.003 (0.015)	-0.003 (0.015)	-0.006 (0.016)	-0.006 (0.015)
Standard deviation			0.064 (0.120)	0.060 (0.095)
Log (distance)				
Mean	-0.232*** (0.036)	-0.232*** (0.036)	-0.199*** (0.051)	-0.194*** (0.049)
Standard deviation			0.519*** (0.106)	0.541*** (0.106)
Tau				
				0.281 (0.458)
gamma				
				0.687 (0.704)
WTP for reform	11.851** (4.805)	11.953** (5.197)	11.419** (5.305)	11.367** (4.990)
Log-likelihood	-34556	-34556	-34549	-34548
N of physicians	15,899	15,899	15,899	15,899

Notes: The full choice set for each physician is the 50 states and the District of Columbia. Each model includes alternative-specific constants (state fixed effects), with New York as the baseline category. Column 1 presents results of conditional logit model. Column 2 presents results of mixed logit model in which coefficients of malpractice reform and wage follow log normal distribution and all other coefficients are fixed. Column 3 presents results of mixed logit model assuming all the coefficients except coefficient on real hourly wage are normally distributed. Column 4 presents results of generalized multinomial logit model. * denotes statistically different from zero at the 0.10 level. ** denotes statistically different from zero at the 0.05 level. *** denotes statistically different from zero at the 0.01 level. Standard errors are in parentheses.

Appendix Table 5
Alternate definitions of high-risk specialty

	(1) Jena et al. (2011)	(2) Klick & Strattmann (2007)
Cap on NE Damages		
Mean	0.129* (0.067)	0.045 (0.066)
Standard deviation	0.019 (0.245)	0.031 (0.304)
Cap on NE Damages \times High-risk Specialists		
Mean	0.024 (0.091)	0.092 (0.097)
Standard deviation	0.006 (0.258)	0.153 (0.345)
Real hourly wage	0.008*** (0.001)	0.010*** (0.001)
Physician density	0.006* (0.003)	0.002 (0.003)
Unemployment rate	0.004 (0.015)	-0.008 (0.015)
Log(distance)	-0.177*** (0.037)	-0.245*** (0.035)
Log-likelihood	-34806	-34245
N Physicians	15,899	15,899
Difference in WTP	3.025 (11.436)	9.333 (9.943)

Notes: The specification is the same as the one used in Table 4 column 1. Each model uses propensity score weights and includes alternative-specific constants (state fixed effects), with New York as the baseline category. (1) Column 1 uses Jena's definition of high-risk specialties based on likelihood of malpractice suits and size of the claim, which include neurosurgery, neurology, obstetrics and gynecology, internal medicine, pulmonary medicine, general surgery, thoracic-cardiovascular surgery, cardiology. (2) Column 2 uses Klick and Strattmann's definition of high-risk specialties (Klick & Strattmann 2007), which include neurological surgery, thoracic surgery, obstetrics and gynecology, emergency room, plastic surgery, radiology, anesthesiology, general surgery and cardiovascular disease and general practice. * denotes statistically different from zero at the 0.10 level. ** denotes statistically different from zero at the 0.05 level. *** denotes statistically different from zero at the 0.01 level. Standard errors are in parentheses.

Appendix Table 6

Interactions between medical human capital and tort reform (unweighted estimates)

Model	(1) High-risk specialties	(2) International medical graduates	(3) Osteopath	(4) Low hospital ranking	(5) Low hospital quality index	(6) No sub- specialty training
Cap on NE Damages						
Mean	-0.005 (0.065)	0.007 (0.061)	0.114** (0.048)	-0.059 (0.080)	-0.052 (0.079)	0.022 (0.136)
Standard deviation	0.009 (0.293)	0.013 (0.269)	0.016 (0.292)	0.048 (0.372)	0.009 (0.393)	0.107 (0.406)
Cap on NE Damages \times Physician characteristics						
Mean	0.259*** (0.092)	0.245*** (0.092)	0.177 (0.174)	0.214* (0.127)	0.209* (0.114)	0.072 (0.146)
Standard deviation	0.011 (0.280)	0.011 (0.244)	0.076 (0.879)	0.286 (0.534)	0.161 (0.588)	0.040 (0.331)
Real hourly wage	0.010*** (0.001)	0.009*** (0.001)	0.010*** (0.001)	0.009*** (0.001)	0.009*** (0.001)	0.010*** (0.001)
Physician density	0.008*** (0.003)	0.002 (0.003)	0.008*** (0.003)	0.007** (0.003)	0.006* (0.003)	0.010*** (0.003)
Unemployment rate	-0.004 (0.015)	-0.005 (0.015)	-0.002 (0.015)	-0.005 (0.017)	-0.006 (0.016)	-0.041** (0.017)
Log (distance)	-0.230*** (0.036)	-0.228*** (0.036)	-0.229*** (0.036)	-0.208*** (0.040)	-0.208*** (0.040)	-0.239*** (0.041)
Log-likelihood	-34493	-34355	-34478	-27774	-27750	-29679
N of physicians	15,899	15,899	15,899	12,766	12,757	13,513
Difference in WTP	25.992*** (9.908)	26.038** (10.504)	17.458 (17.232)	24.052 (14.931)	23.555* (13.427)	7.047 (14.215)

Notes: Each model includes the individual-specific variable in the interaction term. The choice set for each physician is the 50 states and the District of Columbia. Each model includes alternative-specific constants (state fixed effects), with New York as the baseline

category. * denotes statistically different from zero at the 0.10 level. ** denotes statistically different from zero at the 0.05 level. *** denotes statistically different from zero at the 0.01 level. Standard errors are in parentheses.

