Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents

Melanie Wasserman*

forthcoming, *The Review of Economic Studies*

Abstract

Do the long work hours required by many high-paying professions inhibit the entry of women? I investigate this question by studying a 2003 policy that capped the average workweek for medical residents at 80 hours. Using data on the universe of U.S. medical school graduates, I find that when a specialty reduces its weekly hours, more women enter the specialty, whereas there is little change in men's entry. I provide evidence that the increase in women is due to changes in labor supply, rather than labor demand. At the residency program level, I document that baseline female representation predicts female entry after the reform. A back-of-the-envelope calculation suggests that the reallocation of women among medical specialties due to the hours reduction can close the physician gender wage gap by 11 percent.

Keywords: occupational choice, long hours, gender
JEL codes: J16, J24, J44

*UCLA Anderson School of Management. Address: 110 Westwood Plaza, Entrepreneurs Hall C521, Los Angeles, CA 90095. E-mail: melanie.wasserman@anderson.ucla.edu. I thank David Autor, Esther Duflo and Heidi Williams for their extensive guidance and support. I have benefited from discussions with Nikhil Agarwal, Josh Angrist, Martha Bailey, Alex Bartik, Manasi Deshpande, Joseph Doyle, Paola Giuliano, Claudia Goldin, Jon Gruber, Sally Hudson, Erin Johnson, Danielle Li, Brendan Price, Adam Sacarny, Ashish Shenoy, Nico Voigtlaender, Romain Wacziarg, Nils Wernerfelt, and helpful comments from three anonymous referees and the editor as well as many seminar participants. Funding is gratefully acknowledged from the National Science Foundation Graduate Research Fellowship, the National Institute on Aging Grant Number T32-AG000186, NICHD Grant Number HD007339-30, and the MIT Shultz Fund.

1 Introduction

Over the last four decades, there has been a dramatic shift in the occupational choices of women in the U.S., with the female share of graduates in law, medical, and business schools rising by a factor of five (Blau et al., 2013). Despite the current near-equal representation of women and men entering these professional occupations, there remain persistent earnings disparities between male and female professionals. For example, recent statistics show that highly educated, full-time employed women earn 16 to 28 percent less than comparable men. Furthermore, the largest component of this gap now accrues to gender differences *within*, rather than across, broad occupational categories (Goldin, 2014; Blau and Kahn, 2017). This development has prompted researchers to examine the way that jobs within occupations are structured and compensated. One hypothesis, put forth by Claudia Goldin in her 2014 AEA Presidential Address, posits that convex returns to working long, continuous, and particular hours in certain occupations are the main driver of the remaining gender wage gap. Since women tend to work fewer hours than men and sort into positions with greater time flexibility, they may be less likely to reap the returns associated with rigid time requirements (Gicheva, 2013; Goldin, 2014; Cha and Weeden, 2014; Cortés and Pan, 2016). As of yet, there is little evidence on whether time requirements differentially affect women's propensity to enter a job and whether reducing time requirements would indeed narrow the gender wage gap.

This paper investigates whether a job's time requirements—particularly during the early years of individuals' careers—serve as barrier to entry for women. The economics literature has widely theorized that there are gender differences in preferences for occupational attributes, with women differentially valuing those that make working more compatible with actual or anticipated family formation (Polachek, 1981; Gronau, 1988; Adda et al., 2017). Empirical assessment of this hypothesis has presented researchers with a challenge, however. One typically observes equilibrium sorting behavior, i.e., the occupational outcomes of individuals, which is jointly determined by individual preferences, employer preferences, and occupational attributes. Thus, the empirical fact that women are clustered in jobs with lower time requirements does not alone imply gender differences in preferences for time requirements. For example, employer preferences over worker characteristics could give rise to this pattern if women are less likely to be selected for time-intensive, highly compensated positions due to human capital differences between men and women or employer discrimination. Furthermore, even if one is able to abstract from employer preferences, it is not evident whether women select into positions based on time requirements or another unobserved job attribute correlated with time requirements, such as a competitive work environment.

This paper advances the literature by studying a setting in which there was a plausibly exogenous change in the early career time requirements of a large professional occupation: physicians. Patterns in the medical profession mirror the broader trends of male and female professionals. Similar to law and business, starting in the mid-1970s, an influx of women brought the fraction of U.S. medical school graduates who are female to nearly 50 percent. Women and men, however, sort into different career paths within medicine, the first stepping-stone of which is the choice of a medical specialty. A medical specialty represents not only an individual's future earnings potential and the content and style of professional practice, but also the more immediate time demands during the training period, including the length and time intensity of medical residency. Figure 1 provides a snapshot of the heterogeneity in male and female specialty outcomes for the 2002 cohort of U.S. medical school graduates. Panel A plots the share of a medical specialty that is female against the specialty's average hours worked per week during the second year of medical residency, while Panel B plots the female share against the specialty's hourly earnings during professional practice (post-residency).[1] Consistent with the Goldin (2014) hypothesis, both relationships are negative, indicating that women tend to be clustered in less time-intensive and less remunerative specialties.[2]

I formally assess whether a specialty's time demands differentially influence women's career choices by studying the introduction in 2003 of a new policy by the Accreditation Council for Graduate Medical Education (ACGME) that restricted the workweek of medical residents to 80 hours. The impetus for this reform was notably not related to notions of work-life balance or to promoting the participation of women in time-intensive specialties. Rather, its introduction was triggered by mounting concerns regarding the deleterious effects of medical resident fatigue on medical errors and patient safety (ACGME, 2002).[3] The motivation for and nature of this policy make it a particularly attractive setting in which to study the effect of an occupation's time requirements on individuals' propensity to enter.

This paper's empirical strategy exploits the timing of the ACGME reform and the fact that it was differentially binding for medical specialties due to pre-existing differences in specialties' weekly hours. Using detailed data on the universe of U.S. medical school graduates from 1993 through 2010, I find that women are more likely to enter a medical specialty after its residency hours are reduced, whereas there is little change in men's entry behavior. A reduction of four hours per week induces a 5 to 15 percent increase in the share of women in a medical specialty. In contrast, there is, if anything, a slight decrease in the propensity of men to select into time-intensive specialties due to the reduction in hours, which could be a direct consequence of the new entry of women displacing men. The results are robust to various parameterizations of the pre-policy time intensity of medical specialties, the inclusion of time-varying specialty controls, and alternative methods of statistical inference.

The effects of the reform on female specialty entry could arise from changes in labor supply (medical residents' shifting preferences for specialties) or changes in labor demand (medical residency programs shifting preferences or hiring practices for female applicants). By analyzing survey data on the stated preferences of U.S. medical school students, I find that female medical school matriculants shift their preferences for time-intensive specialties in response to the reform. The point estimates, while less precisely estimated, are slightly larger than those from the specialty entry analysis and similarly differentiated by gender. This evidence supports the interpretation that reform-induced changes in medical residents' preferences are the driving force behind the increased entry of women into time-intensive specialties.

This paper also contributes to our understanding of how interactions among non-pecuniary job attributes determine the representation of women in an occupation. Within medical specialties, there remains substantial variation in residency programs' female representation as well as the incidence of family-friendly policies. I document heterogeneity in the effects of the reform on female entry at the residency program level according to baseline residency program attributes. The results demonstrate that gains in female representation due to the reform were larger among programs with established female communities, as proxied by the baseline fraction of residents and full-time faculty members who were female. In contrast, programs with and without family-friendly policies—such as maternity leave and on-site childcare—experienced similar increases in

[1] There is a positive correlation (correlation coefficient of 0.81) between hours worked during residency and professional practice (Iserson, 2006). Fully trained physicians, however, have substantially more discretion over their hours worked through choice of practice setting and volume of patients. Within each specialty, fully trained female physicians work fewer hours, on average, than fully trained male physicians.

[2] I focus on post-residency compensation since it is a proxy for lifetime earnings. The resident salary distribution is highly compressed, with little variation across programs and specialties (Nicholson, 2002; Agarwal, 2015).

[3] It is possible there is a productive purpose to working long hours—such as gains from the continuity of work—and nonlinearities in pay arise from the implied imperfect substitutability of workers (Goldin, 2014). On the other hand, hours could be inefficiently high if used as a screening mechanism (Landers et al., 1996). While I do not take a stand on the economic efficiency of long work hours, evidence from the medical community suggests that the reform had little impact on the quality of physician training and patient health outcomes (Volpp et al., 2013, 2007a,b; Jena et al., 2014a,b).

female representation due to the reform. One interpretation of this result is that programs with established female representation could be more amenable to hiring additional women. An alternative interpretation is that female applicants could view same-gender role models and improvements in work hours as complements rather than substitutes.

As a final exercise, I return to the hypothesis in Goldin (2014) and assess the implications of the rearrangement of women across specialties due to reducing early career time requirements for the physician gender wage gap. A back-of-the-envelope calculation suggests that the entry of women into historically time-intensive and highly compensated specialties due to the reform could close the physician gender wage gap by 11 percent.

To my knowledge, this paper is the first to use a natural experiment to estimate the causal effect of early career hours requirements on the propensity of men and women to select into an occupation. While several papers document that men and women, on average, sort into positions with differing pecuniary and non-pecuniary attributes, we still know relatively little about the extent to which time requirements affect occupational segregation by gender. Recent research shows that highly educated mothers shift away from occupations that experience increases in long hours during 1970 to 2010 (Cortés and Pan, 2016). In addition, there is evidence that when women have children, they transition from occupations characterized by long hours to those with more time flexibility (Pertold-Gebicka et al., 2016). Distinct from the previous literature, the present paper uses a clear source of variation in a job's time requirements stemming from an unanticipated profession-wide policy change. This strategy limits concerns regarding the endogeneity of changes in an occupation's time demands as well as ameliorates threats regarding an unobserved correlate of time requirements confounding the estimated relationship.

This paper also adds to an emerging literature on the relationship between work hours and the gender wage gap. A few recent survey and field experiments investigate gender differences in the valuation of a job's time flexibility (Wiswall and Zafar, 2018; Mas and Pallais, 2017). These studies find that women have a higher willingness to pay for predictable work hours and the availability of part-time work, but there is no difference in men's and women's willingness to pay for the level of work hours. This literature, however, has yet to examine the extremely long hours that are characteristic of many professional occupations. The present paper fills this gap and provides results suggesting that reducing these long work hours spurs the reallocation of women among career paths and could have substantial implications for the gender pay gap.

The structure of the paper is as follows. Section 2 provides background information on the medical profession and the ACGME 2003 duty hour reform. Section 3 describes the data sources. Section 4 discusses the empirical framework for examining the effect of hours requirements on specialty choice, presents the main results on specialty choice and stated specialty preferences, and explores heterogeneity in the effect of the reform by residency program attributes. Section 5 characterizes the implications of the reform for specialty segregation by gender as well as the physician gender wage gap. Section 6 concludes.

2 Medical Profession and the Duty Hour Reform

2.1 Specialty Selection

The decision of which medical specialty to pursue represents the determination of a career path within medicine, one that entails anticipatory human capital investments, a lengthy on-the-job training period, and high switching costs. Acceptance into residency programs hinges on performance during medical school, including scores from the U.S. Medical Licensing Exam (USMLE), medical school grades, letters of recommendation, and evaluations from third and fourth year clinical rotations. Since medical school coursework and the first portion of the USMLE occur early in medical school, as depicted in the medical school timeline in Figure 2, students often plan years in advance in order to emerge a competitive applicant.

Students make their final decision regarding a medical specialty when they apply for residency programs during the beginning of the fourth year of medical school. Residency programs then select applicants to interview. After interviews are complete, programs submit ranked lists of applicants, and students submit ranked lists of programs to the National Residency Matching Program (NRMP).[4] The result of the NRMP is a binding contractual agreement between the resident and the residency program. Selection of a medical specialty is typically considered a precursor to residency program application, although around ten percent of U.S. medical school graduates submit rank lists with residency programs from multiple specialties, meaning residency program rankings can help determine students' specialty outcomes (NRMP, 2000).

In addition to a specialty's monetary payoff, factors that have been cited as influential in specialty choice include the practice setting (hospital, solo practice, group practice), extent of interaction with patients, intellectual content, and lifestyle considerations such as the number of hours and the extent to which the specialty imposes idiosyncratic demands on one's time through being "on call" (USDHHS, 2008; Nicholson, 2002; Newton and Grayson, 2003; Dorsey et al., 2003; Gagné and Léger, 2005). As discussed in the Introduction, patterns of specialty choice differ markedly by gender. Descriptive work by Sasser (2005) finds that women tend to enter specialties with reduced hours, lower monetary penalties for having children, and lower gender earnings gaps. Ku (2011) finds that, upon entry into medical school, men's and women's different preferences for medical specialties are partly explained by women's greater emphasis on the social aspects of medicine and men's greater emphasis on the scientific/technical aspects of medicine. Current research has not addressed whether specialties' non-monetary attributes—in particular, time demands during residency—exert a causal influence on specialty choice and whether these effects differ by gender.[5]

2.2 The Duty Hour Reform as a Natural Experiment

Since its inception in the early 1900s, medical residency has entailed long hours, frequent periods of being "on call," and little time off.[6] Within the U.S. medical community, it was first recognized in the 1960s that these long hours could lead to excessive fatigue. The issue of medical resident work hours rose to national attention after the unexpected death of 18-year old college student Libby Zion in 1984, who was under the care of an allegedly sleep-deprived first year medical resident (Ludmerer, 2015). In 2003, due to mounting concerns regarding medical resident fatigue and sleep deprivation, and the associated heightened risk of medical errors, the Accreditation Council for Graduate Medical Education (ACGME) adopted a set of rules to limit the work hours of medical residents. Characterized as "one of the most substantial redesigns of the country's resident training system in more than a century" and a "watershed event for the ACGME," the new standards represented a departure from the near complete discretion afforded to medical specialties and residency programs in determining the work schedules of their residents (Philibert et al., 2009; Yoon, 2007). While there had been previous attempts at the state and federal level to regulate resident work hours, either

[4]Ophthalmology, Urology, and a small fraction of residency programs conduct their own matching outside of the NRMP.

[5]Agarwal (2015) finds that Family Medicine residents are willing to pay for programs at larger hospitals, located in their home or medical school state, and with a greater a range of cases. It is possible these characteristics vary across specialties and help determine specialty choice. As long as these attributes are stable over time, the empirical strategy in this paper accounts for this variation.

[6]Pre-WWII medical residency has been described as follows: "Whatever the season, house officers worked very long hours. Typically, they were 'on call' (that is admitting new patients and handling unforeseen problems with patients already on the service) every other night. Once or twice a month, they had weekends off, which customarily started Saturdays at noon and continued through the following Monday at 8 a.m" (Ludmerer 2015, p. 104).

these efforts never came to fruition or the regulation was inadequately enforced.[7] The ACGME 2003 duty hour reform had four main provisions:

1. Capped number of hours per week at 80, averaged over a four week period
2. Mandated one day off per week, averaged over a four week period
3. Limited maximum shift length to 30 hours
4. Mandated a minimum 10 hours rest period in between shifts (ACGME, 2002).

Penalties for non-compliance with these provisions included residency program probation and potential loss of accreditation, with monitoring through program audits and periodic surveying of medical residents. In order to comply with the new policy, many residency programs decreased the frequency of being on call, introduced separate day and night shifts (deemed "night float"), and hired physician extenders or medical paraprofessionals to substitute for resident work hours (Philibert et al., 2009).

It is important to note that only residency hours were within the purview of the duty hour reform. After residency, there is no regulation of physician hours, implying that the reduction in hours is confined to the early career period. Although residency comprises a small minority of physicians' expected working years, it tends to coincide with physicians' late 20's and early 30's, which are prime childbearing years for women. Furthermore, during residency individuals have limited means to adjust their labor supply on the extensive and intensive margins to accommodate their family formation choices.[8] The rigid time demands of the residency training period combined with the timing of residency in the lifecycle position women to be more responsive to residency hours requirements when making their specialty choices. While there is a positive correlation between hours worked during and after residency, fully trained physicians have substantially more discretion over their hours worked through choice of practice setting (solo, group practice, hospital, academic) and volume of patients (Iserson, 2006). Consistent with the notion that fully trained physicians can adjust their hours, within every specialty, women work fewer hours per week than men. For example, among fully trained General Surgeons, men work 71.4 while women work 63.9 hours per week. Among Dermatologists, the work hours are 48.2 for men and 38.4 for women.[9]

2.3 Did the Duty Hour Reform Reduce Hours?

To investigate whether the reform was effective in reducing hours worked among medical residents, I require measures of resident work hours before and after its introduction. According to ACGME monitoring data, most residency programs are in compliance with the reform. But it is widely recognized that the monitoring mechanism (resident self-reports of hours) may yield underestimates of hours worked due to the desire to protect the residency program, pressure from residency program directors, or anchoring or recall bias (Landrigan et al., 2006; Szymczak et al., 2010). In line with this conjecture, independent surveys yield non-compliance rates that are substantially higher than those reported by the ACGME (Landrigan et al., 2006). To minimize the potential for misreporting, I examine the effect of the reform on resident work hours

[7] New York state legislated limits on resident duty hours in 1989, but most residency programs were found in violation of the rules in 1998. Bills were introduced in 2002 in Congress to regulate resident hours, and the Occupational Safety and Health Administration (OSHA) considered petitions along similar lines in 2001.

[8] To be eligible for board certification, medical specialty boards stipulate that a resident must not be absent for more than four to six weeks in a given year. Taking additional time off from residency requires special permission. On the intensive margin, at the time the reform was implemented, 10 percent of residency programs offered part-time positions.

[9] I document this using the Practice Patterns of Young Physicians 1991 survey. Results are available upon request.

using the Current Population Survey (CPS), a nationally representative labor force data set collected by the U.S. Census Bureau, and nationally representative surveys of medical residents collected pre-reform or by non-ACGME researchers (U.S. Census Bureau, 2015).

Figure 3 Panel A uses individual reports of hours worked in the previous week from the CPS monthly files to plot the average weekly hours of physicians from 1989 through 2014 for medical residents and non-resident physicians. As medical resident status is not observed in the CPS, I impute it based on an individual's age (<35), occupation (physician), and if the individual works in a hospital.[10] In the years preceding the introduction of the duty hour reform in 2003, medical residents worked, on average, 64 hours in the previous week, well above the average of 50 hours worked by non-resident physicians. While there has been a smooth secular decline in the hours of non-resident physicians, the hours for resident physicians do not mirror this pattern (Staiger et al., 2010). Prior to the introduction of the reform, the hours for medical residents exhibit no clear trend. Right after the introduction of the policy in 2003, there is a discrete drop of 4 hours per week, with the reduction sustained over the subsequent years. Since the reform restricted average hours per week to 80, we expect the upper end of the hours distribution to be primarily affected. Figure 3 Panel B plots the fraction of physicians who worked more than 80 hours in the previous week separately for resident and non-resident physicians. Consistent with the stipulations of the policy, right after its introduction in 2003, there is a precipitous fall of more than ten percentage points in the fraction of medical residents who worked more than 80 hours per week, whereas there is little change among non-resident physicians.