Appendix A. Malpractice and adverse outcomes: Osteopaths vs. Allopaths

The National Practitioner Data Base (NPDB), a database administered by the U.S. Department of Health and Human Services, contains both medical malpractice award and non-malpractice “adverse action” information since September 1, 1990. Our copy of the database runs through September 30, 2018. The NPDB documents the location of these medical incidents and include information on the physician involved, including type of credential, M.D. or D.O. The malpractice information includes the size of the award. Non-malpractice adverse actions are comprehensive, spanning “clinical privileges actions, Federal and state licensure and certification actions, professional society membership actions, peer review organization actions, private accreditation actions, exclusions from Federal or state health care programs, and other adjudicated actions or decisions.” (NPDB Public Use Data File, documentation).

In Table A.1 the first and third columns show the average number of malpractice awards, total malpractice award amounts, and number of adverse actions by state-year by incident physician type. Note that the number of malpractice awards for a state and year is the number of malpractice suits that result in an award concerning incidents that took place within the state and during the focal year. The second and fourth columns show these amounts divided by the number of active non-Federal physicians involved in patient care by state-year. The period covered is 1998 to 2016, with some gap years. Our data include only those malpractice suits that have been resolved so we include only incidents that happened through 2016.

A comparison of columns 1 and 3 show that the total number of malpractice awards, the award amounts, and the number of adverse actions involving MDs is much higher than for DOs. For example, for the average state the number of malpractice awards against MDs was about 141 per year compared to 12 per year for DOs. Note that in all states MDs far outnumber DOs. We use the Area Resource File (2008 release) and the Area Health Resource File (2017-2018 release) which contains counts of practicing physicians by degree type by county and year. (These data are collected and maintained by the U.S. Department of Health and Human Services, Health Resources and Services Administration.) Columns 2 and 4, lines 1 and 2, show that the malpractice risk faced by MDs is lower than the risk faced by DOs, whether the risk is measured as the annual likelihood of a successful suit filed (1% vs. 1.5%) or as annual expected judgment (\$1652 vs \$2488). The per capita rate of adverse actions is twice as high for DOs than MDs (.016 vs. .008).

To test whether the malpractice rate is higher for osteopaths compared to allopaths, we calculated the difference between osteopaths and allopaths at the state-year level. Column 5 reports the mean differences and standard errors. The differences are statistically significant by conventional standards of significance. The malpractice award rate is approximately 0.3 percentage points (or 34%) higher for DOs. The average annual malpractice award per practitioner is about \$630 (or 38%) higher for DOs. The number of actions per osteopath is about .6 percentage points (or 75%) greater for DOs.

<p>Table A.1 Average Annual Malpractice Awards and Adverse Actions Per Physician Type 1998-2016 50 States and Washington D.C.</p>					
	D.O.		M.D.		Difference (Per capita)
State-Year Average	(1) Total (S.D.)	(2) Per capita (S.D.)	(3) Total (S.D.)	(4) Per capita (S.D.)	(5) D.O.-M.D. (S.E.)
Malpractice award (count)	11.45 (20.02)	0.015 (.016)	140.53 (232.02)	0.010 (.008)	0.0034 (.0010)
Malpractice award (in 1982-84 \$)	1,865,030.46 (3463452.14)	2,487.51 (3913.80)	23,829,877.67 (44667573.59)	1,652.16 (1337.164)	630.05 (159.54)
Adverse actions (count)	11.70 (15.98)	0.016 (.015)	102.64 (126.25)	0.008 (.003)	0.0063 (.0007)
<p>Table reports averages for years 1998, 2001, 2003, 2004, 2007, 2010-2016, for all states including Washington, D.C. The counts used for MDs and DOs are those in patient care who are not employed by the federal government.</p> <p>Column 5 shows the difference in malpractice judgment rates and per capita awards and adverse action rates, respectively, based on comparisons within state-years. The standard errors of these differences are provided (in parentheses). All differences are all highly statistically significant by conventional standards.</p>					

Appendix B: Factor analysis of US News hospital rankings¹

We use the U.S. News and World Report (USNWR) rankings of U.S. hospitals. These rankings assess U.S. hospitals by specialty and by “procedures and conditions” and overall.²

Adult specialties: USNWR evaluates hospitals based on their performance in 16 specialties, 12 inpatient and four outpatient specialties. According to the USNWR these 16 specialties reflect how adept hospitals are with complex, high-acuity cases. For the inpatient specialties, the USNWR bases its rankings on patient mortality rates (adjusted by patient characteristics), reputation (based on surveys of board-certified physicians), patient safety (e.g., rates of postoperative hemorrhage or hematoma), and the hospital’s inputs (e.g., nurse staffing). For each of these specialties, an eligible hospital is either ranked one through 50 or unranked, implying a rank outside the top 50.³ Hospitals ranked outside the top 50 in a specialty but placing in the top 10 percent of the scores was designated high performing on the specialty.

The rankings of the four specialties that are primarily outpatient are based on surveys of experts, as they entail low risks of death and safety concerns. Hospitals that receive nominations by at least five percent of physician experts are nationally ranked on the specialty. Hospitals that are nominated by at least three percent but less than five percent are high performing on the specialty.

Children’s specialties: USNWR ranks children’s hospitals and the pediatric centers within major medical centers. These hospitals are ranked based on performance in 10 pediatric specialties.⁴ The specialties are assessed on the basis of clinical outcomes, such as survival rates, infection and complication rates, hospital inputs, procedures (e.g., whether the hospital has a program to prevent infections), and the opinions among pediatric specialists. USNWR publishes how many pediatric specialties in a children’s hospital or hospital with a pediatric center rank in the top 50.⁵

Procedures and conditions: The goal of the procedure and conditions rating is to evaluate how well hospitals treat the full range of patients, not just the high-acuity cases. This rating focusses on common procedures and conditions (e.g., hip replacement and chronic obstructive pulmonary disease) that most U.S. hospitals perform/treat. The ratings are based on hospital’s outcomes (e.g., readmission rates) adjusted for patient characteristics, hospital inputs, patient satisfaction scores, and data transparency (whether hospitals disclose their performance data to patients). The ratings aggregate the measures in each cohort of care into an overall assessment by placing a hospital into one of three composite bands: high performing, average and below average. We use the number of high performing procedures and conditions because the numbers of average and low performing procedures and conditions are difficult to collect from the USNWR website.

¹ This section is taken from USNWR documentation on their Best Hospital rankings: Comarow, 2018, Olmsted et al. 2018a, Olmsted et al. 2018b, and George et al. 2018.

² “U.S. News & World Report’s Best Hospitals” data, copyright ©2018 U.S. News & World Report, L.P., used with permission from U.S. News.

³ The 12 data-based specialties are Cancer, Cardiology and Heart Surgery, Diabetes and Endocrinology, Ear, Nose and Throat, Gastroenterology and GI Surgery, Geriatrics, Gynecology, Neurology and Neurosurgery, Nephrology, Orthopedics, Pulmonology, and Urology. The four reputation-based specialties are Ophthalmology, Psychiatry, Rehabilitation, and Rheumatology.

⁴ These specialties are Cancer, Cardiology & Heart Surgery, Diabetes & Endocrinology, Gastroenterology & GI Surgery, Neonatology, Nephrology, Neurology & Neurosurgery, Orthopedics, Pulmonology and Urology.

⁵ Thus, these rankings are picking up both whether the hospital is or has a dedicated children’s medical center and how well it performs on these specialties. Few hospitals qualify as a child-specializing hospital. USNWR only identifies and therefore ranks 189 such hospitals nation-wide for their 2018-2019 rankings.

Overall rankings: We also used the USNWR hospital rankings for New York State as a measure of performance. Any of the New York hospitals among the top 20 hospitals in the U.S., the USNWR Best Hospitals Honor Roll, appear on the NYS regional rankings at the top of the regional rankings in order of their ranking among the top 20. The USNWR Hospital Honor Roll is based on adult specialty rankings and procedure and condition ratings. In New York State, there were three hospitals among the top 20: New York-Presbyterian Hospital-Columbia and Cornell (No. 10), NYU Langone Hospitals (No. 15), and Mount Sinai Hospital (no. 18).

The New York hospitals on the Honor Roll appear at the top of the New York State ranking in the order they rank in the Honor Roll. The remaining NYS regional rankings include all hospitals that received a national ranking on at least one of the 12 data-determined specialties *or* received at least three high performing designations on the 12 data-determined specialties and nine Best Hospital procedures and conditions are designated Best Regional Hospital within New York state.

In the NY state ranking, a hospital on the Best Hospitals Honor Roll outranked all other hospitals that were not on the Honor Roll regardless of point totals. Other hospitals located in each region were ranked according to the number of points they earned, where points are allocated for the twelve data-driven Best Hospitals specialties in which they were nationally ranked and the twelve specialties and nine procedures and conditions in which they were rated high performing. They are docked points for procedures or conditions in which they were rated below average.

Coverage overlap: While our assessments are based on the USNWR 2018-2019 Best Hospitals rankings, the data upon which the 2018-2019 rankings and ratings were created overlap temporally with the period of our study. Mortality rates and safety measures are based on FY 2014, 2015, and 2016 data. Hospital's process or input data (e.g., volume (discharge), nurse staffing) is based on 2014, 2015, 2016 data or on 2016 data depending on the variable. Reputation-based specialties based on surveys conducted in 2016, 2017, and 2018. Data for Procedures and Conditions comes from 2011-2016 data, though the patient subjective experience data was collected in 2016-2017.

Interpretation: USNWR rankings of the hospitals where physicians trained capture the quality of the physicians in two ways. Better hospitals attract better resident-applicants and their reputation allows them to be more selective. Training hospitals may pass along its practices to residents and hospitals with better patient outcomes may pass along behaviors that raise the performance after the physician leaves.

The table below shows the correlation matrix for the ranking and performance measures and the overall ranking. NY state rankings are based on performance on adult specialties and adult procedures and conditions. Nonetheless, the rankings are not positively correlated with these or the number of nationally ranked child specialties. On the other hand, aside from the correlation between number of nationally ranked and number of high performing adult specialties, the non-NYS ranking measures are correlated: correlation ranges between .35 (number of nationally ranked child specialties and of high performing adult specialties) and .76 (number of nationally ranked adult specialties and number of high performing procedures and conditions). These correlations show that there is information in these constituent metrics that is not reflected in the overall rankings and that they may be picking up a common feature of hospitals.