Figure 4 documents the decline in medical resident work hours, by gender. Before the policy was enacted, female residents worked, on average, five fewer hours per week than male residents, and were four percentage points less likely to work more than 80 hours per week, likely an artifact of gender differences in specialty choice (Panel A). Men experience a reduction of approximately six hours per week, while women experience a more modest reduction of two hours per week (Panel A). The fraction of women and men who work more than 80 hours per week declines to the same level after the reform is enacted (Panel B).

While the duty hour cap was common across all medical specialties, it should have had disproportionate impacts on the most time-intensive specialties, such as General Surgery and Urology, where the typical resident pre-reform worked far in excess of 80 hours per week (Philibert et al., 2009). To confirm whether this hypothesized pattern is substantiated by trends in hours worked by specialty, I use reports of hours worked from three surveys of medical residents that were conducted either pre-reform by the ACGME or by independent researchers. For a measure of pre-policy hours, I use data from a 1999 nationally representative survey of 2,000 second year medical residents conducted by Baldwin Jr et al. (2003). To measure the change in hours before and after the implementation of the reform, I use two nationally representative surveys conducted by Landrigan et al. (2006), which collected reports of hours worked from approximately 2,700 first year residents in 2002 and 1,300 first year residents in 2003. Figure 5 plots for seven specialties the change in hours immediately preceding and succeeding the introduction of the duty hour reform (2002/3 to 2003/4) against pre-policy hours levels in 1999, and confirms the negative relationship between historical hours worked and the change in hours pre/post reform. As expected, average hours declined across all specialties with pre-policy hours near or above 80, with the steepest reductions among the specialties with the highest pre-policy hours.

I formalize the graphical relationship between pre-policy hours and the change in specialty hours before and after the reform by estimating the following regression:

[10] According to the AMA Masterfile sample, 97 percent of physicians under 35 who work in a hospital are medical residents (Staiger et al., 2010). The patterns described in the text are robust to minor adjustments in the age range of medical residents.

$$\text{Hours}_{st} = \delta_0 + \delta_1(\text{Hours}_{s,1999} \times \text{Post}_t) + \alpha_s + \gamma_t + \epsilon_{st} \tag{1}$$

where Hours_{st} is the average hours per week worked in specialty s in year t, where $t = \{2002, 2003\}$, α_s are specialty fixed effects, γ_t are year fixed effects, $\text{Hours}_{s,1999}$ represents the measure of pre-policy hours, taken from the 1999 survey of medical residents, and Post_t is an indicator variable for years after the reform went into effect. From this specification, the estimate of the coefficient of interest, δ_1, is -0.17 (standard error of 0.04), meaning one additional pre-policy hour per week induces a 0.17 hour per week decline post-policy. It is this variation in the extent to which the new standards were binding across specialties, in conjunction with the timing of the reform, that forms the basis of the identification strategy outlined below.[11]

3 Data

3.1 Data Sources

In order to examine the effects of the ACGME duty hour reform on physician specialty choice, I use two data sources. I examine the ramifications of the reform for physicians' medium-term specialty choice using the American Medical Association (AMA) Physician Masterfile ("AMA Masterfile"), which covers the universe of physicians in the U.S. The AMA Masterfile assembles information from a variety of administrative and survey data sources, and includes demographic characteristics (gender, age, and birthplace), medical training history (medical school and year of graduation) and primary specialty, observed at least four years after medical school graduation.[12] An individual's inclusion in the AMA Masterfile is triggered by entry into a U.S. medical school or U.S. medical residency program. Since information on hours is reported for 20 specialty categories,[13] I crosswalk the more detailed specialty information in the AMA Masterfile to the coarser categories, using a classification scheme provided by the Dartmouth Atlas.[14] I assign pre-policy average weekly hours to the 20 broad specialties in the AMA Masterfile using the previously discussed Baldwin Jr et al. (2003) 1999 nationally representative survey of second year medical residents. Since a specialty's pre-policy weekly hours are positively correlated with the length of residency (Appendix Table A.1 columns 1 and 6), I also provide an alternative metric of pre-policy time intensity that represents the pre-policy *total* number of hours over the entire residency period.[15] As a proxy for an individual's medical school quality, I classify medical schools according to whether they were included in U.S. News and World Report's 2014 ranking of U.S. medical schools, which includes about half of U.S. medical schools in its ranking.

I examine the initial specialty choices of medical residents using the American Association of Medical Colleges' National Graduate Medical Education Census Track survey of residency program directors ("GME

[11]It is possible that the reform also affected the variance of weekly hours or the predictability of schedules, which might be job attributes differentially valued by women. Unfortunately, I do not have data on the variance of weekly hours. While the reform was not enforced beyond residency, the reform could affect post-residency hours by conditioning physicians to work fewer hours or through compositional changes to specialties. The smooth evolution of non-resident physician hours after the reform suggests no major change in post-residency hours (Figure 3).

[12]According to specialty-specific studies of attrition from residency, the vast majority of attrition occurs during the first three years of residency (Everett et al., 2007; Sullivan et al., 2013; Yeo et al., 2010; Hatton and Loewenstein, 2004). Most residents who drop out of a residency program choose to pursue another specialty rather than leave medicine altogether.

[13]The 20 specialties are: Anesthesiology, Dermatology, Emergency Medicine, Family Practice, Internal Medicine, Internal Medicine/Pediatrics, Neurological Surgery, Neurology, Obstetrics/Gynecology, Ophthalmology, Orthopedic Surgery, Otolaryngology, Pathology, Pediatrics, Physical Medicine/Rehabilitation, Psychiatry, Radiation Oncology, Radiology, General Surgery, and Urology. I also exclude Preventive Medicine from the analysis since the Baldwin Jr et al. (2003) survey sample size is fewer than five individuals and Preventive Medicine residency programs accept no more than 10 individuals each year.

[14]The Dartmouth Atlas is available at: http://www.dartmouthatlas.org/downloads/methods/research_methods.pdf.

[15]$\text{TotalHours}_{s,1999} = \text{Hours}_{s,1999} \times 52 \times \text{ResidencyLength}_s$.

Census Track"). This survey collects information on program-level attributes for every ACGME-accredited U.S. residency program 1996-2010, including the number of female/male residents in each year of the program. I use the specialty choices of first-year residents and classify the 21 broad specialties in the GME Census Track data based on their pre-policy time intensity using Baldwin Jr et al. (2003).[16] I also use the GME Census Track to explore the allocation of new entrants across residency programs according to baseline program attributes in the survey, including program-level provision of paid maternity leave, program-level availability of onsite childcare, and the gender composition of full-time faculty. While the AMA Masterfile permits analysis of physicians' eventual specialty outcome—taking into account attrition from specialties during residency—the GME Census Track permits analysis of physicians' initial specialty choice. There are other distinctions between the GME Census Track and AMA Masterfile data that I discuss below.

To test the robustness of the effects of the reform, I allow specialty choices to evolve over time according to various baseline specialty attributes. First, I control for the baseline availability of paid maternity leave and onsite childcare, computed using the GME Census Track. I also control for baseline female representation in each specialty, computed using either the AMA Masterfile or the GME Census Track, depending on the data set used in the analysis. Last, I control for the fraction of applicants to a specialty that do not obtain a position—a metric of competitiveness—sourced from a specialty choice guide, Freeman (2003).

To test whether underlying preferences for specialties change as a result of the duty hour reform, I use data from the AAMC Matriculating Student Questionnaire (MSQ), which is administered to all first year U.S. medical school students. In survey years 1998-2006 and 2009-2010, students are asked the specialty category they are considering upon enrollment in medical school. Twenty-six specialties are represented in the MSQ and I again crosswalk these specialties to the 20 specialty categories described above. Additional information on all data sources is available in Online Appendix B.

3.2 Sample Restrictions and Summary Statistics

For the analysis of medium-term specialty entry using the AMA Masterfile, I limit the sample to U.S. medical school graduates from 1993 to 2010, which permits a ten-year and eight-year window before and after the introduction of the reform, respectively. The sample ends in 2010 in order to not confound the effects of the 2003 reform with a subsequent reform implemented in 2011, which limited the maximum shift length of first year medical residents to 16 hours. The timing of a physician's residency training governs their exposure to the duty hour reform. In the AMA Masterfile, I do not observe residency start date, so I use medical school graduation date as a proxy, which is an excellent approximation for U.S. medical school graduates (USMG), more than 90 percent of whom proceed directly from medical school to residency training. Medical school graduation date is a poor proxy for residency start date for foreign medical school graduates, many of whom train initially in their home countries before training in the U.S. (Appendix Figure A.1). For this reason, I exclude foreign graduates from the main analysis. I also exclude individuals who graduated from osteopathic medical schools but participated in an M.D. residency program, as there is a high incidence of missing specialty information among this population, which increases throughout the 1993-2010 period. I additionally exclude the 1.5 percent of individuals who do not have valid information on a primary specialty

[16] The 21 specialties are the same as in the AMA Masterfile data, but include Transitional Year. Due to the focus on first-year residents, I include residents who are completing a Transitional Year, which is one to two years of general training required by certain residency programs/specialties. I classify the hours of Transitional Year residents using information in Baldwin Jr et al. (2003). The exclusion of Transitional Year residents from this analysis does not change the results.

or have a medical school graduation date/year of birth that would imply graduating from medical school at an unreasonable age (<16 or >60 years old). The final sample for the specialty entry analysis is 281,433 U.S. medical school graduates.[17]

For the analysis of initial specialty entry using the GME Census Track, I limit the sample to U.S. medical school graduates who are observed in the first year of a residency program, 1996 to 2010. The sample starts in 1996 due to a major change in the survey format between 1995 and 1996. I exclude survey year 2000 due to its low response rate. Since the GME Census Track surveys residency program directors, I observe the allocation of men and women across residency programs in each year, for each specialty. The final sample for the initial specialty entry analysis is 270,887 residents who graduated from U.S. medical schools and are in their first year of a residency program.[18]

Table 1 presents summary statistics for the AMA Masterfile (Panel I) and GME Census Track (Panel II) samples, with columns 1 to 3 reporting summary statistics for the full sample, which includes foreign medical school graduates and osteopaths, and columns 4 to 6 reporting summary statistics for the USMG sample, which is the main sample used for analysis. In each case, the sample is almost half female, with an average age at medical school graduation of about 28 years. The exclusion of foreign graduates substantially increases the fraction of the sample that is U.S. born and attended a ranked medical school, as expected. In the USMG sample, the vast majority were born in the U.S. and about half attended medical schools included in U.S. News and World Report's 2014 rankings.[19] Since foreign medical school and osteopathic graduates comprise 32 percent of the full sample, their exclusion is a nontrivial sample restriction. I therefore reproduce the specialty entry analysis with the inclusion of foreign and osteopathic medical school graduates. All results are robust to their inclusion. Appendix Table A.1 reports the pre-policy distribution of physicians across medical specialties, by gender. For both men and women, across initial and medium-term choices, the largest specialty is Internal Medicine. After that, we observe a divergence by gender in specialty entry patterns. For initial specialty choice (columns 4 and 5), the next largest specialties for men are General Surgery, Family Practice, and Emergency Medicine, while for women they are Family Practice, Pediatrics, and Obstetrics/Gynecology. Since General Surgery serves as preliminary training for other specialties, many initial entrants in General Surgery do not continue on in the specialty.

Summary statistics for the GME Census Track sample used in the residency program-level analysis and the computation of specialty baseline characteristics are presented in Appendix Table A.3. I aggregate the program-level data to compute baseline specialty characteristics: the presence of a paid maternity leave policy and onsite childcare. For the program-level analysis, I exclude the broad specialty Internal Medicine-Pediatrics due to inconsistent coding over survey years and residency programs that are small (1 resident) in any survey year. None of these sample restrictions is consequential for the results.

4 The Effect of the Duty Hour Reform on Specialty Choice

In order to identify the effect of a specialty's hours requirements on specialty choice, I rely on two sources of variation: (1) the extent to which the duty hour reform was binding for a particular specialty, and (2) the

[17] I validate the final USMG AMA Masterfile sample using official AAMC data on U.S. medical school graduates. My restrictions reduce the number of U.S. medical school graduates by at most three percent, with no trend 1993-2010 (Appendix Table A.2).

[18] Note that some specialties, such as Urology and Ophthalmology, require preliminary residency training prior to entry (known as "preliminary years"). This means that a physician's first year of a residency program in a given specialty may diverge from a physician's first year of residency training overall. Due to the shuffling of students among specialties during the first one to two years of residency training, it is of particular importance to examine both the short- and medium-term specialty choices of physicians.

[19] The results are robust to using U.S. News and World Report rankings from 2000.

extent to which an individual was exposed to the policy change through the timing of residency training. The first source of variation captures a specialty's potential exposure to the provisions of the duty hour reform, based on the specialty's pre-policy time intensity. As shown in Figure 5, average hours declined across all specialties with pre-policy hours near or above 80, with the steepest reductions among the specialties with the highest pre-policy hours. The second source of variation stems from the timing of an individual's residency training. Medical school graduates who started residency training before 2003 were not aware of the hours restrictions at the time they chose a specialty for medical residency. Medical school graduates who started residency from 2003 onward, however, had the capacity to select their specialty, taking into consideration the reduction in hours associated with the reform.

To transparently display the variation in specialty choice and hours, by gender, I compute the change in the natural log share of individuals in each specialty before the policy went into effect (1993-2002) and after the policy went into effect (2003-2010), separately for men and women, using AMA Masterfile data. I plot the change in the log share against the average hours of the specialty, as measured in the 1999 survey conducted by Baldwin Jr et al. (2003). Figure 6 Panel A shows a positive correlation between changes in the female share and the pre-policy hours of a specialty, whereas Panel B depicts a negative relationship for men. These patterns are consistent with women increasingly entering more time-intensive specialties after the reform went into effect, and the potential displacement of men. In order to incorporate the full variation in specialty entry over this time period and account for the polychotomous nature of specialty outcomes, I use a conditional logit model of specialty choice.

4.1 Conditional Logit Specification

I model the probability that individual i's specialty outcome C_i is specialty s using the following conditional logit formulation:

$$\Pr(C_i = s) = \frac{\exp(\beta \text{TimeIntensity}_s + \text{X}_i'\delta_s + \text{D}_s'\eta)}{\sum_{s' \in S} \exp(\beta \text{TimeIntensity}_{s'} + \text{X}_i'\delta_{s'} + \text{D}_{s'}'\eta)} \tag{2}$$

where TimeIntensity_s is the time intensity of specialty s during residency, X_i is a vector of observable individual characteristics, and D_s is a vector of other observable specialty attributes. An individual's choice set consists of the 20 specialties in the AMA Masterfile and 21 specialties in the GME Census Track data. The goal of the empirical analysis is to estimate β, and compute the marginal effect of a specialty's time intensity during the training period on the probability of entering a specialty. The main challenge in obtaining an unbiased estimate of β relates to omitted variable bias. Absent exogenous variation in specialty time intensity, it is unclear whether it is this specific attribute or an omitted correlate that is driving an individual's specialty outcome.[20] High hours specialties could, for example, possess other disamenities, including poor work conditions or a "macho" culture, or could be specialties for which individuals have a lower unobserved taste.

In order to isolate exogenous variation in the time intensity of medical specialties, I use the empirical strategy outlined above that relies on residency timing and specialty variation. It leads to the following specification:

[20] A similar concern arises in the industrial organization literature regarding potential biases in estimation of price elasticities of demand in the absence of fully accounting for product characteristics.

$$\Pr(C_{it}=s)=\frac{\exp(\lambda_1(\text{Hours}_{s,1999}\times\text{Transition}_t)+\lambda_2(\text{Hours}_{s,1999}\times\text{Post}_t)+\text{X}_i'\delta_s+\alpha_s)}{\sum_{s'\in S}\exp(\lambda_1(\text{Hours}_{s',1999}\times\text{Transition}_t)+\lambda_2(\text{Hours}_{s',1999}\times\text{Post}_t)+\text{X}_i'\delta_{s'}+\alpha_{s'})} \quad (3)$$

where C_{it} represents the specialty outcome of individual i who graduated from medical school/entered a residency program in year t, $\text{Hours}_{s,1999}$ represents a specialty's pre-policy hours, Transition_t is an indicator for medical school cohorts 2003-2005, and Post_t is an indicator for medical school cohorts 2006-2010. Following other studies, I allow the effect of the reform on specialty entry to evolve over time by splitting up the post-reform period into a transition period (2003-2005) and a post period (2006-2010) (Babu et al., 2014; Jena et al., 2014b). The coefficients λ_1 and λ_2 on the interaction terms capture the effect of a reduction in hours on the log odds of entering a specialty. A positive λ_1 or λ_2 indicates that individuals are more likely to choose high hours specialties relative to low hours specialties after the reform is introduced. The introduction of the time dimension permits the inclusion of specialty fixed effects, α_s, which capture time invariant characteristics of or tastes for a particular specialty. In the empirical analysis, I additionally include controls for the effect of other specialty attributes D_s—measured pre-policy—over time through the interaction of D_s with time fixed effects. Time fixed effects are not identified in this model, since they enter as additive shifters in both the numerator and denominator.[21] I estimate the model using maximum likelihood, separately for men and women.

The identification strategy depends on the assumption that absent the reform, the log odds of entering more and less time-intensive specialties would have followed similar paths over time. Although this assumption is not directly testable, I probe its validity in a few ways. First, I examine the sorting patterns in the years leading up to the reform through estimation of an event study version of equation 3, with medical school graduates/first-year residency entrants from 2002 serving as the reference group. Second, I include individual controls, including age at medical school graduation and rank of medical school attended, to test the sensitivity of the effects to potential changes in the composition of the USMG population over this time period.

Third, I test the robustness of the estimates to the inclusion of trends for certain specialties and specialty groups that exhibited strong pre-existing trends that are unrelated to the reform. Ob/Gyn is an outlier specialty in that it is highly time-intensive specialty, yet women have entered en masse over the last thirty years. There has also been an extensively documented declining interest in the Primary Care specialties: Family Practice, Internal Medicine, and Pediatrics. Next, I control for the interaction of year fixed effects with baseline specialty characteristics, including the presence of family-friendly policies, female representation, and the competitiveness of obtaining a position. Finally, I test the robustness of the results to the inclusion of extrapolated specialty-specific linear pre-trends, by estimating a linear trend for each specialty during the period prior to the reform and including an interaction of the estimated coefficients with a linear trend in equation 3.[22] Due to the sensitivity of the results to the inclusion of specialty trends, I also provide descriptive evidence of the extent to which there is overlap between the specialties women differentially entered pre- and post-reform.

[21] A time shifter affects all specialties in an individual's choice set uniformly and has no effect on the choice of a particular specialty.

[22] Other papers that use a similar approach include Bhuller et al. (2013).