Table B.1 Correlation Matrix N=1880					
	NYS rank	Nationally ranked adult specialties	Nationally ranked child specialties	High performing Adult specialties	High-performing Procedures and conditions
NYS hospital rank	1				
No. of nationally ranked adult specialties	-0.2882	1			
No. of nationally ranked child specialties	-0.1419	0.7006	1		
No. of high performing adult specialties	0.1039	0.0472	0.3522	1	
No. of high-performing procedures and conditions	0.0429	0.7596	0.5523	0.3700	1

Constructing an index: We use factor analysis to identify the combination of non-NYS rank measures that best characterizes this common factor. Table B.2 reports the results from the factor analysis of the four specialty and procedure and condition measures. Note that the first factor explains 85 percent of the variation. Table B.3 below shows the extent by which the first factor loads on each of the measures. The loading on the number of ranked adult specialties is highest at .87, with the loadings on ranked child specialties (.77) and procedures and conditions (.83) also high. The loading on the number of highly performing adult specialties is the lowest (.36) but also positive. We use these factor loadings to create a single specialty and condition index for each residency of each physician.

Table B.2 Factor Analysis of USNWR metrics				
Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	2.16746	1.68205	0.8449	0.8449
Factor2	0.48541	0.37754	0.1892	1.0341
Factor3	0.10787	0.30327	0.0420	1.0762
Factor4	-0.19540	.	-0.0762	1.0000

Table B.3 Factor loadings	
Variable	Factor 1 loading
No. of nationally ranked adult specialties	0.8662
No. of nationally ranked child specialties	0.7701
No. of high performing adult specialties	0.3605
No. of high-performing procedures and conditions	0.8332

Appendix C: Balance tables before and after propensity score weighting

Table C.1

Specialty risk

Covariates	Unweighted			Propensity Score Weighting		
	High-risk	Non high-risk	p-value	High-risk	Non high-risk	p-value
International medical graduates	0.403 (0.491)	0.271 (0.444)	0.000	0.336 (0.472)	0.336 (0.472)	0.965
Attended medical school in NYS	0.324 (0.468)	0.344 (0.475)	0.008	0.335 (0.472)	0.333 (0.471)	0.846
Osteopath	0.114 (0.318)	0.091 (0.288)	0.000	0.102 (0.302)	0.100 (0.300)	0.784
Female	0.515 (0.500)	0.387 (0.487)	0.000	0.449 (0.497)	0.449 (0.497)	0.978
Age	33.4 (4.781)	33.7 (4.418)	0.000	33.528 (4.587)	33.531 (4.573)	0.970
Missing age	0.045 (0.206)	0.046 (0.210)	0.600	0.045 (0.207)	0.045 (0.208)	0.928
Age<=25	0.001 (0.028)	0.000 (0.000)	0.013	0.000 (0.021)	0.000 (0.000)	0.016
Age>=50	0.008 (0.090)	0.008 (0.093)	0.690	0.008 (0.090)	0.008 (0.091)	0.961
Solo	0.017 (0.119)	0.027 (0.154)	0.000	0.021 (0.136)	0.021 (0.137)	0.928
Missing solo	0.115 (0.319)	0.109 (0.312)	0.236	0.113 (0.317)	0.113 (0.317)	0.998
Number of physicians	7,628	8,271		7,628	8,271	

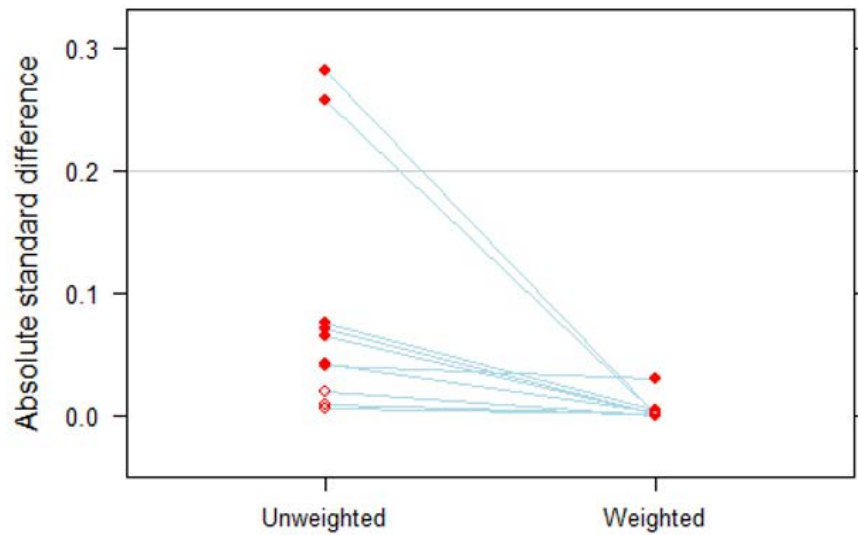


Figure C.1 High-risk vs. non-high risk. The standard difference is defined as the difference between the treatment group mean and the control group mean divided by the pooled sample standard deviation. Filled circles indicate that difference is statistically significant. All differences are not significant after propensity score weighting except for age less than 25.

Table C.2

International and U.S. medical graduates

Covariates	Unweighted			Propensity Score Weighting		
	International medical graduates	U.S. medical graduates	p-value	International medical graduates	U.S. medical graduates	p-value
Osteopath	0.003 (0.055)	0.152 (0.359)	0.000	0.089 (0.285)	0.103 (0.304)	0.612
Female	0.417 (0.493)	0.464 (0.499)	0.000	0.452 (0.498)	0.450 (0.497)	0.900
Family and General Practice	0.081 (0.273)	0.061 (0.239)	0.000	0.052 (0.223)	0.068 (0.251)	0.002
Internists	0.471 (0.499)	0.235 (0.424)	0.000	0.311 (0.463)	0.313 (0.464)	0.916
OB/GYN	0.043 (0.203)	0.081 (0.273)	0.000	0.069 (0.253)	0.069 (0.253)	0.990
Pediatricians	0.071 (0.257)	0.065 (0.246)	0.137	0.065 (0.246)	0.067 (0.249)	0.794
Psychiatrists	0.089 (0.284)	0.053 (0.225)	0.000	0.069 (0.254)	0.065 (0.246)	0.588
Surgeons	0.024 (0.154)	0.123 (0.328)	0.000	0.081 (0.273)	0.091 (0.287)	0.319
All other specialties	0.145 (0.352)	0.302 (0.459)	0.000	0.274 (0.446)	0.251 (0.433)	0.278
Age	35.729 (5.337)	32.421 (3.713)	0.000	33.573 (4.488)	33.481 (4.524)	0.456
Missing age	0.063 (0.243)	0.037 (0.188)	0.000	0.055 (0.229)	0.045 (0.207)	0.235
Age<=25	0.000 (0.019)	0.000 (0.019)	0.995	0.000 (0.018)	0.000 (0.017)	0.893
Age>=50	0.018 (0.132)	0.004 (0.061)	0.000	0.008 (0.089)	0.007 (0.083)	0.574
Solo	0.024 (0.147)	0.021 (0.134)	0.171	0.019 (0.129)	0.021 (0.135)	0.430
Missing solo	0.103 (0.304)	0.117 (0.322)	0.006	0.110 (0.313)	0.112 (0.315)	0.862
Number of physicians	5,318	10,581		5,318	10,581	

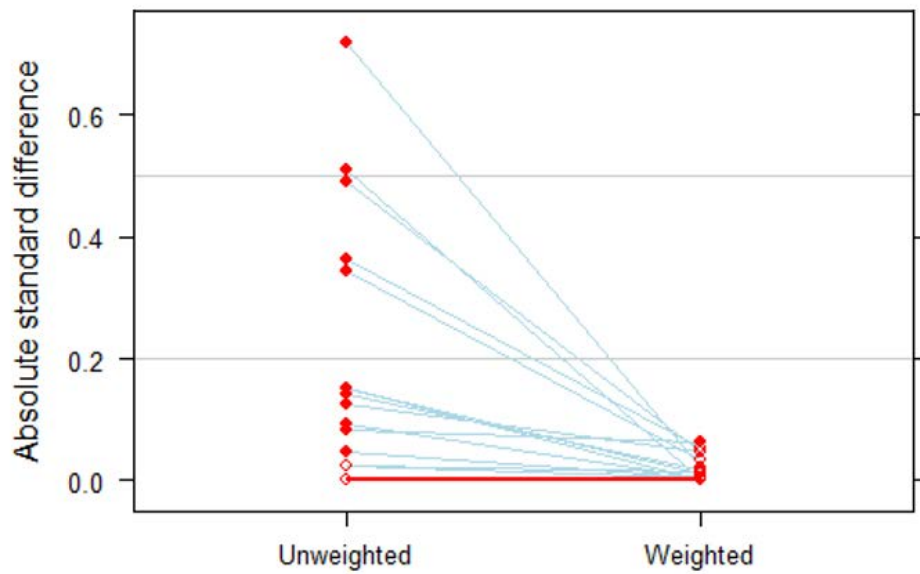


Figure C.2 International vs. US medical graduate. The standard difference is defined as the difference between the treatment group mean and the control group mean divided by the pooled sample standard deviation. Filled circles indicate that difference is statistically significant. All differences are not significant after propensity score weighting except for number of physicians in family and general practice.

Table C.3
Osteopaths and allopaths

Covariates	Unweighted			Propensity Score Weighting		
	Osteopaths	Allopaths	p-value	Osteopaths	Allopaths	p-value
International medical graduates	0.010 (0.099)	0.371 (0.483)	0.000	0.293 (0.455)	0.335 (0.472)	0.434
Attended medical school in NYS	0.545 (0.498)	0.311 (0.463)	0.000	0.355 (0.479)	0.335 (0.472)	0.491
Female	0.500 (0.500)	0.442 (0.497)	0.000	0.453 (0.498)	0.448 (0.497)	0.917
age	32.661 (4.391)	33.626 (4.611)	0.000	33.135 (3.788)	33.515 (4.576)	0.280
Missing age	0.043 (0.203)	0.046 (0.209)	0.633	0.068 (0.251)	0.045 (0.208)	0.355
Age<=25	0.001 (0.035)	0.000 (0.017)	0.280	0.000 (0.021)	0.000 (0.017)	0.668
Age>=50	0.006 (0.078)	0.009 (0.093)	0.230	0.001 (0.035)	0.008 (0.090)	0.000
Solo	0.022 (0.139)	0.022 (0.138)	0.935	0.014 (0.109)	0.022 (0.138)	0.013
Missing solo	0.093 (0.291)	0.114 (0.318)	0.006	0.102 (0.302)	0.113 (0.316)	0.632
Family and general practice Internists	0.199 (0.400)	0.052 (0.223)	0.000	0.083 (0.276)	0.067 (0.250)	0.561
	0.213 (0.409)	0.326 (0.469)	0.000	0.269 (0.443)	0.315 (0.464)	0.225
OB/GYN	0.075 (0.264)	0.068 (0.251)	0.260	0.066 (0.249)	0.068 (0.252)	0.881
Pediatricians	0.085 (0.278)	0.065 (0.247)	0.007	0.053 (0.223)	0.067 (0.250)	0.116
Psychiatrists	0.040 (0.196)	0.068 (0.252)	0.000	0.057 (0.232)	0.065 (0.247)	0.680
Surgeons	0.039 (0.195)	0.096 (0.294)	0.000	0.082 (0.274)	0.090 (0.286)	0.493
All other specialties	0.269 (0.444)	0.247 (0.431)	0.058	0.285 (0.452)	0.250 (0.433)	0.283
Number of physicians	1,621	14,278		1,621	14,278	