4.2 Initial and Medium-Term Specialty Choice

I start by estimating the effect of the reform on physicians' medium-run specialty choice, as measured by the AMA Masterfile. Table 2 reports the average marginal effects associated with the coefficients λ_1 and λ_2 on the interaction terms ($\text{Hours}_{s,1999} \times \text{Transition}_t$) and ($\text{Hours}_{s,1999} \times \text{Post}_t$) from maximum likelihood estimation of the conditional logit model specified in equation 3, separately for men and women. To compute average marginal effects, I compute the marginal effects of ($\text{Hours}_{s,1999} \times \text{Transition}_t$) and ($\text{Hours}_{s,1999} \times \text{Post}_t$) for each specialty. Then I average over the specialty-specific marginal effects to arrive at an average marginal effect of the reform on specialty entry. Standard errors are heteroskedastic robust and computed using the delta method.[23] Coefficients are reported in Appendix Table A.4.

From Table 2 Panel I.A column 1, we observe that for women the effect of the reform during the Post period is positive and statistically significant, indicating that after the introduction of the reform, women were more likely to choose high hours specialties. An additional hour pre-policy increases the probability that women choose a specialty post-reform by 0.023 percentage points. I discuss the magnitude of these results below in Section 4.4. During the Transition period, the effect of the reform is positive but one-tenth as large as the Post period effect. For men, there is little change in their specialty entry due to the reform during both the Transition and Post periods. Based on these estimates, I report the predicted probability of choosing each specialty pre-reform (1993-2002) and post-reform (2006-2010) in Appendix Table A.5.

The reform could affect physicians' medium-run specialty choices by increasing entry into or decreasing attrition from more time-intensive specialties.[24] To examine whether medium-run specialty choice is driven by increased entry, I turn to the GME Census Track data on first-year residents. From Table 2 Panel II.B column 1, we observe that the effects of the reform increased women's initial entry into more time-intensive specialties. Interestingly, the initial entry effects are smaller than the medium-run effects, implying the reform could have differentially reduced women's attrition from more time-intensive specialties. There are also other explanations for the smaller effects. First, the sample period is slightly different for the AMA Masterfile (1993-2010) and GME Census Track (1996-2010) samples. The AMA results are invariant to limiting the start year to 1996, however. Second, the GME Census Track data captures first-year residency program entry, which may not coincide with longer term specialty intentions, if a specialty requires preliminary training. Since the hours of specialties commonly used for preliminary training (Transitional Year, General Surgery, Internal Medicine) tend to be higher than the eventual specialty chosen, this initial allocation would, if anything, produce upward bias in the initial specialty entry results. Given the limited role for these other explanations, it is likely that the reform's effect on women's medium-term specialty choice are due to a combination of increased entry and decreased attrition.

To test the validity of the identifying assumption—absent the introduction of the reform, the log odds of choosing a more time-intensive specialty relative to a less time-intensive specialty would have been stable over time—I estimate an event study version of equation 3, separately for men and women. Figures 7 and 8 plot the year-by-year average marginal effects associated with the coefficients on interactions of $\text{Hours}_{s,1999}$ with medical school cohort/residency start year fixed effects, with 2002 serving as the reference year. Starting with the estimates for medium-run specialty choice, in Figure 7 Panel A, in the years before the reform, there was no change in female entry into more time-intensive specialties. Almost immediately after the reform was introduced in 2003, the average marginal effects begin to increase, which indicates a shift toward more

[23] See Online Appendix C for more details on the computation of average marginal effects and standard errors.

[24] Time-intensive specialties such as General Surgery are known to have attrition rates of 25 percent during residency, which leaves scope for the attrition mechanism to be operative. Since most specialty switching occurs during the first three years of residency, the AMA Masterfile should capture most within residency specialty switching.

time-intensive specialties. For men, there is a negative pre-trend in specialty entry in the years preceding the reform, meaning men's entry into more time-intensive specialties was decreasing relative to their entry into less time-intensive specialties. In the years after the reform, there is no apparent visual change in men's sorting behavior. The patterns are strikingly similar when examining initial specialty choice, depicted in Figure 8.

Next I test the robustness of the initial and medium-run results to the inclusion of various controls, including individual and specialty attributes. Table 2 column 2 controls for the decline in choosing Primary Care specialties, as well as the outlier status of Ob/Gyn, and finds similar results. In column 3, I find the results are unaffected by controls for the age and quality of medical school students. Columns 4-6 control for baseline specialty characteristics—including the available of family-friendly policies, fraction female, and competitiveness of securing a residency position—interacted with cohort fixed effects. Across all specifications, there remains a positive and statistically significant effect of the reform on women's choice of more time-intensive specialties. Furthermore, all tests reject equality for the female and male Post effects.

To explore whether specialties' pre-existing trends account for the increased entry of women post-reform, I include controls for specialty-specific extrapolated pre-trends. Specifically, using data from the pre-reform period, I estimate a linear trend for each specialty and create a new variable with these coefficients. I control for the interaction of this new variable with a linear time trend in equation 3. The results are reported in Table 2 column 7, which shows that the coefficient on ($\text{Hours}_{s,1999} \times \text{Post}_t$) declines a bit relative to the baseline specification. Based on these estimates, a continuation of specialty-specific pre-trends can account for 39 percent of the increase in women's medium-run specialty choice and 27 percent of the increased initial specialty choice. Note that the inclusion of extrapolated pre-trends (column 7) is a weaker test than directly including all specialty-specific linear trends. The direct inclusion of these trends over the entire sample period leaves little residual variation and causes the standard errors to substantially increase, however. In light of the sensitivity of the results to specialty-specific trends, in Online Appendix D, I provide descriptive evidence that the specialties experiencing increased entry of women pre-reform do not overlap with the specialties with the largest growth post-reform.

For men, the inclusion of the additional controls yields effects that are overall negative, which could be evidence of the increased entry of women displacing men.[25] One exception to this pattern are the specifications that include specialty-specific extrapolated pre-trends, which correct for the substantial negative pre-trend among men (Panel B of Figure 7) and flip the sign of the male effect.

Both event study figures show that the effect of the reform on female specialty choice increases over time. There are three possible reasons for this upward trend. First, consistent with the discussion in Section 2.1 on human capital investments necessary for specialty selection, it would have been difficult for students who were in the final years of medical school when the reform was introduced to change their specialties. Second, in line with the discussion in Section 2.3, there may have been weaker effects of the reform immediately after the roll out due to lagged implementation or imperfect compliance from residency programs. While there was an immediate drop in the hours of medical residents in 2003, there were further declines for the subsequent two to three years (Figure 3). Survey and anecdotal evidence also reveal widespread noncompliance in the initial years post-reform (Landrigan et al., 2006). Third, it may have taken some time for information regarding the efficacy of the reform to disseminate among medical school students.

[25] Female entry does not have to be perfectly offset by male exit. There is the possibility for modest specialty growth from limited growth in residency slots over this time period and unfilled slots in time-intensive specialties pre-reform.

4.3 Robustness

OLS evidence: I test the robustness of the conditional logit evidence with the following OLS difference-in-differences specification:

$$\ln sh_{st} = \beta_0 + \beta_1(\text{Hours}_{s,1999} \times \text{Transition}_t) + \beta_2(\text{Hours}_{s,1999} \times \text{Post}_t) + \alpha_s + \gamma_t + \epsilon_{st} \tag{4}$$

where the dependent variable is the natural logarithm of the share of individuals from medical school cohort t in specialty s. I use the natural logarithm of the share rather than the linear share due to the right-skewed distribution of specialty shares (Appendix Figure A.2). The independent variables are specialty fixed effects α_s and medical school cohort/residency start year fixed effects γ_t. All other variables are as defined above in equation 3. The regression is estimated separately for men and women. Appendix Table A.6 demonstrates that the OLS specification yields results that are consistent with those of conditional logit. For the medium-run specialty outcomes, an additional hour pre-policy causes a 0.63 percent increase in the share of women that enter a specialty from 2006 to 2010 (Panel I.A column 1). The 0.63 percent increase is similar to the marginal effect of 0.023 percentage points estimated in the conditional logit specification, which represents a 0.46 percent increase over the average specialty share of 5 percent. Turning to Panel I.B, the coefficients reveal there is little change in the specialty choices of men due to the reform and—if anything—there is a negative response. The event study also shows qualitatively similar patterns (Appendix Figure A.3).

Parameterizations of hours: The results are robust to various parameterizations of pre-policy hours requirements during residency (Appendix Table A.7), including (1) the pre-policy fraction of residents in a specialty who report working more than 80 hours per week, also sourced from Baldwin Jr et al. (2003), (2) the pre-policy total time investment for each specialty by multiplying the average pre-policy hours per week by the number of years of training each specialty requires, and (3) a binary classification of specialties based on whether their pre-policy average weekly hours exceeded 80. The results for women remain positive and statistically significant in the Post period. In the conditional logit, the classification based on total residency hours and the binary measure yield positive estimates for both men and women, with the female estimates larger than the male estimates.

Alternative methods of statistical inference: Appendix Table A.8 provides alternative methods of statistical inference for the conditional logit average marginal effects, including clustering standard errors at the medical school cohort/residency start year level (for the 18 cohorts analyzed) and clustering standard errors at the medical school level (for the 128 schools in the sample). The conclusions remain unchanged.

Alternative sample restrictions: The results are robust to the inclusion foreign and osteopathic medical school graduates (Appendix Table A.9 and Appendix Figure A.4). Since individuals who were already in residency at the time the reform was enacted (cohorts 2001/2) also experienced a reduction in hours, in Appendix Table A.7 column 5, I omit these cohorts who are partially affected by the reform and find similar results.

Independence of irrelevant alternatives: The conditional logit specification assumes that the error term is independently distributed, which leads to well-known limitations on the substitution patterns between specialties. Specifically, the independence of irrelevant alternatives (IIA) assumption in this context imposes the restriction that the choice between two specialties is independent of other specialties in the choice set. The IIA assumption is violated, for example, if the reduction of hours in Neurological Surgery increases the probability of choosing Neurological Surgery overall, and differentially decreases the probabilities of choosing General Surgery and Internal Medicine. Intuitively, this could occur if individuals view Neurological Surgery and General Surgery as closer substitutes than Internal Medicine and Neurological Surgery. In order to address this concern, I estimate a nested logit specification, which relaxes the IIA assumption by permitting the substitution elasticities between specialties within a given nest to differ from the substitution elasticities in other nests. I test three nest structures, using the Berry logit version of conditional logit (Berry, 1994).[26] The intuition behind these nests draws on a sequential decision-making process, in which individuals first decide whether they want to work in the broad categories of, for example, surgical or non-surgical specialties. Individuals then decide which of the specialties within the broad category is most appealing. The first nest structure defines two nests based on whether a specialty is (1) surgical: Neurological Surgery, General Surgery, Orthopedic Surgery or Urology, or (2) non-surgical. The second nest structure defines three nests based on whether a specialty is (1) surgical (as previously defined), (2) primary care: Family Practice, Internal Medicine, or Pediatrics, or (3) neither surgical nor primary care. The third nest structure defines four nests based on whether a specialty is (1) surgical (as previously defined), (2) primary care (as previously defined), (3) E-ROAD (lifestyle specialties): Emergency Medicine, Radiology, Ophthalmology, Anesthesiology, or Dermatology or (4) in none of the previous categories. Appendix Table A.10 reports the results. For women, the coefficients in the post-period remain positive and statistically significant. The coefficients for men are negative, with the exception of the first nest. In all cases, the effect of the reform is larger for women and statistically distinguishable from the effect for men.

4.4 Magnitude of the Effects

How large is the effect of the duty hour reform on medium-run female specialty choice? To answer this question, I use the OLS estimates to compute the indirect least squares version of the instrumental variables estimator, which is the ratio of the reduced form and first stage relationships (Angrist and Pischke, 2008).[27] As a lower bound (in magnitude) on the first stage relationship, I take the estimate of equation 1 using the survey data on seven specialties immediately pre/post reform.[28] As reported in Section 2.3, for each additional weekly hour pre-reform, the policy causes a -0.17 decline in post-reform weekly hours. I construct an upper bound on the first stage by assuming perfect compliance with the reform and arrive at an estimate of -0.55. I compute the ratio of the effect of the reform on female specialty choice estimated in Appendix Table A.6 column 1 of 0.63 percent and the lower (upper) bound first stage relationship of -0.17 (-0.55). Then I scale the per-hour effect by four hours, which is the average reduction in weekly hours due to the reform across all specialties. Putting these components together implies that a reduction of four hours per week in a specialty's residency hours causes a 5 to 15 percent increase in the share of women who choose the

[26] I use the Berry logit version of conditional logit due to the fact that the maximum likelihood estimation of nested logit did not converge for various nest structures. The Berry logit regression is: $\ln sh_{st} - \ln sh_{0t} = \beta_0 + \beta_1[(\text{Hours}_{s,1999} - \text{Hours}_{0,1999}) \times \text{Transition}_t] + \beta_2[(\text{Hours}_{s,1999} - \text{Hours}_{0,1999}) \times \text{Post}_t] + \alpha_s + \gamma_t + \epsilon_{st}$, where all variables have been normalized by the outside option specialty (Pathology) in year t (e.g. $\ln sh_{0t}$ is the log share of Pathology in year t). The nested logit version includes the control $\ln nestsh_{st}$, which is the natural logarithm of the share of specialty s of its respective nest in year t.

[27] I use the OLS estimates to maintain linear relationships across the first stage and reduced form. I cannot estimate two stage least squares since I only have data on hours worked after the reform for seven specialties.

[28] This is a lower bound on the effect of the reform on specialty hours since hours were measured in the year the reform went into effect. According to the qualitative literature, a considerable fraction of residency programs required additional time to adjust their residents' schedules.

specialty.[29]

4.5 Labor Supply or Labor Demand?

The above analysis demonstrates the reform increased women's propensity to enter historically time intensive specialties. This result could arise due to changes in physicians' preferences for specialties (labor supply) or changes in residency programs' preferences for the types of physicians they hire (labor demand). The reform could improve the hiring prospects of female applicants, if work hours were previously considered an impediment to program success for women. On the other hand, if the willingness to opt into high hours requirements was used by residency programs as a proxy for applicant quality and women are perceived as having a lower willingness to work long hours (on average), then eliminating this screening mechanism could have detrimental effects on female applicants' prospects.[30] One might surmise that residency programs could expand in response to the reform and thus accommodate additional interest. Since residency slots are subsidized by the U.S. federal government and this funding did not expand after the reform, programs had limited financial resources to increase capacity. Using the GME Census Track residency program level data on program size 1996-2010, I find the reform does not spur differential expansion in the availability of residency slots in high versus low hours specialties (Appendix Table A.11).

These capacity constraints give rise to an additional reason the specialty choice analysis may not reflect underlying changes in physicians' preferences for specialties. When an individual decides on a medical specialty, she could weigh the competitiveness of her application relative to the expected applicant pool. If an individual anticipates having a low chance of being accepted into a residency program in a specialty, then she may decide to pursue a specialty other than her unconstrained utility-maximizing choice. If individuals would like to enter time-intensive specialties after the reform but cannot due to capacity constraints, the resulting specialty choice changes will be downward biased relative to the unconstrained case.[31]

To disentangle the labor supply and the labor demand channels, I assess whether how the reform affected stated preferences for specialties using data from the Association of American Medical Colleges (AAMC) Matriculating Student Questionnaire (MSQ) on the specialty students are considering upon enrollment in medical school, for students entering in years 1998-2006 and 2009-2010.[32] Radiation Oncology and Internal Medicine-Pediatrics excluded from the survey specialty options, so 18 of the 20 specialties used in the main analysis are represented. I use these data to estimate the OLS specification outlined in equation 4, where the dependent variable is the natural log share of women (men) in a medical school cohort who express interest in each specialty and the medical school cohort indicates the year entering rather than exiting medical school.[33]

The results in Table 3 broadly support the interpretation that changes in residents' preferences—rather than residency programs' preferences—are the primary driver of the effect of the reform on female specialty entry. There is a positive and statistically significant effect of the reform on women's stated preferences for more time-intensive specialties. Analogous to the specialty choice results, the female coefficients are consistently larger than the male coefficients, although due to large standard errors I cannot reject the null

[29] The calculations are: $\frac{0.63}{-0.17} \times -4 = 14.82$ and $\frac{0.63}{-0.55} \times -4 = 4.58$.

[30] This is similar to the unintended consequences of "ban the box" policies (Agan and Starr, 2018; Doleac and Hansen, 2020). Affirmative action for female applicants could imply that similar labor supply responses among men and women translate into greater female entry. In contrast, discriminatory attitudes toward female applicants would yield the opposite result.

[31] Nicholson (2002) shows that incorporating capacity constraints into estimation of medical students' elasticity of specialty choice with respect to specialty income yields estimates that are substantially smaller than the unconstrained scenario. For similar reasons, rationing of slots in the most time-intensive specialties may introduce downward bias in the effect of a reduction in hours on specialty choice.

[32] I was not able to obtain data on applications or NRMP rankings, by gender, to test these metrics of labor supply.

[33] I allow the reform's effects to evolve over time due to the possibility of lagged implementation or information dissemination.

hypothesis that the effects for women and men are the same in some specifications. Using the estimates from column 1 and scaling by the first stage lower and upper bounds, I find that a four hour per week reduction causes a 6 to 21 percent increase in interest for women and a 2 to 8 percent (insignificant) increase for men. The male and female coefficients in Table 3 are larger than the OLS estimates of the effect of the reform on specialty choice in Appendix Table A.6, which suggests that post-reform, the increased interest—for both women and men—could not be fully accommodated by available residency slots.[34]

4.6 Implications for Talent Allocation

A growing literature suggests that the underrepresentation of women in certain occupations could arise from occupational frictions that prevent women from pursuing their comparative advantage in human capital accumulation and occupational choice, and lead to the misallocation of talent (Hsieh et al., 2019). If this is the case, we would expect specialties with the lowest female representation to have the most talented women. Indeed, before the reform was introduced, there was a negative relationship between a specialty's female representation and the fraction of women in the specialty who attended a ranked medical school. In addition, women in time-intensive specialties were more likely to have attended a ranked medical schools than their male counterparts. The talent misallocation hypothesis also has implications for the marginal female entrant. If the hours cap relaxes an occupational friction, then the marginal woman induced to enter a time-intensive specialty due to the reform should be of lower quality than the average female entrant.[35]

To test this prediction regarding the quality of the marginal female entrant, I estimate the following specification using the AMA Masterfile, separately for men and women:

$$Ranked_{ist} = \beta_0 + \beta_1(\text{Hours}_{s,1999} \times \text{Transition}_t) + \beta_2(\text{Hours}_{s,1999} \times \text{Post}_t) + \alpha_s + \gamma_t + \epsilon_{ist} \tag{5}$$

where $Ranked_{ist}$ is an indicator for whether individual i in specialty s from medical school cohort t attended a ranked medical school. From the results in Appendix Table A.12, we observe that the marginal female entrants are less likely to come from a ranked medical school than the inframarginal women; a four hour reduction leads marginal female entrants to be 0.69 to 2.24 percentage points less likely to come from a ranked school. It is interesting to note that this compositional change is also present among male entrants, particularly in the period immediately after the reform's introduction, even absent an effect of the reform on men's specialty choice.[36]

4.7 Residency Program Drivers of Female Entry

Given the positive effect of the reform on women's specialty choice, we would expect that after the reform, female representation in residency programs in more time-intensive specialties would rise relative to programs in less time-intensive specialties. In this section, I document this changing gender composition at the residency program level and additionally investigate heterogeneity in the effects of the reform on female entry into programs according to baseline program attributes. Building on the literature on the effects of same-gender role models on career path choice, the first set of attributes is related to programs' baseline female

[34] Given the increase in the effects the reform over time, it is possible that the omission of years 2007 and 2008 from the MSQ data attenuates the estimates of the effect of the reform on specialty preference in the post-period.

[35] Another possibility is hours requirements serve as a screening mechanism for unobserved talent. In this scenario, we might expect an observable characteristic such as medical school ranking to hold more weight in residency placement after the reform's implementation, leading to the new female entrants being positively selected on this dimension.

[36] I also find some evidence that male and female marginal entrants into time-intensive specialties are older.

representation, including female representation among full-time physician faculty who supervise residents and among first-year residents. The second set of attributes captures whether programs have official policies to accommodate having children during residency, including whether the program has a paid maternity leave policy and provides onsite childcare. Since 15 to 30 percent of female physicians (depending on the specialty) have children during residency, it is possible that female residents take these policies into consideration when selecting residency programs (Finch, 2003; Hutchinson et al., 2011; Turner et al., 2012).

For this analysis, I employ residency-program level data from the GME Census Track survey of residency program directors for years 1996-2010. The sample consists of 2,781 programs and 31,969 program-year observations, spanning 19 of the 20 broad specialties from the above analysis (Appendix Table A.3).[37] Within medical specialties, there is substantial variation in residency programs' female representation. Since each program sets its own parental leave policies, there is within-specialty variation in the provision of paid maternity leave. While onsite childcare is provided by the hospital, hospitals may not have residency programs in every specialty, giving rise to within-specialty variation in onsite childcare as well. Before the reform, there is no relationship between the female representation in a program and the program's provision of maternity leave and onsite childcare.[38] Within specialty, there is a positive correlation between the fraction of the faculty and the fraction of the residents in a program who are female.

I use the following specification to estimate the effect of the reform on a program's gender composition:

$$FractionFemale_{pst} = \beta_0 + \beta_1(\text{Hours}_{s,1999} \times \text{Transition}_t) + \beta_2(\text{Hours}_{s,1999} \times \text{Post}_t) + \alpha_p + \gamma_t + \epsilon_{pst} \quad (6)$$

where $FractionFemale_{pst}$ is the fraction of first year residents in program p in specialty s in year t who are female and α_p are program fixed effects. The reform had a positive and statistically significant effect on the fraction of residents in a program who are female; an additional hour pre-policy causes a 0.078 percentage point increase in the Post period (Table 4 Panel A).

Next I examine whether the effects of the reform vary with baseline program attributes. I estimate equation 6 separately for programs above/below the specialty-specific average for female representation and for programs with/without family-friendly policies, where each of these classifications stems from the program's 1996 survey response. In Table 4 Panel B, we observe that the effect of the reform on the fraction of residents who are female is larger among programs with higher baseline female representation among their full-time faculty and residents. This result indicates that new female entrants into specialties select programs with established female communities or, alternatively, programs with established female communities are more amenable to hiring women after the reform.[39] While this heterogeneous response is not causal, it is consistent with the literature on the positive effects of female role models on women's

[37] Similar to the specialty entry analysis using the GME Census Track, I exclude fellowship programs in sub-specialties (e.g. Cardiology, Vascular Surgery) that require completion of another specialty's residency program before entry. I also exclude the broad specialty Internal Medicine-Pediatrics, since its programs are inconsistently defined across years. I also exclude Transitional Year. I include residency programs that require preliminary years of training in other specialties (e.g. Urology, Diagnostic Radiology). Residency programs with one resident (in any year) are excluded. An unbalanced panel of programs is used in the main specifications, with 94 percent of programs that meet the other sample restrictions observed at least 14 times throughout the 1996-2010 time period. Dropping the 6 percent of programs with fewer than 14 observations yields similar results.

[38] The provision of onsite childcare is stable over the 1996-2010 period and the reform does not have an effect on its provision. The fraction female among faculty rose over this period, but not differentially across more and less time-intensive programs. Due to changes in the definition of maternity leave policies between survey rounds, I cannot chart its pattern over time.

[39] Another possibility is that programs with initially higher female representation experienced larger reductions in hours and therefore attracted more women. Conditional on specialty, there is no relationship between pre-reform female faculty/resident representation and hours. To assess whether compliance varies with female representation, I estimate the heterogeneity specifications using a program-specific measure of hours, rather than the specialty-specific measure, and find similar patterns. Note that the program-specific hours are reported by program directors and could be biased downward post-reform.

propensity to enter STEM careers and points to a potential externality of increased female representation (Kahn and Ginther, 2018; Carrell et al., 2010; Canes and Rosen, 1995). This result also suggests that female role models could be complements to—rather than substitutes for—work hour requirements. Turning to the family-friendly policies, the reform has a slightly larger effect on female representation in programs with a maternity leave policy, but there is little difference in the reform's effects based on the presence of onsite childcare (Table 4 Panel C). Overall, there is less heterogeneity in the effects of the reform on female representation in residency programs based on official policies that make residency more compatible with having children. This pattern could arise because women do not value or anticipate using family-friendly policies. Alternatively, the existence of a policy may not capture the generosity of the policy or norms surrounding policy take-up.[40]

5 Implications for Occupational Segregation and Pay

5.1 Specialty Segregation by Gender

Did the reform contribute to the convergence of male and female specialty choices? To provide an aggregate calculation of specialty segregation by gender, I compute the Duncan Segregation Index, defined as:

$$D_t = 0.5 \times \sum_s \left| \frac{f_{st}}{F_t} - \frac{m_{st}}{M_t} \right|$$

where f_{st} is the number of women in specialty s from medical school cohort t and F_t is the total number of women from cohort t. The variables are analogous for men. This is a metric of specialty segregation by gender, whose range is between 0 and 1. The metric represents the fraction of women (or men) who would have to change specialties in order for the specialty distribution of men and women to be identical. Figure 9 plots the Duncan Index from 1993 to 2010 using the AMA Masterfile. We observe that prior to the reform, men's and women's specialty distributions were growing increasingly dissimilar, with peak dissimilarity in 2002, at which point nearly 30 percent of female physicians would have needed to change their specialty in order to reach gender parity in specialty choice. After the reform the segregation index stabilizes and decreases a bit, suggesting that the reform halted the growing segregation of men's and women's specialty choices.

5.2 Physician Gender Wage Gap

As discussed in the Introduction, Goldin (2014) posits that a key contributor to the remaining gender wage gap is the presence of convex returns to working long hours, and women's lower likelihood of reaping these returns, either through their choice of job or choice of hours within a job. Using the estimates of the effect of the reform on specialty choice and the average wage associated with each specialty, I assess the implications of the entry of women into high hours, high compensation specialties due to the reform for physician gender wage gap. For this exercise, I use specialty-specific average hourly earnings from the Community Tracking Study, as reported in Leigh et al. (2010). One limitation of these data is the specialties covered are those that require direct patient care and therefore exclude Anesthesiology, Pathology, and Radiology. I simulate the gender pay gap before the introduction of the policy by computing the weighted average of specialty-specific

[40] From a survey of female surgeons who completed residency in 2007 or later, 39 percent of those who had children during residency reported having seriously considered dropping out of their program, with the 72 percent citing inadequate maternity leave (of no more than 6 weeks) and lack of childcare support as contributing factors (Rangel et al., 2018).

pay, w_s, where the weights are the shares of men and women in each specialty pre-reform, $sh_{sg} = \frac{n_{sg}}{N_g}$: $\bar{w}_g = \sum_s w_s sh_{sg}$. According to this calculation, the pre-reform gender hourly earnings gap is $4.80 ($76.60 for men and $71.80 for women, all in 2004 dollars) or 6.3 percent. This is smaller than the gap in average salaries for male and female physicians reported elsewhere, likely due to the other estimates incorporating gender differences in choice of practice setting and hours worked (Lo Sasso et al., 2011; Sasser, 2005). This measure, however, isolates the portion of physician gender pay gap that is due to hourly pay differences across specialties alone.[41]

Quantifying the contribution of the reform to average female (male) hourly earnings entails a sum of specialty-specific female (male) share changes due to the reform, weighted by specialty earnings: $\Delta \bar{w}_g = \sum_s w_s \Delta sh_{sg}$.[42] The back-of-the-envelope calculation suggests that, through specialty selection, the reform could increase women's average hourly earnings by $0.53. The change in male average hourly earnings implied by this reform is zero, since men are unresponsive in their specialty choices. Thus, the rearrangement of women among medical specialties due to the reform could close the physician gender wage gap by 11 percent.[43]

While this contribution to the physician pay gap may seem sizable, other estimates in the literature indicate that women have a substantial willingness to pay for jobs with the availability of part-time work or flexible scheduling. For example, Mas and Pallais (2017) find that female applicants are willing to pay about two dollars more per hour than their male counterparts for a more flexible job, a difference amounting to more than 20 percent of the offered wage. Cortés and Pan (2019) calculate that a one standard deviation decrease in the gender gap in working long hours decreases the gender earnings gap by 30 percent. Wiswall and Zafar (2018) estimate that, even after controlling for college major, the gender earnings gap would be reduced by at least 25 percent if men and women had the same preferences over job attributes. Adding to this body of evidence, my estimates suggest that modest reductions in early career time requirements could be instrumental in narrowing the gender pay gap.

6 Conclusion

Recent public debate on the gender wage gap has focused on two explanations. The first contends that earnings differentials between men and women are primarily driven by women's behavior or decision-making in the workplace, such as their level of confidence and propensity to negotiate salaries or apply for promotions (Sandberg, 2013). The second explanation cites institutional or organizational factors, such as inflexible job characteristics, the absence of low-cost childcare, and lack of paid parental leave, which may disproportionately impede women's entry into and upward mobility within occupations (Slaughter, 2015). This paper informs the debate by empirically examining whether one non-monetary attribute of jobs in high-paying professions—long, inflexible time requirements during early career years—differentially deters women from entering. Using plausibly exogenous variation in weekly hours worked during medical residency stemming from the introduction of the 2003 ACGME duty hour reform, I find that reducing a medical specialty's hours induces more women to enter, but has little effect on men's entry. Furthermore, the entry of women appears

[41] Note that this computation does not take into consideration how hourly wages, conditional on specialty, vary by gender. Sasser (2005) finds that there is only a small gender gap in hourly earnings among fully trained physicians, conditional on specialty. Furthermore, across high and low hour specialties, female physicians are able to reduce their weekly work hours without a substantial penalty to hourly earnings.

[42] I compute the difference between the predicted specialty shares before and after the reform. I assume that specialty-specific pay is time invariant.

[43] This calculation assumes that new female entrants would have earnings similar to prior cohorts of women working in these specialties. The direction of the bias from compositional changes is theoretically ambiguous.

to be due to changes in physician preferences for specialties, rather than a shift in residency program preferences for hiring women. I estimate that the entry of women into historically time-intensive and high-paying specialties due to a four hour per week reduction could close the physician gender pay gap by 11 percent.

Excessively long hours are not unique to the medical profession; early career investment bankers work, on average, 74 hours per week and 18 percent of tenure-track assistant professors at research universities work 70 or more hours per week (Bertrand et al., 2010; NCES, 1999). These long work hours may serve a productive purpose, such as fostering skill acquisition, producing gains from the continuity of work, or addressing the imperfect substitutability of workers. This paper demonstrates, however, that the hours requirements ubiquitous among professional occupations have important implications for occupational segregation by gender. From an equity perspective, reducing hours requirements could be considered an effective tool alongside the many gender diversity initiatives enacted by employers. From an efficiency perspective, as discussed by Hsieh et al. (2019), it is possible that long hours are an occupational friction restricting the optimal allocation of talent in the labor market.

Data Availability Statement

The data underlying this article were (1) derived from sources in the public domain, (2) were provided by the American Medical Association (AMA) directly, or (3) were provided by the AMA through its licensing with Medical Marketing Service (MMS), Inc. Data provided by the AMA are available at a cost to other researchers. The code and publicly available data underlying this research are available on Zenodo at https://doi.org/10.5281/zenodo.6578263.

References

ACGME, "Report of the ACGME Work Group on Resident Duty Hours, June 11, 2002," Technical Report 2002.

Adda, J, Christian Dustmann, and K Stevens, "The Career Costs of Children," *Journal of Political Economy*, 2017, *125* (2), 293–337.

Agan, Amanda and Sonja Starr, "Ban the box, criminal records, and racial discrimination: A field experiment," *Quarterly Journal of Economics*, 2018, *133* (1), 191–235.

Agarwal, Nikhil, "An Empirical Model of the Medical Match," *American Economic Review*, 2015, *105* (7), 1939–1978.

Angrist, Joshua D. and Jörn-Steffen Pischke, *Mostly Harmless Econometrics*, Princeton, NJ: Princeton University Press, 2008.

Babu, Ranjith, Steven Thomas, Matthew A Hazzard, Yuliya V Lokhnygina, Allan H Friedman, Oren N Gottfried, Robert E Isaacs, Maxwell Boakye, Chirag G Patil, Carlos A Bagley, Michael M Haglund, and Shivanand P Lad, "Morbidity, Mortality, and Health Care Costs for Patients undergoing Spine Surgery following the ACGME Resident Duty-Hour Reform," *Journal of Neurosurgery. Spine*, 2014, *21* (4), 502–15.

Baldwin Jr, Dewitt C, Steven R Daugherty, Ray Tsai, and Michael J Scotti Jr, "A National Survey of Residents Self-Reported Work Hours: Thinking Beyond Specialty," *Academic Medicine*, 2003, *78* (11), 1154–1163.

Berry, Steven T., "Estimating Discrete Choice Models of Product Differentiation," *The RAND Journal of Economics*, 1994, *25* (2), 242–262.

Bertrand, Marianne, Claudia Goldin, and Lawrence F Katz, "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors," *American Economic Journal: Applied Economics*, 2010, *2* (July), 228–255.

Bhuller, Manudeep, Tarjei Havnes, Edwin Leuven, and Magne Mogstad, "Broadband internet: An information superhighway to sex crime?," *Review of Economic Studies*, 2013, *80* (4), 1237–1266.

Blau, Francine D. and Lawrence M. Kahn, "The Gender Wage Gap: Extent, Trends, and Explanations," *Journal of Economic Literature*, 2017, *55* (3), 789–865.

_ , Peter Brummund, and Albert Yung Hsu Liu, "Trends in Occupational Segregation by Gender 1970-2009: Adjusting for the Impact of Changes in the Occupational Coding System," *Demography*, 2013, *50* (2), 471–492.

Canes, Brandice J. and Harvey S. Rosen, "Following in Her Footsteps? Faculty Gender Composition and Women's Choices of College Majors," *Industrial and Labor Relations Review*, 1995, *48* (3), 486–504.

Carrell, Scott E, Marianne E Page, and James E West, "Sex and Science: How professor Gender Perpetuates the Gender Gap," *The Quarterly Journal of Economics*, 2010, *125* (3), 1101–1144.

Cha, Y. and K. A. Weeden, “Overwork and the Slow Convergence in the Gender Gap in Wages,” *American Sociological Review*, 2014, *79* (3), 457–484.

Cortés, Patricia and Jessica Pan, “Prevalence of Long Hours and Women’s Job Choices: Evidence across Countries and within the U. S.,” *IZA Discussion Paper No. 10225*, 2016, (April), 1–42.

_ **and** _ , “When Time Binds: Returns to Working Long Hours and the Gender Wage Gap among the Highly Skilled,” *Journal of Labor Economics*, 2019, *37* (2).

Doleac, Jennifer L and Benjamin Hansen, “The unintended consequences of "ban the box": Statistical discrimination and employment outcomes when criminal histories are hidden,” *Journal of Labor Economics*, 2020, *38* (2), 321–374.

Dorsey, E. Ray, David Jarjoura, and Gregory W. Rutecki, “Influence of Controllable Lifestyle on Recent Trends in Specialty Choice by US Medical Students,” *JAMA*, 2003, *290* (9), 1173–1178.

Everett, Christopher B., Stephen D. Helmer, Jacqueline S. Osland, and R. Stephen Smith, “General surgery resident attrition and the 80-hour workweek,” *American Journal of Surgery*, 2007, *194*, 751–757.

Finch, Susan J, “Pregnancy during residency: a literature review.,” *Academic Medicine*, 2003, *78* (4), 418–428.

Freeman, Brian, *The Ultimate Guide to Choosing a Medical Specialty*, 2 ed., McGraw Hill, 2003.

Gagné, Robert and Pierre Thomas Léger, “Determinants of Physicians’ Decisions to Specialize,” *Health Economics*, 2005, *14* (7), 721–35.

Gicheva, Dora, “Working Long Hours and Early Career Outcomes in the High-End Labor Market,” *Journal of Labor Economics*, 2013, *31* (4), 785–824.

Goldin, Claudia, “A Grand Gender Convergence: Its Last Chapter,” *American Economic Review*, 2014, *104* (4), 1091–1119.

Gronau, Reuben, “Sex-Related Wage Differentials and Women’s Interrupted Labor Careers-the Chicken or the Egg,” *Journal of Labor Economics*, 1988, *6* (3), 277–301.

Hatton, Mark P. and John Loewenstein, “Attrition from ophthalmology residency programs,” *American Journal of Ophthalmology*, 2004, *138* (Table 1), 863–864.

Hsieh, Chang-Tai, Erik Hurst, Charles Jones, and Peter Klenow, “The Allocation of Talent and U.S. Economic Growth,” *Econometrica*, 2019, *87* (5), 1439–1474.

Hutchinson, Anne M., Norman S. Anderson, Gregory L. Gochnour, and Connie Stewart, “Pregnancy and childbirth during family medicine residency training,” *Family Medicine*, 2011, *43* (3), 160–165.

Iserson, Kenneth V., *Iserson’s Getting into a Residency: A Guide for Medical Students*, Tucson, Ariz: Galen Press, 2006.

Jena, Anupam B, Lena Schoemaker, and Jay Bhattacharya, "Exposing Physicians to Reduced Residency Work Hours Did Not Adversely Affect Patient Outcomes After Residency," *Health Affairs*, 2014, *33* (10), 1832–40.

_ , Vinay Prasad, and John A Romley, "Long-term Effects of the 2003 ACGME Resident Duty Hour Reform on Hospital Mortality," *Mayo Clinic Proceedings*, 2014, *89* (7), 1023–5.

Kahn, Shulamit and Donna Ginther, "Women and STEM," in Susan L. Averett, Laura M. Argys, and Saul D. Hoffman, eds., *Oxford Handbook on the Economics of Women*, Oxford University Press, 2018.

Ku, M. C., "When Does Gender Matter?: Gender Differences in Specialty Choice Among Physicians," *Work and Occupations*, 2011, *38*, 221–262.