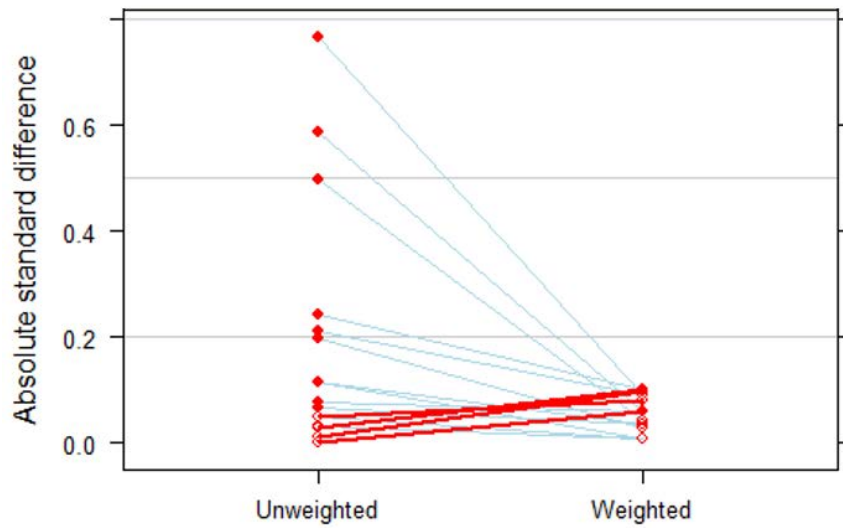


Figure C.3 Osteopath vs. allopath. The standard difference is defined as the difference between the treatment group mean and the control group mean divided by the pooled sample standard deviation. Filled circles indicate that difference is statistically significant. All differences are not significant after propensity score weighting except for age more than 50 and solo practice.

Table C.4

Training hospital's US News ranking

Covariates	Unweighted			Propensity Score Weighting		
	At or Below median	Above median	p-value	At or Below median	Above median	p-value
International medical graduates	0.338 (0.473)	0.321 (0.467)	0.046	0.330 (0.470)	0.329 (0.470)	0.944
Attended medical school in NYS	0.325 (0.468)	0.350 (0.477)	0.002	0.337 (0.473)	0.338 (0.473)	0.929
Osteopath	0.108 (0.311)	0.097 (0.296)	0.034	0.103 (0.303)	0.103 (0.304)	0.946
Female	0.440 (0.496)	0.451 (0.498)	0.219	0.445 (0.497)	0.445 (0.497)	0.932
Family and General Practice	0.073 (0.260)	0.063 (0.242)	0.019	0.067 (0.251)	0.068 (0.251)	0.944
Internists	0.314 (0.464)	0.316 (0.465)	0.800	0.315 (0.465)	0.316 (0.465)	0.914
OB/GYN	0.068 (0.252)	0.068 (0.252)	0.954	0.068 (0.251)	0.068 (0.252)	0.948
Pediatricians	0.066 (0.248)	0.065 (0.246)	0.772	0.065 (0.247)	0.065 (0.247)	0.923
Psychiatrists	0.063 (0.244)	0.064 (0.245)	0.843	0.063 (0.243)	0.063 (0.243)	0.998
Surgeons	0.091 (0.287)	0.095 (0.293)	0.438	0.093 (0.290)	0.093 (0.291)	0.927
All other specialties	0.246 (0.430)	0.256 (0.436)	0.187	0.250 (0.433)	0.251 (0.434)	0.931
Age	33.592 (4.700)	33.440 (4.514)	0.063	33.504 (4.586)	33.504 (4.589)	0.998
Missing age	0.046 (0.211)	0.043 (0.202)	0.308	0.045 (0.207)	0.044 (0.206)	0.912
Age<=25	0.001 (0.025)	0.000 (0.013)	0.190	0.001 (0.023)	0.000 (0.017)	0.491
Age>=50	0.010 (0.097)	0.007 (0.086)	0.185	0.008 (0.090)	0.008 (0.090)	0.932
Solo	0.021 (0.135)	0.023 (0.142)	0.385	0.022 (0.138)	0.022 (0.138)	0.973
Missing solo	0.112 (0.316)	0.114 (0.318)	0.759	0.112 (0.316)	0.112 (0.316)	0.962
Number of physicians	6,520	6,246		6,520	6,246	