Landers, By Renee M, James B Rebitzer, and Lowell J Taylor, "Rat Race Redux : Adverse Selection in the Determination of Work Hours in Law Firms," *The American Economic Review*, 1996, *86* (3), 329–348.

Landrigan, Christopher P, Laura K Barger, Brian E Cade, Najib T Ayas, and Charles A Czeisler, "Interns Compliance with Accreditation Council For Graduate Medical Education Work-Hour Limits," *JAMA*, 2006.

Leigh, J. Paul, Daniel Tancredi, Anthony Jerant, and Richard L. Kravitz, "Physician Wages Across Specialties," *Arch Intern Med*, 2010, *170* (19), 1728–1734.

Lo Sasso, Anthony T., Michael R. Richards, Chiu Fang Chou, and Susan E. Gerber, "The $16,819 Pay Gap for Newly Trained Physicians: The Unexplained Trend of Men Earning more than Women," *Health Affairs*, 2011, *30* (2), 193–201.

Ludmerer, Kenneth M., *Let Me Heal: The Opportunity to Preserve Excellence in American Medicine*, Oxford University Press, 2015.

Mas, Alexandre and Amanda Pallais, "Valuing Alternative Work Arrangements," *The American Economic Review*, 2017, *107* (12), 3722–3759.

NCES, "National Center for Education Statistics, National Study of Postsecondary Faculty," 1999.

Newton, Dale A and Martha S Grayson, "Trends in Career Choice by US medical School Graduates," *JAMA*, 2003, *290* (9), 1179–1182.

Nicholson, Sean, "Physician Specialty Choice under Uncertainty," *Journal of Labor Economics*, 2002, *20* (4), 816–847.

NRMP, "National Residency Matching Program Results and Data 2000 Match," Technical Report 2000.

Pertold-Gebicka, Barbara, Filip Pertold, and Nabanita Datta Gupta, "Employment Adjustments around Childbirth," *IZA DP No. 9685*, 2016.

Philibert, Ingrid, Betty Chang, Timothy Flynn, Paul Friedmann, Rebecca Minter, Eric Scher, and W. T. Williams, "The 2003 Common Duty Hour Limits: Process, Outcome, and Lessons Learned," *Journal of Graduate Medical Education*, 2009, *1* (2), 334–337.

Polachek, Solomon, "Occupational Self-selection: A Human Capital Approach to Sex Differences in Occupational Structure," *The Review of Economics and Statistics*, 1981, *63* (1), 60–69.

Rangel, Erika L., Douglas S. Smink, Manuel Castillo-Angeles, Gifty Kwakye, Marguerite Changala, Adil H. Haider, and Gerard M. Doherty, "Pregnancy and Motherhood During Surgical Training," *JAMA Surgery*, 2018.

Sandberg, Sheryl, *Lean In*, New York: Alfred A. Knopf, 2013.

Sasser, Alicia C, "Gender Differences in Physician Pay Tradeoffs Between Career and Family," *Journal of Human Resources*, 2005, *60* (2), 477–504.

Slaughter, Anne Marie, *Unfinished Business: Women Men Work Family*, New York: Random House, 2015.

Staiger, Douglas O, David I Auerbach, and Peter I Buerhaus, "Trends in the Work Hours of Physicians in the United States," *JAMA*, 2010, *303* (8), 747–753.

Sullivan, Michael C, Heather Yeo, Sanziana a Roman, Maria M Ciarleglio, Xiangyu Cong, Richard H Bell, and Julie a Sosa, "Surgical residency and attrition: defining the individual and programmatic factors predictive of trainee losses.," *Journal of the American College of Surgeons*, mar 2013, *216* (3), 461–71.

Szymczak, Julia E, Joanna Veazey Brooks, Kevin G Volpp, and Charles L Bosk, "To Leave or to Lie? Are Concerns about a Shift-Work Mentality and Eroding Professionalism as a Result of Duty-Hour Rules Justified?," *The Millbank Quarterly*, 2010, *88* (3), 350–381.

Turner, Patricia L, Kimberly Lumpkins, Joel Gabre, Maggie J Lin, Xinggang Liu, and Michael Terrin, "Pregnancy Among Women Surgeons: Trends over Time," *Archives of Surgery*, 2012, *147* (5), 474–9.

U.S. Census Bureau, "Current Population Survey 1989-2014 [Data Set]," Technical Report, Accessed via National Bureau of Economic Research [Distributor] 2015. https://data.nber.org/cps-basic2/raw/ and https://data.nber.org/cps-basic2/dta/.

USDHHS, "The Physician Workforce : Projections and Research into Current Issues Affecting Supply and Demand," *U.S. Department of Health and Human Services*, 2008, (December).

Volpp, Kevin G, Amy K Rosen, Paul R Rosenbaum, Patrick S Romano, Orit Even-Shoshan, Anne Canamucio, Lisa Bellini, Tiffany Behringer, and Jeffrey H Silber, "Mortality Among Patients in VA Hospitals in the First 2 Years Following ACGME Resident Duty Hour Reform," *JAMA*, 2007, *298* (9), 984–992.

_ , _ , _ , _ , _ , Yanli Wang, Lisa Bellini, Tiffany Behringer, and Jeffrey H Silber, "Mortality Among Hospitalized Medicare Beneficiaries in the First 2 Years Following ACGME Resident Duty Hour Reform," *JAMA*, 2007, *298* (9), 975–83.

Volpp, Kevin G., Dylan S. Small, Patrick S. Romano, Kamal M F Itani, Amy K. Rosen, Orit Even-Shoshan, Yanli Wang, Lisa Bellini, Michael J. Halenar, Sophia Korovaichuk, Jingsan Zhu, and Jeffrey H. Silber, "Teaching Hospital Five-year Mortality Trends in the Wake of Duty Hour Reforms," *Journal of General Internal Medicine*, 2013, *28* (8), 1048–1055.

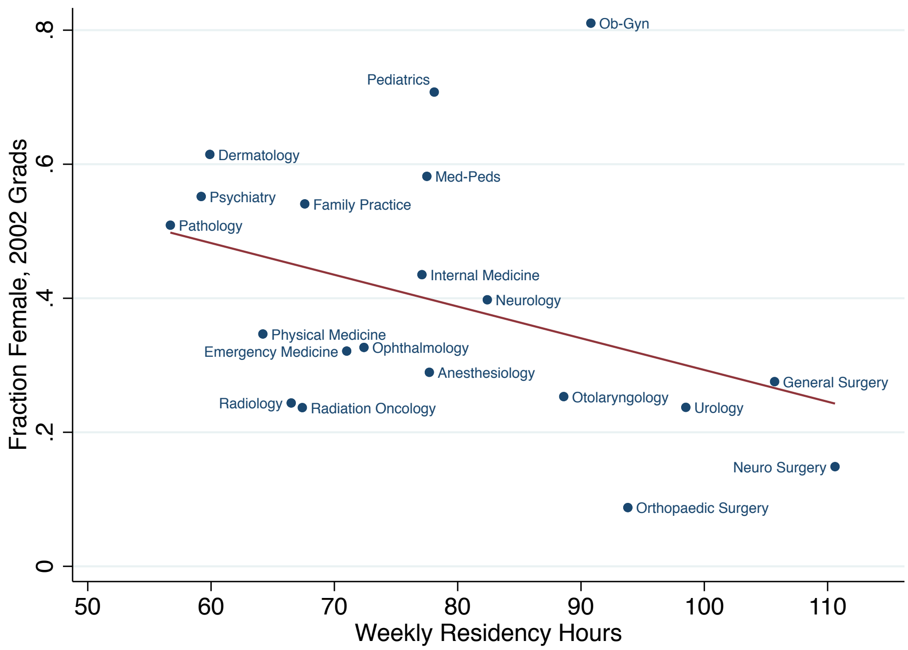
Wiswall, Matthew and Basit Zafar, "Preference for the Workplace, Human Capital, and Gender," *The Quarterly Journal of Economics*, 2018, *133* (1), 457–507.

Yeo, Heather, Emily Bucholz, Julie Ann Sosa, Leslie Curry, Frank R Lewis, Andrew T Jones, Kate Viola, Zhenqui Lin, and Richard H Bell, "A national study of attrition in general surgery training: which residents leave and where do they go?," *Annals of Surgery*, 2010, *252* (3), 529–534; discussion 534–536.

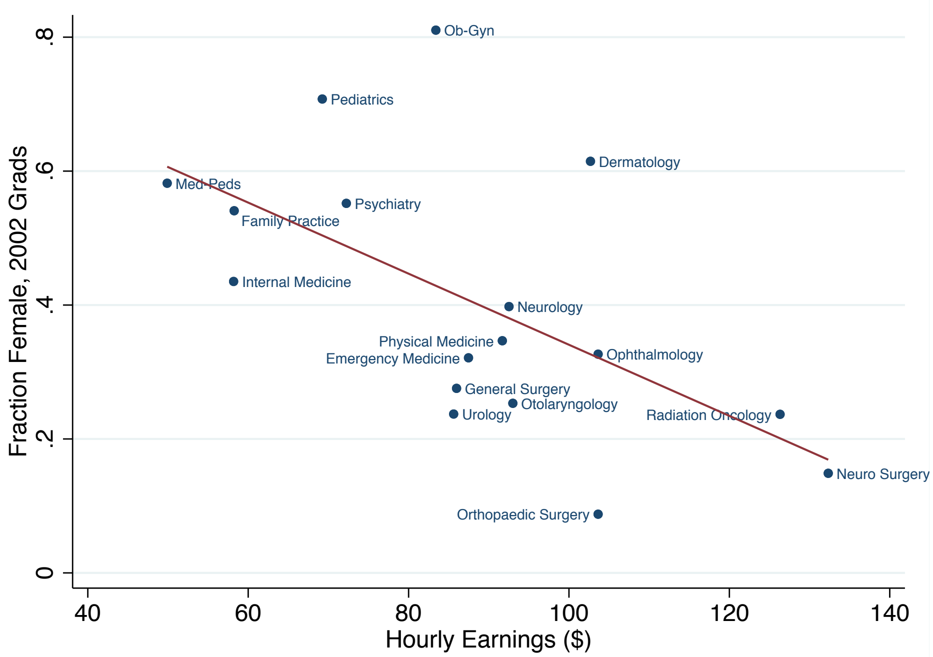
Yoon, Harry H., "Adapting to Duty-Hour Limits: Four Years On," *New England Journal of Medicine*, 2007, *356*, 2668–2670.

Figure 1: Relationship between Female Representation in Medical Specialties and Selected Specialty Characteristics

A. Female Share of Specialties and Average Weekly Hours during Residency

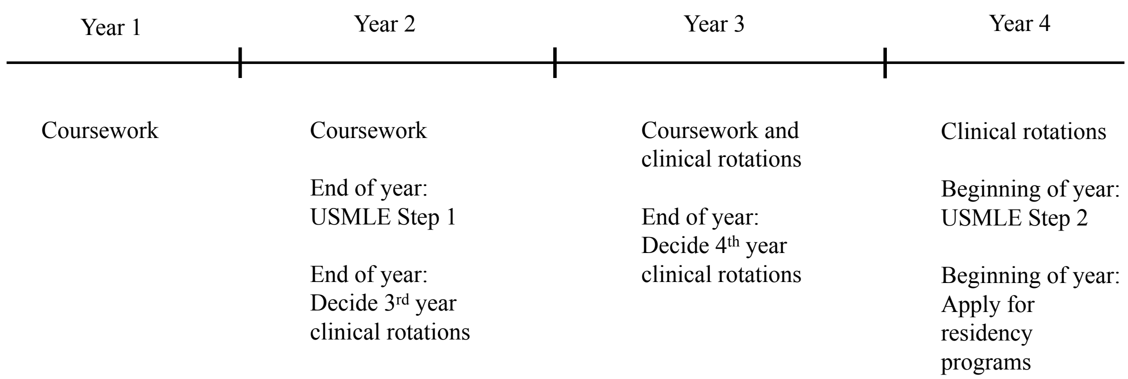


B. Female Share of Specialties and Average Earnings Post-residency



Source: AMA Physician Masterfile, Baldwin Jr et al. (2003), Leigh et al. (2010). Note: This figure plots the fraction of each specialty that is female, using the 2002 U.S. medical school graduation cohort, against: in Panel A, the average hours per week worked during the second year of medical residency from a survey of medical residents in 1998/9 by Baldwin Jr et al. (2003); and in Panel B, average post-residency hourly earnings from the Community Tracking Study, as reported in Leigh et al. (2010). The solid line in Panel A (B) represents the line of best fit from a regression of female share on hours (earnings).

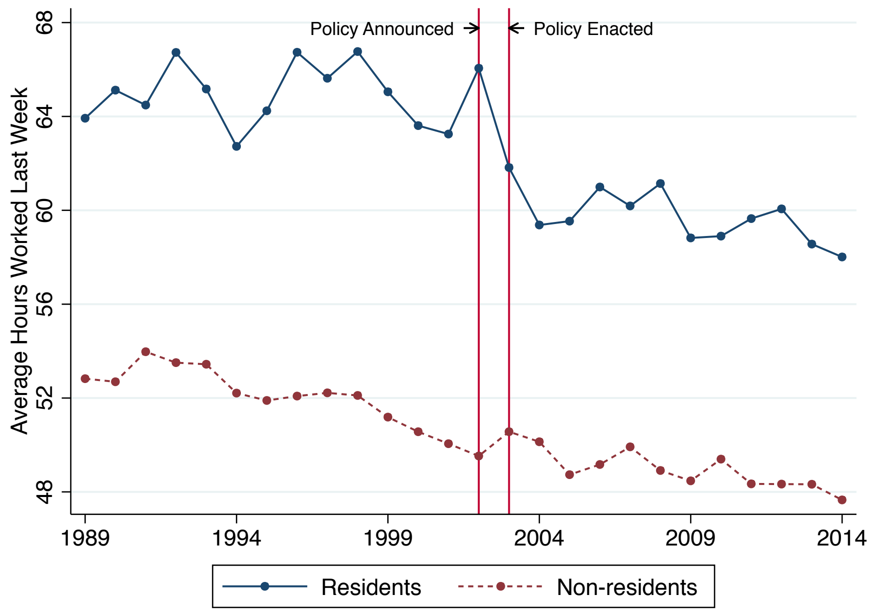
Figure 2: Medical School Timeline



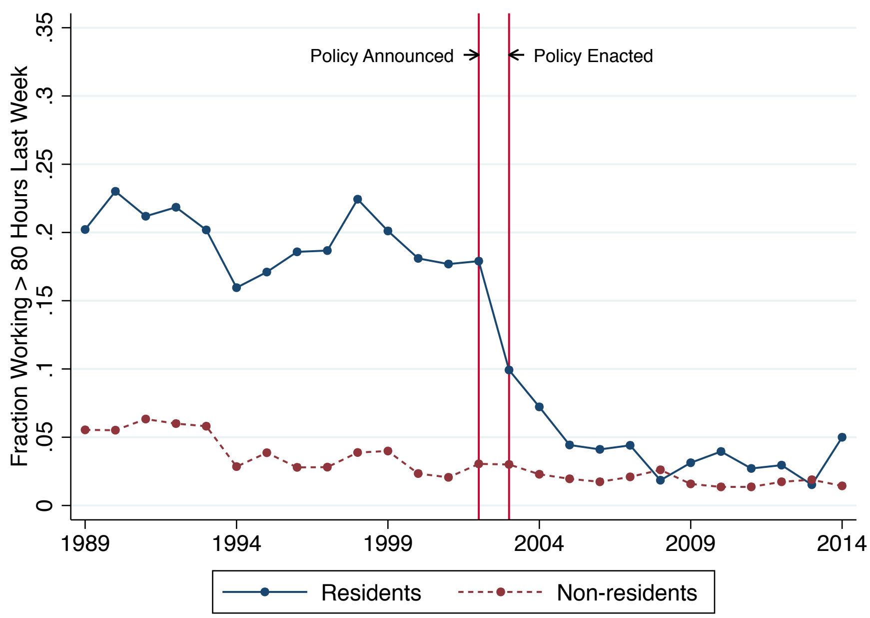
Source: Association of American Medical Colleges *Medical School Admission Requirements United States and Canada 2001-2002.*

Figure 3: Hours Worked in the Prior Week among Resident and Non-Resident Physicians, 1989-2014

A. Average Hours



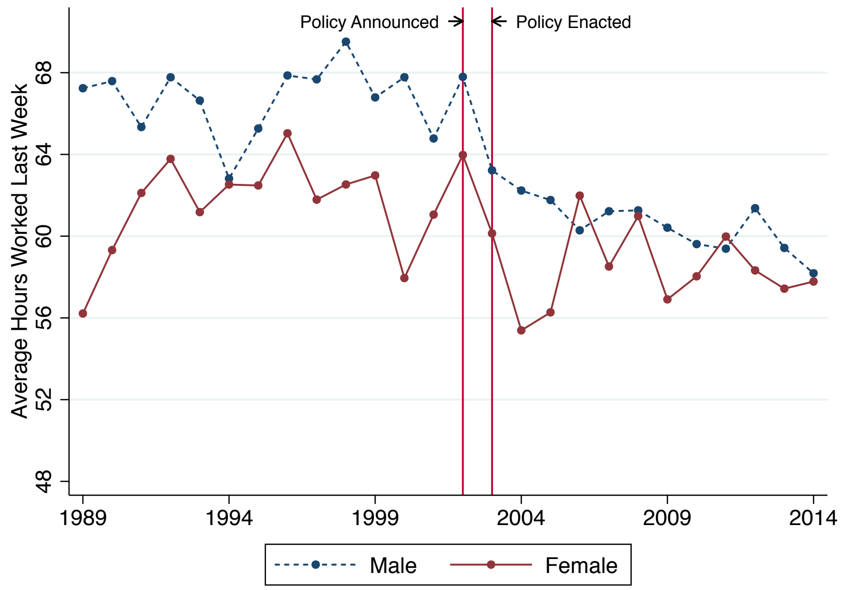
B. Fraction Working > 80 Hours



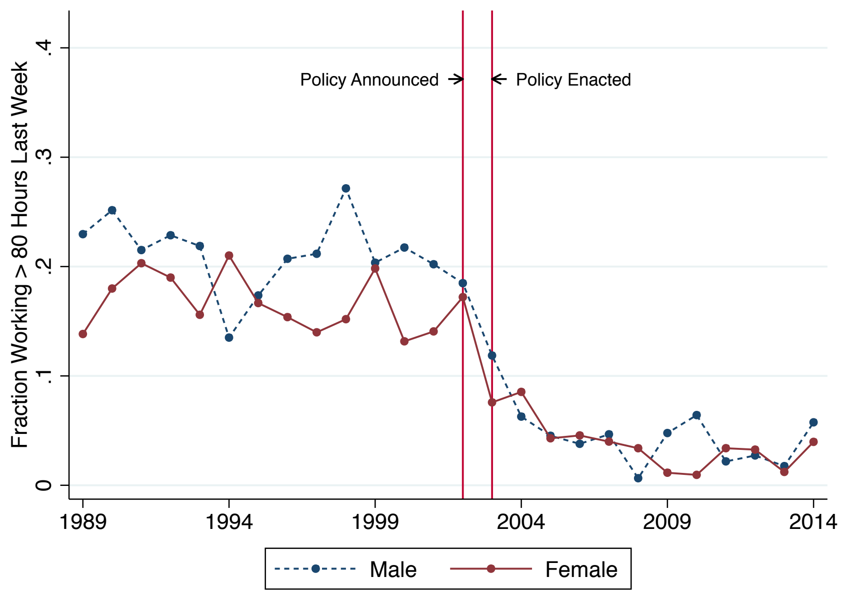
Source: Current Population Survey, monthly files January 1989-December 2014. Note: Panel A plots the average number of hours worked last week for physicians, separately for residents and non-residents. Panel B plots the fraction of physicians who worked more than 80 hours last week, separately for residents and non-residents. Resident status is imputed based on age (<35) and whether the individual works in a hospital. CPS sampling weights are used.

Figure 4: Hours Worked in the Prior Week among Male and Female Resident Physicians, 1989-2014

A. Average Hours

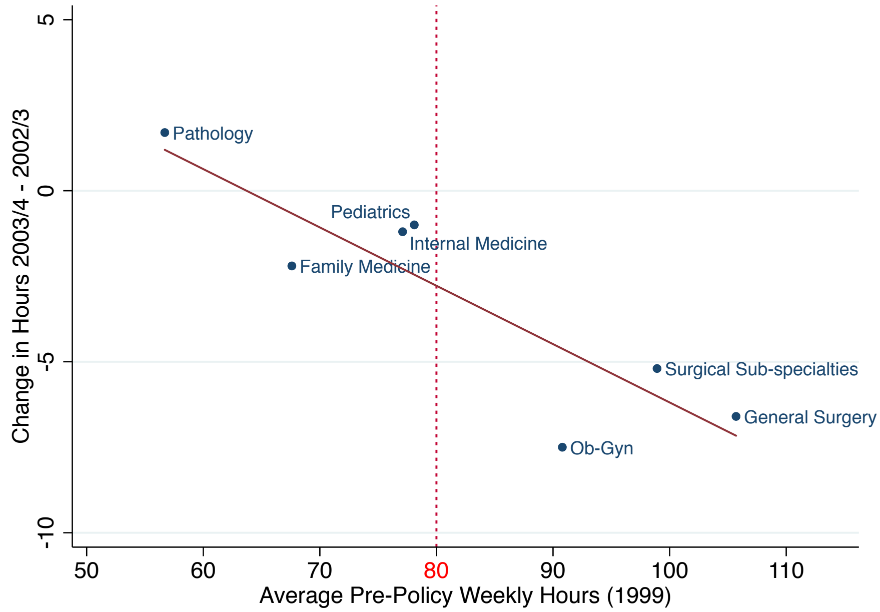


B. Fraction Working > 80 Hours



Source: Current Population Survey, monthly files January 1989-December 2014. Panel A plots the average number of hours worked last week for resident physicians, separately for men and women. Panel B plots the fraction of physicians who worked more than 80 hours last week, separately for men and women. Resident status is imputed based on age (<35) and whether the individual works in a hospital. CPS sampling weights are used.

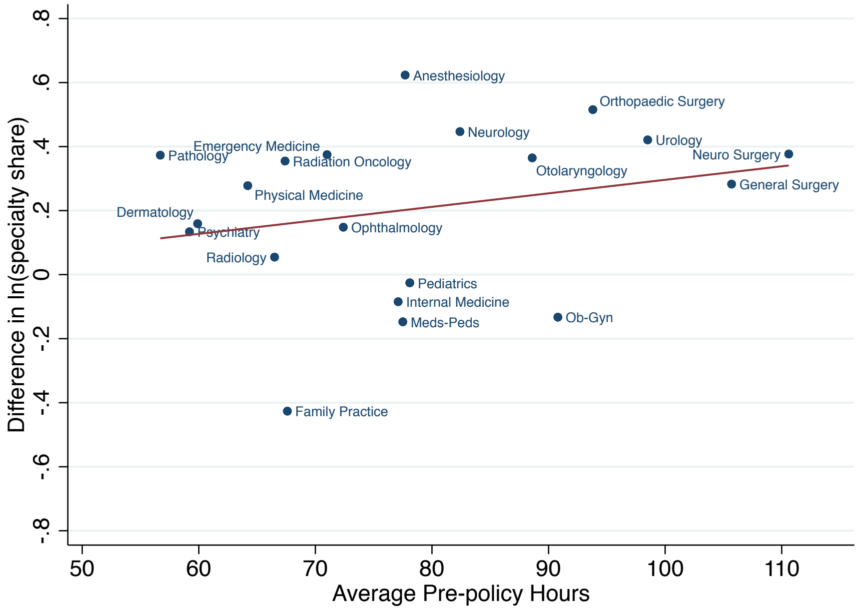
Figure 5: Relationship between 1999 Hours and the Change in Hours 2002/3-2003/4, by Specialty



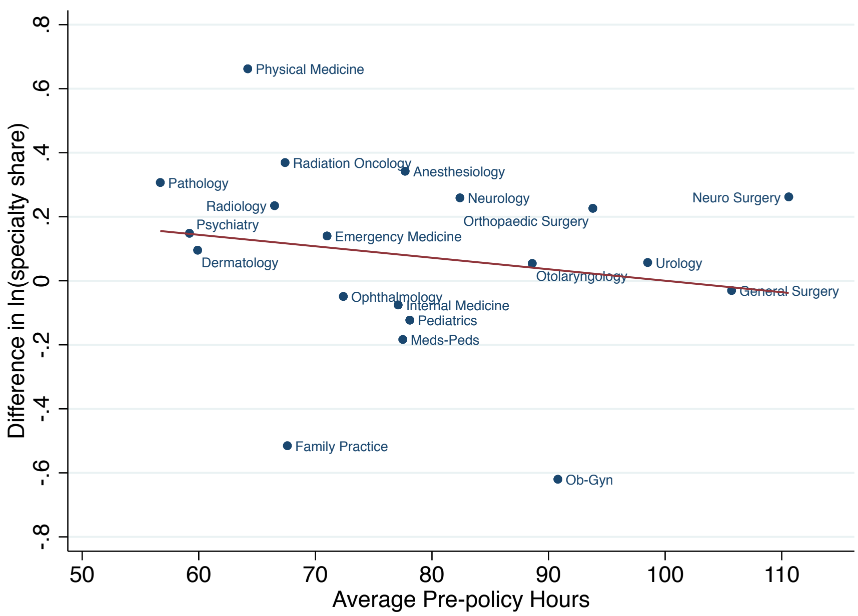
Source: Baldwin Jr et al. (2003), Landrigan et al. (2006) and personal correspondence with author. Note: This figure plots the average number of hours worked per week for second year medical resident physicians in 1999 on the x-axis. The change in average hours per week between residency years 2003/4 and 2002/3 for first year residents is plotted on the y-axis.

Figure 6: Relationship between Change in Specialty Shares and Pre-Policy Specialty Hours

A. Women



B. Men

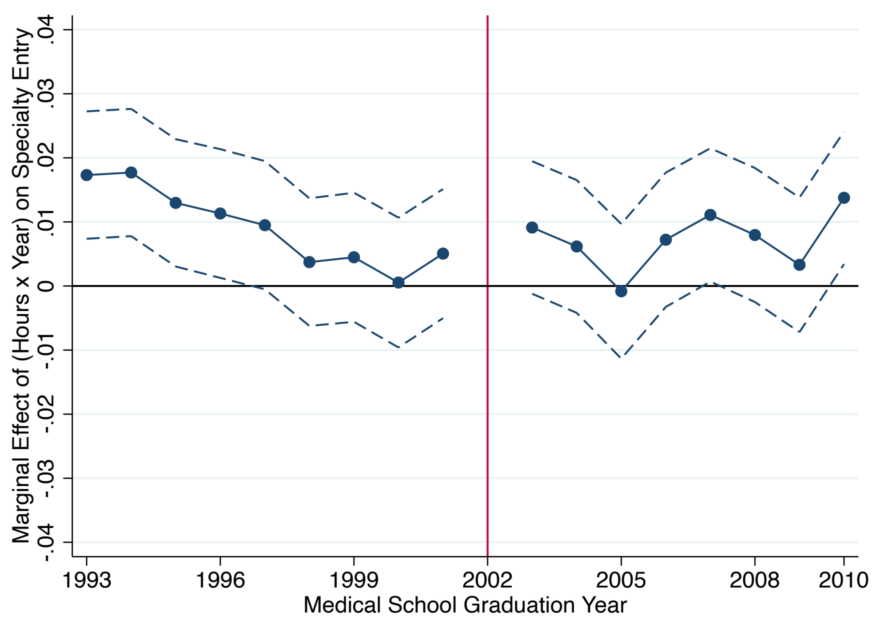


Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This figure plots the change in the natural log share of women and men entering a medical specialty before (1993-2002) and after (2003-2010) the reform against a specialty's average pre-policy hours, for the sample of U.S. medical school graduates, 1993-2010. The line of best fit is from a regression of the difference in log shares pre/post reform on pre-policy hours.

Figure 7: Effect of the Reform on Medium-Term Specialty Choice: Conditional Logit Event Study

A. Women

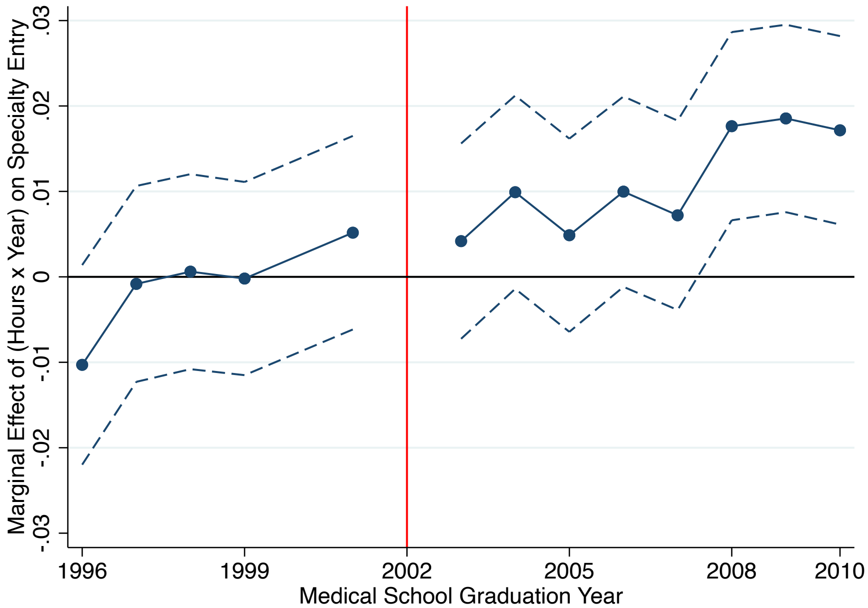
B. Men



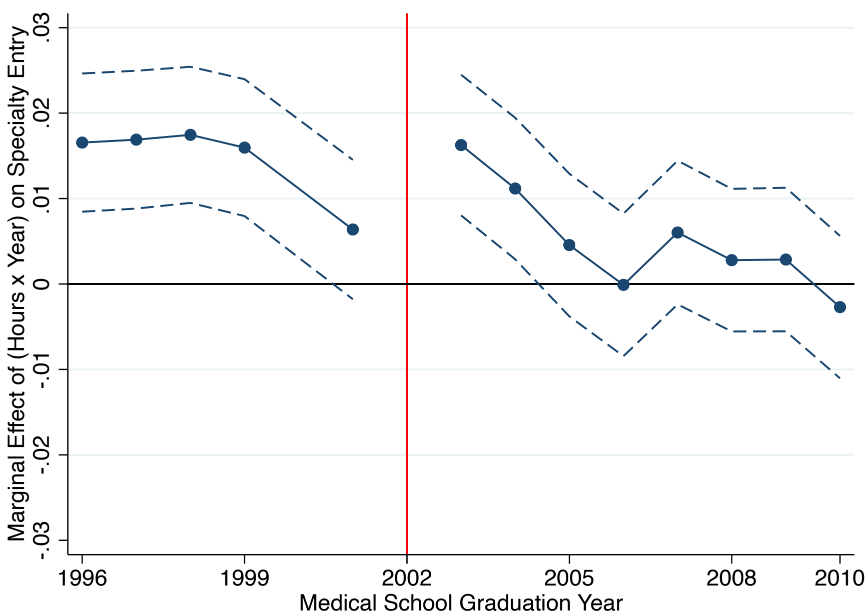
Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This figure plots the average marginal effects associated with the coefficients from the conditional logit event study. The sample is U.S. medical school graduates, 1993-2010. The dependent variable is specialty outcome and the explanatory variables are specialty fixed effects, and interactions of specialty pre-policy average hours with medical school cohort fixed effects. Cohort 2002 is omitted as the reference year. The solid line plots the average marginal effects of the interaction term ($\text{Hours}_{s,1999} \times \text{Year}$). The dashed lines plot the 95% confidence intervals based on heteroskedastic robust standard errors.

Figure 8: Effect of the Reform on Initial Specialty Choice: Conditional Logit Event Study

A. Women

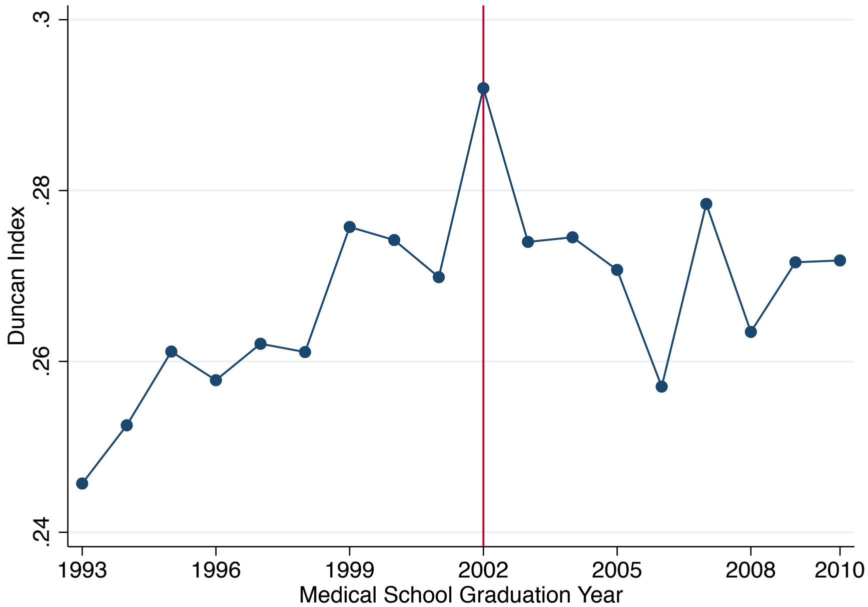


B. Men



Source: GME Census Track, Baldwin Jr et al. (2003). Note: This figure plots the average marginal effects associated with the coefficients from the conditional logit event study. The sample is U.S. medical school graduates in their first year of a residency program, for years 1996-2010. The dependent variable is specialty outcome and the explanatory variables are specialty fixed effects, and interactions of specialty pre-policy average hours with residency start year fixed effects. Residency year 2002 is omitted as the reference year. The solid line plots the average marginal effects of the interaction term ($\text{Hours}_{s,1999} \times \text{Year}$). The dashed lines plot the 95% confidence intervals based on heteroskedastic robust standard errors.

Figure 9: Duncan Segregation Index, 1993-2010



Source: AMA Physician Masterfile. Note: This figure plots the Duncan Segregation Index for the sample is U.S. medical school graduates, 1993-2010. The metric represents the fraction of women (or men) who would have to change specialties in order for the specialty distribution of men and women to be identical.

Table 1: Summary Statistics: Full and U.S. Medical School Graduate Samples

	(1)	(2)	(3)	(4)	(5)	(6)
	Full Sample			USMG Sample		
	All	Female	Male	All	Female	Male
	I: AMA Masterfile					
Female	0.44	-	-	0.44	-	-
Age at Medical School Graduation	27.90	27.73	28.04	28.27	28.15	28.36
	(3.67)	(3.69)	(3.64)	(3.35)	(3.44)	(3.28)
U.S. Born	0.63	0.63	0.63	0.83	0.83	0.83
Attended Ranked Medical School	0.33	0.33	0.32	0.48	0.48	0.48
Foreign Medical School	0.24	0.24	0.24	-	-	-
Osteopathic Medical School	0.08	0.08	0.08	-	-	-
N	413,989	181,826	232,163	281,433	124,797	156,636
	II: GME Census Track					
Female	0.43	-	-	0.43	-	-
Foreign or Osteopathic Medical School	0.32	0.32	0.32	-	-	-
N	396,674	171,831	224,843	270,887	117,629	153,258

Source AMA Physician Masterfile, GME Census Track. Note: The AMA Masterfile includes individuals who graduated from medical school years 1993-2010. The GME Census Track includes individuals who were in the first year of a residency program 1996-2010 (excluding 2000). The Full Sample includes all medical school graduates, including foreign medical school graduates and osteopaths. The USMG Sample includes only U.S. medical school graduates. Medical school rank is determined by the inclusion of the medical school in U.S. News and World Report 2014 rankings. Standard deviations are reported in parentheses.

Table 2: Effect of the Reform on Specialty Choice:
Conditional Logit Average Marginal Effects

Dependent Variable: Specialty Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I: Medium-Run Specialty Outcome, AMA Masterfile							
Panel A: Female							
Average Weekly Hours × Transition	0.002	0.012***	0.000	-0.007**	0.000	0.002	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Average Weekly Hours × Post	0.023***	0.037***	0.019***	0.014***	0.019***	0.023***	0.014***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Panel B: Male							
Average Weekly Hours × Transition	-0.004	-0.007***	-0.003	-0.016***	-0.031***	-0.010***	0.007***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Average Weekly Hours × Post	0.000	-0.006***	0.001	-0.012***	-0.034***	-0.009***	0.018***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
P-value for test of equality of male/female coeff.							
Average Weekly Hours × Transition	0.169	0.000	0.591	0.031	0.000	0.002	0.004
Average Weekly Hours × Post	0.000	0.000	0.000	0.000	0.000	0.000	0.239
II: Initial Specialty Entry, GME Census Track							
Panel A: Female							
Average Weekly Hours × Transition	0.007**	0.008***	-	-0.002	0.002	0.006**	0.004
	(0.003)	(0.003)		(0.003)	(0.003)	(0.003)	(0.003)
Average Weekly Hours × Post	0.015***	0.016***	-	0.005**	0.007***	0.013***	0.011***
	(0.002)	(0.003)		(0.002)	(0.002)	(0.002)	(0.002)
Panel B: Male							
Average Weekly Hours × Transition	-0.002	-0.007***	-	-0.010***	-0.029***	-0.005**	0.007***
	(0.002)	(0.002)		(0.002)	(0.003)	(0.002)	(0.002)
Average Weekly Hours × Post	-0.011***	-0.020***	-	-0.022***	-0.049***	-0.017***	0.005***
	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
P-value for test of equality of male/female coeff.							
Average Weekly Hours × Transition	0.011	0.000	-	0.025	0.000	0.003	0.413
Average Weekly Hours × Post	0.000	0.000	-	0.000	0.000	0.000	0.055
Specialty FE	X	X	X	X	X	X	X
Ob/Gyn., Primary Care time trends		X					
Controls for age, medical school rank			X				
Specialty characteristics X cohort FE				X			
Fraction female X cohort FE					X		
Competitiveness X cohort FE						X	
All specialty linear pre-trends							X

Source: AMA Physician Masterfile, GME Census Track, Baldwin Jr et al. (2003). Note: This table reports the results of maximum likelihood estimation of a conditional logit model in which the baseline specification has specialty outcome as the dependent variable and the explanatory variables include specialty fixed effects and specialty pre-policy hours ($\text{Hours}_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). The average marginal effects associated with the coefficient on the interaction terms from Appendix Table A.4 are reported. The average marginal effect is the average of individual specialty-specific marginal effects for all individuals in the sample. Heteroskedastic robust standard errors are reported in parentheses. Column 1 reports the results of the baseline specification with no additional controls. Columns 2-7 progressively include specialty-specific and demographic controls. The p-values at the bottom of each panel are from a Wald test of the null hypothesis that the male and female coefficients are equal.