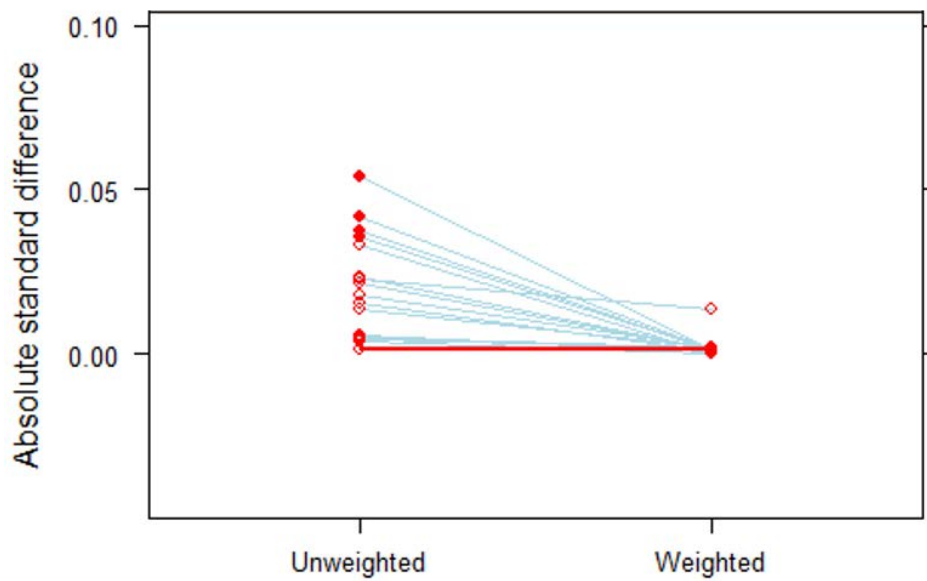


Figure C.4 Low vs. high hospital ranking. The standard difference is defined as the difference between the treatment group mean and the control group mean divided by the pooled sample standard deviation. Filled circles indicate that difference is statistically significant. All differences are not significant after propensity score weighting.

Table C.5

Training hospital's quality index based on factor analysis

Covariates	Unweighted			Propensity Score Weighting		
	At or Below median	Above median	p-value	At or Below median	Above median	p-value
International medical graduates	0.338 (0.473)	0.321 (0.467)	0.048	0.330 (0.470)	0.329 (0.470)	0.921
Attended medical school in NYS	0.324 (0.468)	0.350 (0.477)	0.002	0.337 (0.473)	0.338 (0.473)	0.970
Osteopath	0.108 (0.310)	0.098 (0.297)	0.058	0.103 (0.304)	0.103 (0.304)	0.968
Female	0.439 (0.496)	0.451 (0.498)	0.165	0.444 (0.497)	0.445 (0.497)	0.913
Family and General Practice	0.073 (0.259)	0.063 (0.243)	0.036	0.067 (0.251)	0.068 (0.252)	0.936
Internists	0.314 (0.464)	0.316 (0.465)	0.862	0.316 (0.465)	0.316 (0.465)	0.999
OB/GYN	0.068 (0.252)	0.068 (0.252)	0.935	0.068 (0.251)	0.068 (0.252)	0.917
Pediatricians	0.066 (0.248)	0.065 (0.247)	0.913	0.065 (0.247)	0.066 (0.247)	0.937
Psychiatrists	0.063 (0.243)	0.065 (0.246)	0.752	0.063 (0.243)	0.063 (0.243)	0.958
Surgeons	0.092 (0.288)	0.094 (0.292)	0.610	0.093 (0.290)	0.093 (0.291)	0.947
All other specialties	0.246 (0.431)	0.255 (0.436)	0.242	0.250 (0.433)	0.251 (0.434)	0.908
Age	33.583 (4.690)	33.450 (4.529)	0.099	33.504 (4.585)	33.503 (4.584)	0.990
Missing age	0.046 (0.210)	0.043 (0.203)	0.348	0.045 (0.207)	0.044 (0.206)	0.910
Age<=25	0.001 (0.025)	0.000 (0.013)	0.185	0.001 (0.023)	0.000 (0.016)	0.444
Age>=50	0.009 (0.097)	0.007 (0.086)	0.214	0.008 (0.090)	0.008 (0.090)	0.946
Solo	0.021 (0.135)	0.023 (0.142)	0.396	0.022 (0.138)	0.022 (0.138)	0.918
Missing solo	0.111 (0.314)	0.116 (0.320)	0.410	0.112 (0.316)	0.113 (0.316)	0.931
Number of physicians	6,447	6,310		6,447	6,310	

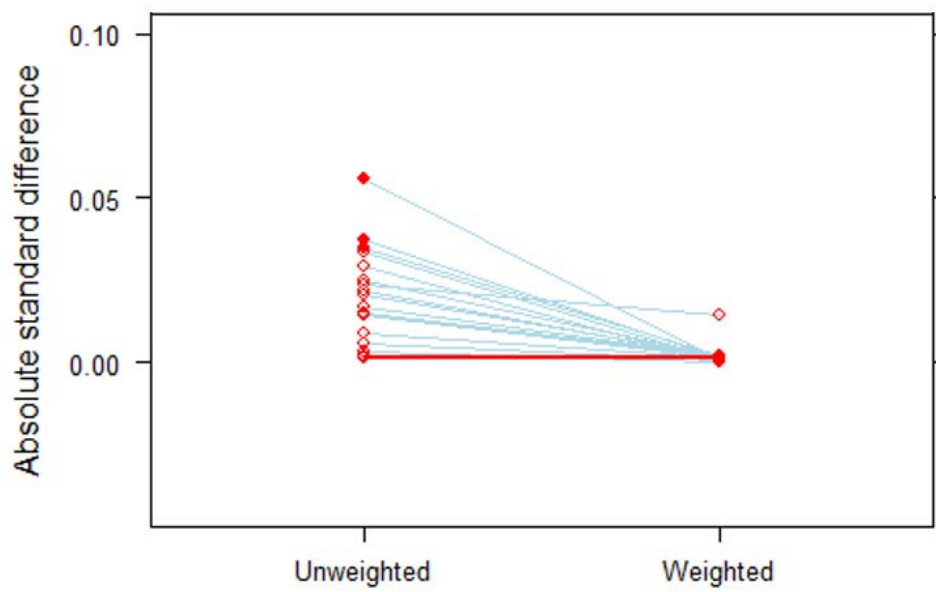


Figure C.5 Low vs. high hospital quality index. The standard difference is defined as the difference between the treatment group mean and the control group mean divided by the pooled sample standard deviation. Filled circles indicate that difference is statistically significant. All differences are not significant after propensity score weighting.

Table C.6

Sub-specialty training

Covariates	Unweighted			Propensity Score Weighting		
	Without sub-specialty training	With sub-specialty training	p-value	Without sub-specialty training	With sub-specialty training	p-value
International medical graduates	0.349 (0.477)	0.272 (0.445)	0.000	0.338 (0.473)	0.337 (0.473)	0.948
Attended medical school in NYS	0.320 (0.467)	0.342 (0.475)	0.048	0.324 (0.468)	0.319 (0.466)	0.649
Osteopath	0.100 (0.300)	0.135 (0.342)	0.000	0.105 (0.307)	0.107 (0.309)	0.837
Female	0.449 (0.497)	0.493 (0.500)	0.000	0.455 (0.498)	0.460 (0.499)	0.714
Family and General Practice	0.064 (0.245)	0.073 (0.260)	0.144	0.065 (0.246)	0.069 (0.253)	0.499
Internists	0.327 (0.469)	0.256 (0.437)	0.000	0.316 (0.465)	0.309 (0.462)	0.544
OB/GYN	0.063 (0.242)	0.088 (0.283)	0.000	0.067 (0.250)	0.067 (0.249)	0.950
Pediatricians	0.057 (0.232)	0.079 (0.270)	0.000	0.061 (0.238)	0.063 (0.244)	0.590
Psychiatrists	0.058 (0.234)	0.114 (0.318)	0.000	0.066 (0.249)	0.068 (0.251)	0.788
Surgeons	0.090 (0.286)	0.054 (0.225)	0.000	0.084 (0.278)	0.083 (0.276)	0.896
All other specialties	0.259 (0.438)	0.252 (0.435)	0.550	0.258 (0.437)	0.260 (0.439)	0.837
age	33.435 (4.519)	33.951 (4.648)	0.000	33.507 (4.537)	33.486 (4.464)	0.855
Missing age	0.050 (0.219)	0.054 (0.225)	0.547	0.051 (0.221)	0.050 (0.218)	0.804
Age<=25	0.000 (0.021)	0.000 (0.022)	0.958	0.000 (0.021)	0.000 (0.022)	0.938
Age>=50	0.008 (0.091)	0.011 (0.105)	0.227	0.009 (0.093)	0.008 (0.090)	0.831
Solo	0.020 (0.136)	0.033 (0.176)	0.001	0.022 (0.143)	0.022 (0.143)	0.964
Missing solo	0.050 (0.218)	0.063 (0.244)	0.016	0.051 (0.221)	0.047 (0.212)	0.408
Number of physicians	11,370	2,143		11,370	2,143	

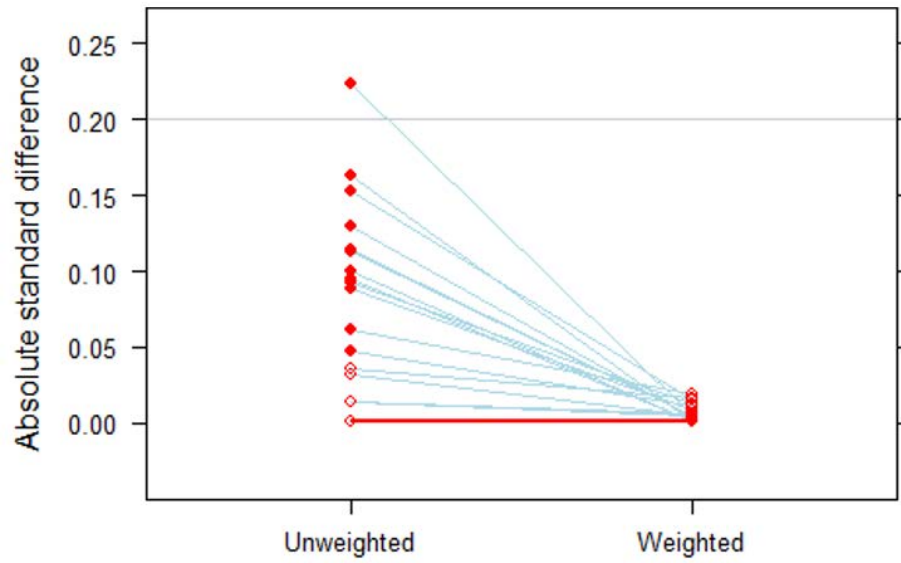


Figure C.6 With vs. without subspecialty training. The standard difference is defined as the difference between the treatment group mean and the control group mean divided by the pooled sample standard deviation. Filled circles indicate that difference is statistically significant. All differences are not significant after propensity score weighting.