Table 3: Effect of the Reform on Stated Specialty Preference

Dependent Variable: $\ln(\text{Share}_{st})$*100	(1)	(2)	(3)	(4)
Panel A: Female				
Average Weekly Hours × Transition	0.32	0.30	0.40	0.36
	(0.43)	(0.33)	(0.45)	(0.35)
Average Weekly Hours × Post	0.88*	0.85*	0.99*	0.97**
	(0.50)	(0.44)	(0.52)	(0.37)
Panel B: Male				
Average Weekly Hours × Transition	0.27	0.26	0.33	0.27
	(0.44)	(0.35)	(0.47)	(0.37)
Average Weekly Hours × Post	0.32	0.31	0.39	0.32
	(0.52)	(0.41)	(0.58)	(0.51)
P-value for test of equality of male/female coeff.				
Average Weekly Hours × Transition	0.843	0.872	0.824	0.826
Average Weekly Hours × Post	0.085	0.111	0.076	0.278
N	198	198	198	198
Specialty FE	X	X	X	X
Ob/Gyn., Primary Care time trends		X		
Specialty characteristics X cohort FE			X	
All specialty linear pre-trends				X

Source: AAMC Matriculating Student Questionnaire 1998-2006, 2010, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-differences specification for specialty stated preference, estimated separately for men and women. The dependent variable is the natural logarithm of the share of women (men) who reported considering a specialty in the MSQ, in a given survey year, multiplied by 100. The explanatory variables include specialty fixed effects, medical school graduation cohort fixed effects and specialty pre-policy hours ($\text{Hours}_{s,1999}$) interacted with an indicator for entering medical school 2003-2005 (Transition) and 2006-2010 (Post). The coefficients on the interaction terms are reported. All specifications have 198 observations stemming from the analysis of 18 specialties over 11 years. Radiation Oncology and Internal Medicine-Pediatrics are not included in the MSQ. Column 1 reports the results of the baseline specification with no additional controls. Columns 2-4 include time-varying specialty controls. Standard errors clustered at the specialty level are reported in parentheses. The p-values are from a Wald test of the null hypothesis that the male and female coefficients are equal.

Table 4: Residency Program Drivers of Female Entry

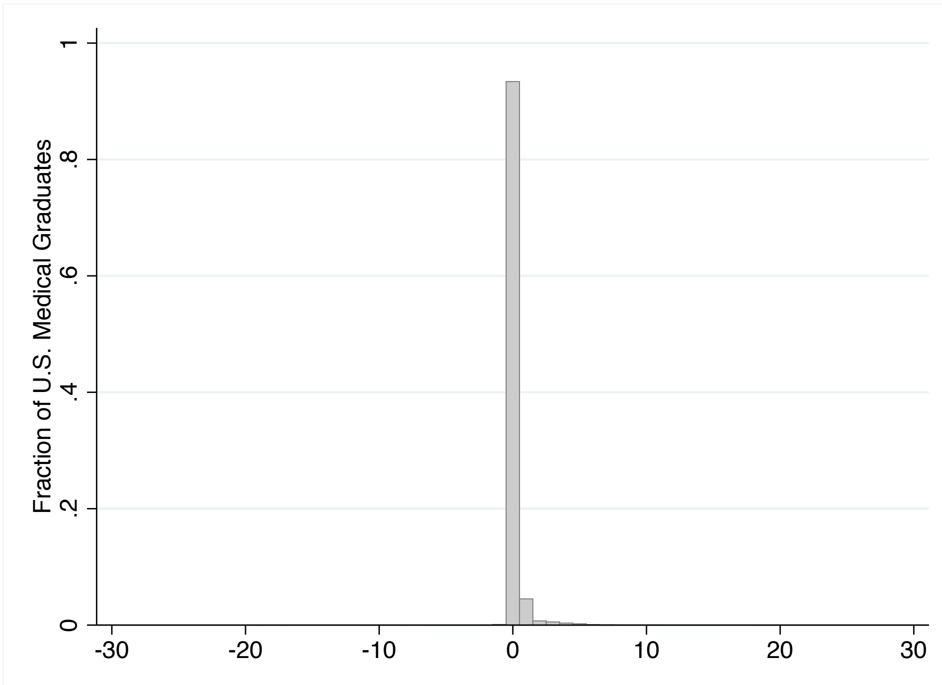
Dependent Variable: $\text{FractionFemale}_{pst}$	(1)	(2)	(3)	(4)
Panel A: Overall				
Average Weekly Hours × Transition	0.008	-0.007	-0.009	
	(0.044)	(0.039)	(0.039)	
Average Weekly Hours × Post	0.078**	0.053	0.048*	
	(0.034)	(0.040)	(0.026)	
N	31,969	31,969	31,969	
Ob/Gyn. and Primary Care specialty time trends		X		
All specialty linear pre-trends			X	
Panel B: By Baseline Female Representation				
	Faculty High	Faculty Low	Residents High	Residents Low
Average Weekly Hours × Transition	0.060	-0.033	0.007	-0.001
	(0.042)	(0.067)	(0.051)	(0.058)
Average Weekly Hours × Post	0.116***	0.062	0.115***	0.037
	(0.038)	(0.047)	(0.039)	(0.043)
N	12,167	15,670	13,112	14,062
Panel C: By Family-Friendly Policies				
	Paid Maternity Leave Policy	No Paid Maternity Leave Policy	Onsite Childcare	No Onsite Childcare
Average Weekly Hours × Transition	-0.021	0.030	-0.019	0.020
	(0.042)	(0.053)	(0.061)	(0.042)
Average Weekly Hours × Post	0.050	0.099**	0.091**	0.077**
	(0.034)	(0.044)	(0.036)	(0.036)
N	12,088	15,867	10,944	16,150

Source: GME Census Track, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-differences specification estimating the effect of the duty hour reform on residency program gender composition. The dependent variable is the fraction of first year US medical school graduate residents who are female in a given residency program for a given calendar year multiplied by 100. The explanatory variables include program fixed effects, calendar year fixed effects and specialty-specific pre-policy hours interacted with an indicator for entering residency in 2003-2005 (Transition) and 2006-2010 (Post). In Panel A, column 1 reports the results of the baseline specification with no additional controls. Columns 2-3 progressively include specialty- and program-specific controls. Panel B examines heterogeneity in the effects of the reform on program gender composition based on baseline program gender composition among full-time faculty and residents. Panel C examines heterogeneity in the effects of the reform by baseline program availability of paid maternity leave and onsite childcare. Standard errors clustered at the specialty level are reported in parentheses.

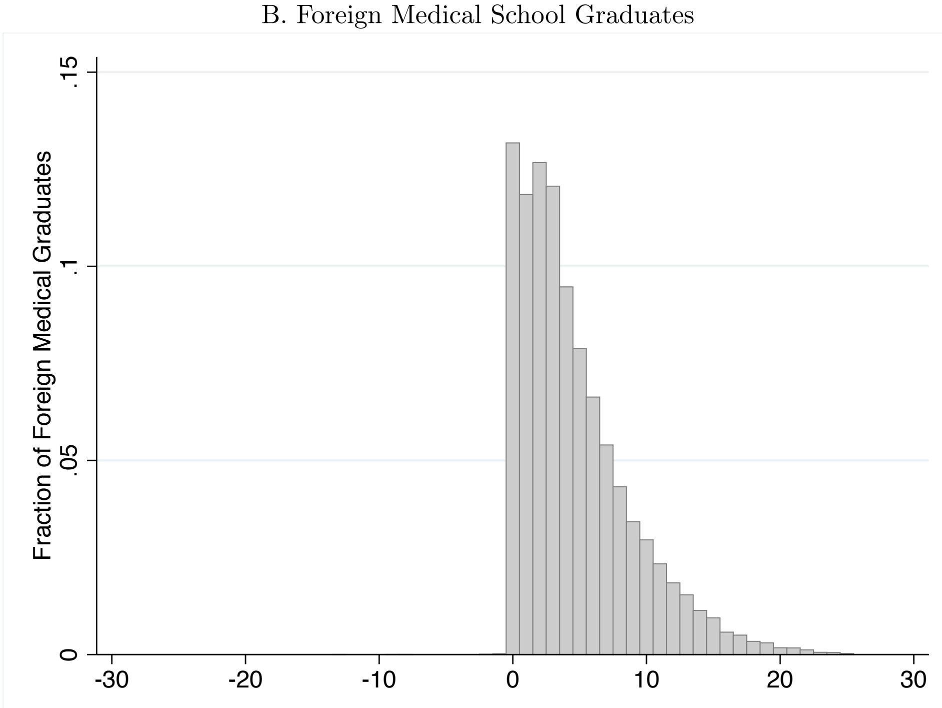
A Supplemental Figures and Tables

Figure A.1: Distribution of Difference between Residency Start Year and Medical School Graduation Year

A. U.S. Medical School Graduates



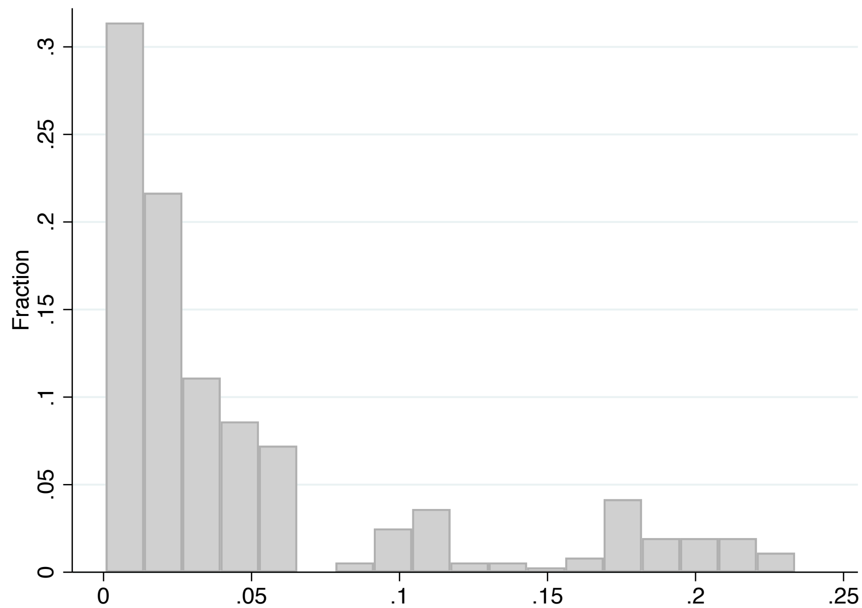
B. Foreign Medical School Graduates



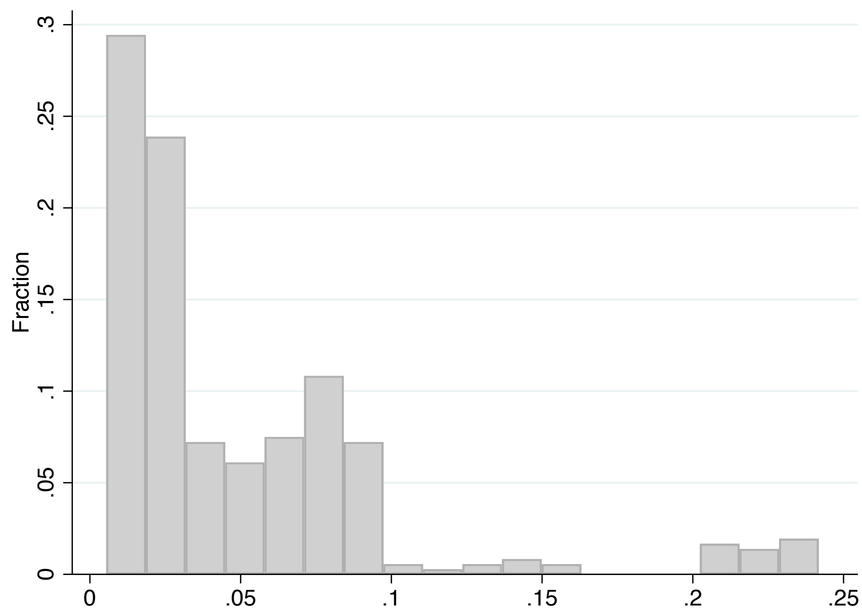
Source: AMA Physician Masterfile Training Supplement for California and Texas Note: This figure plots distribution of the difference between medical school graduation year and residency start year for individuals who graduated from U.S. medical schools and foreign medical schools and attended residency training in California or Texas.

Figure A.2: Distribution of Specialty Shares

A. Female



B. Male

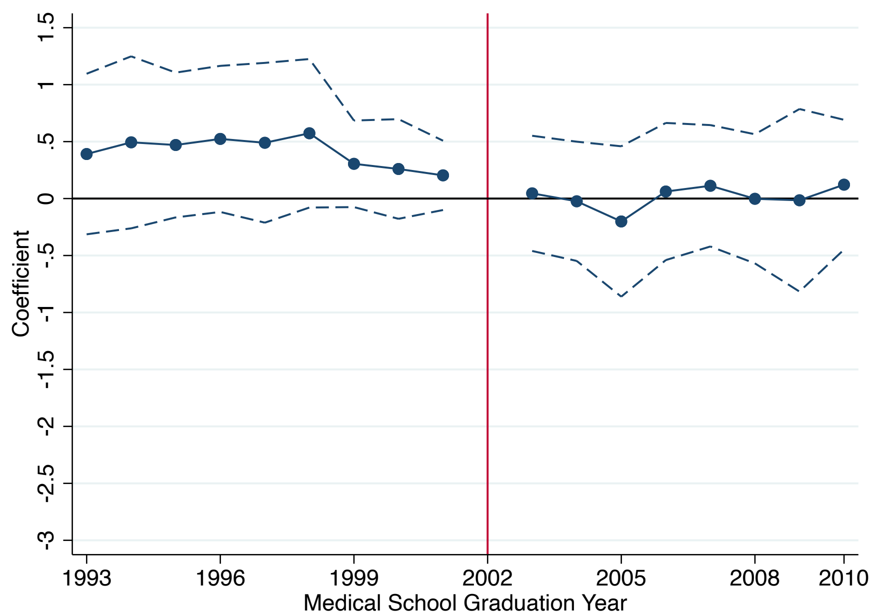


Source: AMA Physician Masterfile Note: This figure plots distribution of specialty shares for male and female U.S. medical school graduates who graduated from medical school 1993 to 2010.

Figure A.3: Effect of the Reform on Specialty Entry: OLS Event Study

A. Women

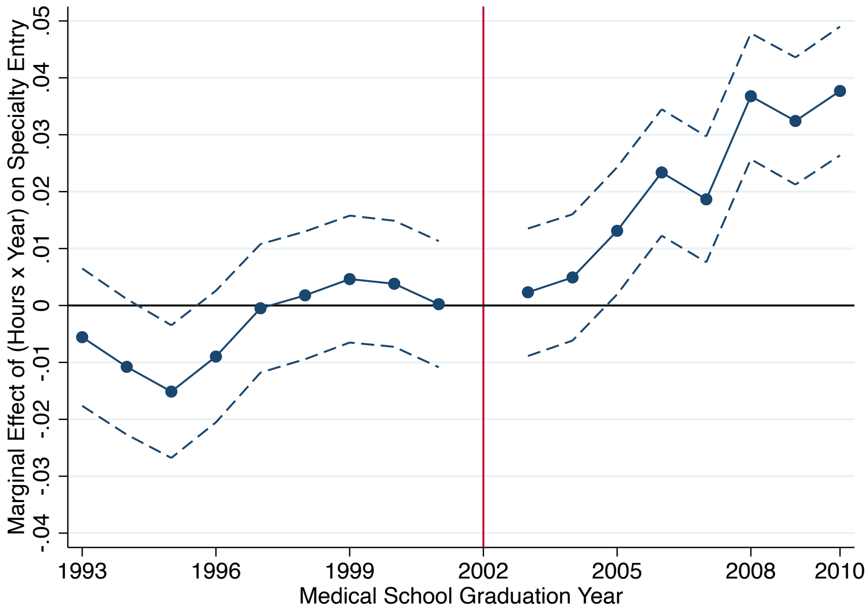
B. Men



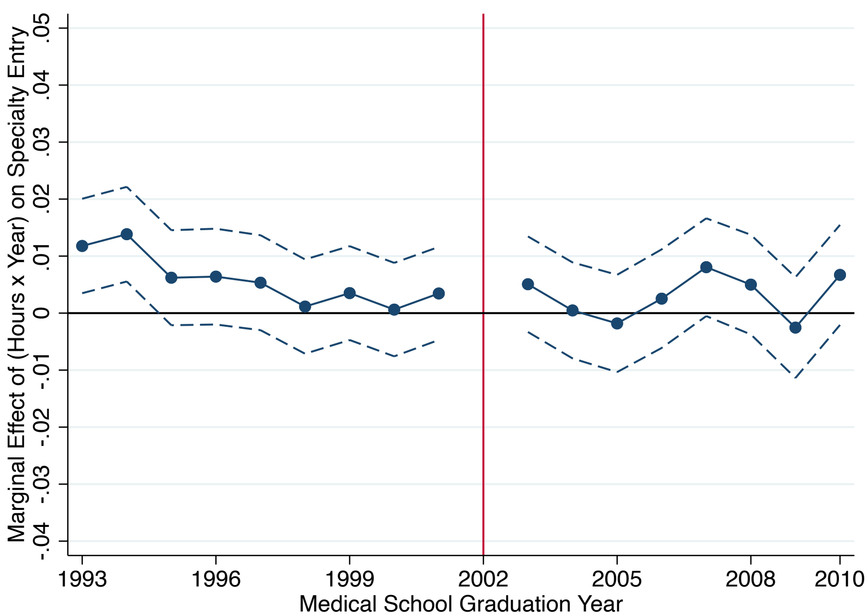
Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This figure plots the coefficients from the OLS event study model. The dependent variable is the natural logarithm of the share of women (men) from a given medical school graduation cohort in a specialty, multiplied by 100. The explanatory variables include specialty fixed effects, medical school graduation cohort fixed effects, and specialty pre-policy hours interacted with indicator variables for medical school cohorts. Cohort 2002 serves as the reference year. The solid line plots the coefficients on the interaction term ($\text{Hours}_{s,1999} \times \text{Year}$). The dashed lines plot the 95% confidence intervals based on standard errors clustered at the specialty level.

Figure A.4: Effect of the Reform on Medium-Term Specialty Entry: Conditional Logit Event Study Including Foreign and Osteopathic Medical School Students

A. Women



B. Men



Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This figure plots the average marginal effects from maximum likelihood estimation of the conditional logit event study model for the sample of all medical school graduates, 1993-2010. The dependent variable is specialty outcome and the explanatory variables are specialty fixed effects, and interactions of specialty pre-policy average hours with medical school cohort fixed effects. Cohort 2002 is omitted as the reference year. The solid line plots the average marginal effects of the interaction term ($\text{Hours}_{s,1999} \times \text{Year}$). The dashed lines plot the 95% confidence intervals based on heteroskedastic robust standard errors.

Table A.1: Pre-Policy Specialty Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		AMA Masterfile 1993-2002		GME Census Track 1996-2002		Freeman (2003)	
Specialty	Avg Pre-Policy Weekly Hours	Share of Men (%)	Share of Women (%)	Share of Men (%)	Share of Women (%)	Modal Residency Length	% Residency Applicants Unmatched
Pathology	56.7	1.7	1.9	1.4	1.7	4	2.1
Psychiatry	59.2	3.5	5.1	2.7	4.3	4	4.3
Dermatology	59.9	1.3	2.5	1.0	1.8	4	16.1
Physical medicine/rehab	64.2	1.3	1.1	1.0	0.9	4	2.7
Radiology	66.5	6.7	3.3	4.9	2.6	5	11.1
Radiation oncology	67.4	0.8	0.5	0.6	0.3	5	17.9
Family practice	67.6	13.2	17.1	10.9	14.9	3	1.9
Emergency medicine	71.0	7.6	4.2	6.3	3.9	3.5	6.5
Ophthalmology	72.4	3.0	1.9	2.3	1.5	4	22.0
Internal medicine	77.1	23.1	22.0	25.1	25.1	3	1.6
Internal medicine/peds	77.5	1.2	1.7	1.6	2.2	4	-
Anesthesiology	77.7	5.7	2.9	4.1	2.2	4	5.2
Pediatrics	78.1	6.4	17.8	5.5	16.4	3	1.3
Neurology	82.4	1.5	1.2	1.1	1.0	4	3.0
Otolaryngology	88.6	2.0	0.7	1.7	0.6	5	15.0
Obstetrics/Gynecology	90.8	3.3	10.8	2.5	9.8	4	4.3
Orthopedic surgery	93.8	5.4	0.6	4.4	0.7	5	14.9
Urology	98.5	2.1	0.5	1.5	0.3	5	20.0
General surgery	105.7	9.0	3.9	15.0	6.4	5	5.4
Neurological surgery	110.6	1.2	0.2	1.0	0.2	6	15.0
Transitional year	80.1	-	-	5.5	3.3	1	-

Source: AMA Physician Masterfile, GME Census Track, Baldwin Jr et al. (2003). Note: This table reports the distribution of men and women across medical specialties for the pre-policy period (1993-2002 for the AMA Masterfile and 1996-2002 for the GME Census Track). For each specialty, it reports the modal residency length (in years) and percentage of residency applicants to the specialty who are unmatched, sourced from Freeman (2003).

Table A.2: Comparison of Masterfile Sample and AAMC Data on Medical School Graduates

	(1) USMG Sample 1993-2010	(2) AAMC Graduation Data	(3) % difference
Medical School Graduation Year			
1993	15,237	15,474	1.53
1994	15,239	15,504	1.71
1995	15,646	15,883	1.49
1996	15,613	15,895	1.77
1997	15,678	15,894	1.36
1998	15,669	15,972	1.90
1999	15,639	16,006	2.29
2000	15,390	15,716	2.07
2001	15,582	15,796	1.35
2002	15,364	15,676	1.99
2003	15,342	15,531	1.22
2004	15,777	15,829	0.33
2005	15,367	15,760	2.49
2006	15,677	15,927	1.57
2007	15,798	16,140	2.12
2008	15,939	16,168	1.42
2009	16,148	16,467	1.94
2010	16,328	16,838	3.03

Source: AMA Physician Masterfile, American Association of Medical Colleges Data Warehouse Student Section. This table reports the AAMC official number of graduates from U.S. medical schools and the AMA Masterfile sample of graduates of U.S. medical schools, by medical school graduation year.

Table A.3: Residency Program Summary Statistics for Program-Level Analysis

	1996-2010		1996				
	# Programs-Years	Fraction Female among First Year Residents	# Programs	Fraction Female among First Year Residents	Fraction Female among Full-time Faculty	Paid Maternity Leave Policy	Onsite Childcare
All Programs	31,969	0.43	2,541	0.35	0.21	0.43	0.41
		(0.27)		(0.27)	(0.16)		
By Specialty							
Anesthesiology	1,318	0.32	146	0.29	0.23	0.43	0.45
		(0.17)		(0.22)	(0.11)		
Dermatology	782	0.60	97	0.48	0.29	0.53	0.43
		(0.28)		(0.32)	(0.18)		
Emergency medicine	1,453	0.34	113	0.27	0.21	0.50	0.37
		(0.17)		(0.14)	(0.12)		
Family practice	4,467	0.51	453	0.46	0.26	0.35	0.32
		(0.22)		(0.23)	(0.18)		
General surgery	3,452	0.26	361	0.18	0.09	0.44	0.43
		(0.20)		(0.20)	(0.13)		
Internal medicine	4,178	0.41	407	0.36	0.21	0.54	0.39
		(0.14)		(0.16)	(0.12)		
Neurological surgery	143	0.11	98	0.08	0.05	0.41	0.38
		(0.23)		(0.20)	(0.08)		
Neurology	657	0.43	115	0.32	0.18	0.38	0.52
		(0.25)		(0.27)	(0.11)		
Obstetrics/Gynecology	2,860	0.75	263	0.64	0.27	0.44	0.37
		(0.22)		(0.26)	(0.15)		
Ophthalmology	1,113	0.36	131	0.29	0.17	0.53	0.37
		(0.27)		(0.28)	(0.16)		
Orthopedic surgery	1,600	0.10	156	0.06	0.05	0.38	0.43
		(0.16)		(0.13)	(0.10)		
Otolaryngology	573	0.25	105	0.19	0.10	0.40	0.41
		(0.26)		(0.26)	(0.10)		
Pathology	961	0.51	178	0.40	0.26	0.42	0.41
		(0.25)		(0.28)	(0.14)		
Pediatrics	2,269	0.68	208	0.62	0.37	0.46	0.43
		(0.17)		(0.17)	(0.12)		
Physical medicine/rehab	509	0.39	78	0.33	0.32	0.27	0.43
		(0.24)		(0.26)	(0.13)		
Psychiatry	1,623	0.53	193	0.46	0.27	0.44	0.43
		(0.21)		(0.23)	(0.12)		
Radiation oncology	138	0.31	79	0.22	0.28	0.46	0.51
		(0.29)		(0.32)	(0.17)		
Radiology	1,873	0.28	204	0.26	0.20	0.47	0.43
		(0.21)		(0.24)	(0.11)		
Transitional Year	1,524	0.33	142	0.32	0.21	0.39	0.49
		(0.19)		(0.20)	(0.19)		
Urology	476	0.18	120	0.11	0.04	0.38	0.35
		(0.25)		(0.24)	(0.09)		
N	31,969	31,969	3,647	3,398	3,673	3,417	3,295

Source: GME Census Track. Note: This table reports summary statistics for the sample of residency programs used for the analysis of the effect of the duty hour reform on program gender composition. The data are comprised of an unbalanced panel of residency programs 1996-2010 for 21 broad specialties, omitting 2000. Fellowship programs and small programs (those with one resident in any survey year) are dropped. The table also reports the GME Census Track baseline (1996) specialty characteristics that are used as controls in the main analysis. All programs, regardless of size, are included in the computation.

Table A.4: Effect of the Reform on Specialty Entry: Conditional Logit Coefficients

Dependent Variable: Specialty Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			II: Medium-Run Outcomes, AMA Masterfile				
Panel A: Female							
Average Weekly Hours × Transition	0.046	0.277***	-0.009	-0.152**	-0.011	0.056	-0.109
	(0.073)	(0.075)	(0.073)	(0.074)	(0.073)	(0.072)	(0.059)
Average Weekly Hours × Post	0.522***	0.834***	0.440***	0.307***	0.429***	0.522***	0.307***
	(0.058)	(0.062)	(0.059)	(0.059)	(0.058)	(0.058)	(0.073)
Panel B: Male							
Average Weekly Hours × Transition	-0.079	-0.154***	-0.058	-0.346***	-0.695***	-0.231***	0.156***
	(0.055)	(0.055)	(0.056)	(0.057)	(0.064)	(0.056)	(0.056)
Average Weekly Hours × Post	0.005	-0.133***	0.031	-0.271***	-0.746***	-0.194***	0.389***
	(0.046)	(0.045)	(0.046)	(0.047)	(0.053)	(0.047)	(0.048)
P-value for test of equality of male/female coeff.							
Average Weekly Hours × Transition	0.172	0.000	0.598	0.037	0.000	0.002	0.004
Average Weekly Hours × Post	0.000	0.000	0.000	0.000	0.000	0.000	0.284
			II: Initial Specialty Entry, GME Census Track				
Panel A: Female							
Average Weekly Hours × Transition	0.173**	0.198***	-	-0.052	0.047	0.147**	0.100
	(0.069)	(0.072)		(0.070)	(0.070)	(0.070)	(0.069)
Average Weekly Hours × Post	0.360***	0.379***	-	0.121**	0.177***	0.318***	0.271***
	(0.058)	(0.062)		(0.059)	(0.059)	(0.058)	(0.058)
Panel B: Male							
Average Weekly Hours × Transition	-0.045	-0.177***	-	-0.247***	-0.676***	-0.112**	0.168***
	(0.050)	(0.051)		(0.052)	(0.060)	(0.051)	(0.050)
Average Weekly Hours × Post	-0.258***	-0.480***	-	-0.512***	-1.161***	-0.395***	0.128***
	(0.043)	(0.044)		(0.045)	(0.052)	(0.043)	(0.045)
P-value for test of equality of male/female coeff.							
Average Weekly Hours × Transition	0.011	0.000	-	0.026	0.000	0.003	0.422
Average Weekly Hours × Post	0.000	0.000	-	0.000	0.000	0.000	0.052
Specialty FE	X	X	X	X	X	X	X
Ob/Gyn., Primary Care time trends		X					
Controls for age, medical school rank			X				
Specialty characteristics X cohort FE				X			
Fraction female X cohort FE					X		
Competitiveness X cohort FE						X	
All specialty linear pre-trends							X

Source: AMA Physician Masterfile, GME Census Track, Baldwin Jr et al. (2003). Note: This table reports the results of maximum likelihood estimation of a conditional logit model in which the baseline specification has specialty outcome as the dependent variable and the explanatory variables include specialty fixed effects and specialty pre-policy hours ($\text{Hours}_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). Coefficients are reported, with heteroskedastic robust standard errors in parentheses. Column 1 reports the results of the baseline specification with no additional controls. Columns 2-7 progressively include specialty-specific and demographic controls. The p-values at the bottom of each panel are from a Wald test of the null hypothesis that the male and female coefficients are equal.

Table A.5: Predicted Probability of Entering Specialties, Pre- and Post-Reform

	Women		Men	
	Pre-reform (1993-2002)	Post-reform (2006-2010)	Pre-reform (1993-2002)	Post-reform (2006-2010)
Anesthesiology	0.039	0.040	0.065	0.065
Dermatology	0.027	0.025	0.014	0.014
Emergency medicine	0.052	0.051	0.080	0.080
Family practice	0.144	0.137	0.109	0.109
General surgery	0.043	0.050	0.089	0.089
Internal medicine	0.211	0.212	0.224	0.224
Neurological surgery	0.002	0.003	0.014	0.014
Neurology	0.016	0.016	0.017	0.017
Obstetrics/Gynecology	0.099	0.107	0.026	0.026
Ophthalmology	0.021	0.020	0.029	0.029
Orthopedic surgery	0.008	0.009	0.060	0.060
Otolaryngology	0.008	0.009	0.021	0.021
Pathology	0.024	0.022	0.019	0.019
Pediatrics	0.175	0.176	0.061	0.061
Physical medicine/rehab	0.013	0.012	0.019	0.019
Psychiatry	0.056	0.052	0.037	0.037
Radiation oncology	0.005	0.005	0.010	0.010
Radiology	0.034	0.033	0.074	0.074
Urology	0.006	0.006	0.022	0.022
Internal Medicine-Pediatrics	0.016	0.016	0.011	0.011

Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This table reports the predicted probabilities of entering each specialty, by gender, based on the results of maximum likelihood estimation of the baseline conditional logit model. The predicted probabilities are computed based on the pre-period (1993-2002) and the Post period (2006-2010).

Table A.6: Effect of the Reform on Specialty Choice: OLS

Dependent Variable: ln(Share$_{st}$)*100	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I: Medium-Run Specialty Outcome, AMA Masterfile							
Panel A: Female							
Average Weekly Hours × Transition	0.07	0.08	0.16	0.08	-0.57**	-0.06	0.02
	(0.28)	(0.22)	(0.28)	(0.32)	(0.25)	(0.27)	(0.27)
Average Weekly Hours × Post	0.63**	0.64***	0.72**	0.66*	-0.08	0.51*	0.55
	(0.29)	(0.19)	(0.27)	(0.35)	(0.26)	(0.28)	(0.36)
Panel B: Male							
Average Weekly Hours × Transition	-0.43	-0.34	-0.30	-0.46	-1.17***	-0.49	-0.25
	(0.39)	(0.27)	(0.38)	(0.44)	(0.37)	(0.40)	(0.36)
Average Weekly Hours × Post	-0.32	-0.17	-0.19	-0.32	-1.19***	-0.44	-0.02
	(0.43)	(0.27)	(0.43)	(0.49)	(0.38)	(0.43)	(0.40)
P-value for test of equality of male/female coeff.							
Average Weekly Hours × Transition	0.133	0.176	0.173	0.129	0.102	0.193	0.234
Average Weekly Hours × Post	0.002	0.003	0.004	0.004	0.006	0.003	0.039
N	360	360	360	360	360	360	360
II: Initial Specialty Entry, GME Census Track							
Panel A: Female							
Average Weekly Hours × Transition	0.12	0.12	-	0.17	-0.36*	-0.00	0.30
	(0.25)	(0.19)		(0.29)	(0.20)	(0.22)	(0.27)
Average Weekly Hours × Post	0.63*	0.63**	-	0.69	-0.07	0.43	0.95*
	(0.36)	(0.30)		(0.42)	(0.29)	(0.31)	(0.47)
Panel B: Male							
Average Weekly Hours × Transition	-0.35	-0.30*	-	-0.27	-0.91***	-0.41	-0.11
	(0.27)	(0.17)		(0.30)	(0.19)	(0.30)	(0.27)
Average Weekly Hours × Post	-0.30	-0.21	-	-0.22	-1.10***	-0.45	0.12
	(0.38)	(0.27)		(0.43)	(0.25)	(0.41)	(0.43)
P-value for test of equality of male/female coeff.							
Average Weekly Hours × Transition	0.069	0.098		0.073	0.025	0.103	0.063
Average Weekly Hours × Post	0.002	0.004		0.002	0.001	0.004	0.003
N	294	294		294	294	294	294
Specialty FE	X	X	X	X	X	X	X
Ob/Gyn., Primary Care time trends		X					
Age, Rank			X				
Specialty characteristics X cohort FE				X			
Fraction female X cohort FE					X		
Competitiveness X cohort FE						X	
All specialty linear pre-trends							X

Source: AMA Physician Masterfile, GME Census Track, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-differences OLS log share specification for specialty entry, estimated separately for men and women. The dependent variable is the natural logarithm of the share of women (men) from a medical school graduation cohort in a specialty, multiplied by 100. The explanatory variables include specialty fixed effects, medical school graduation cohort fixed effects, and specialty pre-policy hours (Hours$_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). Panel I uses the AMA Masterfile data, 1993-2010, and specifications have 360 observations stemming from the analysis of 20 specialties over 18 years. Panel II uses the GME Census Track data, 1996-2010 (omitting 2000), and specifications have 294 observations stemming from the analysis of 21 specialties over 14 years. Column 1 reports the results of the baseline specification with no additional controls. Columns 2-7 progressively include specialty-specific and demographic controls. Standard errors clustered at the specialty level are reported in parentheses. The p-values at the bottom of each panel are from a Wald test of the null hypothesis that the male and female coefficients are equal.

Table A.7: The Effect of the Duty Hour Reform on Specialty Entry:
Conditional Logit Average Marginal Effects
Additional Robustness Checks

Dependent Variable: Specialty Outcome	(1) Avg Weekly Hours	(2) % Residents >80 hours	(3) Total Residency Hours	(4) Avg Weekly Hours >80	(5) Omit Cohorts 2001/2
I: Medium-Run Specialty Outcome, AMA Masterfile					
Panel A: Female					
Average Weekly Hours × Transition	0.002	0.001	0.004***	0.164*	0.002
	(0.003)	(0.002)	(0.000)	(0.090)	(0.003)
Average Weekly Hours × Post	0.023***	0.011***	0.007***	0.677***	0.023***
	(0.003)	(0.001)	(0.000)	(0.071)	(0.003)
Panel B: Male					
Average Weekly Hours × Transition	-0.004	0.000	0.003***	0.039	-0.005*
	(0.002)	(0.001)	(0.000)	(0.075)	(0.002)
Average Weekly Hours × Post	0.000	0.001	0.003***	0.186***	-0.001
	(0.002)	(0.001)	(0.000)	(0.062)	(0.003)
P-value for test of equality of male/female coeff.					
Average Weekly Hours × Transition	0.169	0.659	0.000	0.282	0.075
Average Weekly Hours × Post	0.000	0.000	0.000	0.000	0.000
II: Initial Specialty Entry, GME Census Track					
Panel A: Female					
Average Weekly Hours × Transition	0.007**	0.003**	0.002***	0.296***	0.009***
	(0.003)	(0.001)	(0.000)	(0.078)	(0.003)
Average Weekly Hours × Post	0.015***	0.006***	0.004***	0.465***	0.017***
	(0.002)	(0.001)	(0.000)	(0.065)	(0.003)
Panel B: Male					
Average Weekly Hours × Transition	-0.002	0.000	0.000	0.003	-0.006***
	(0.002)	(0.001)	(0.000)	(0.061)	(0.002)
Average Weekly Hours × Post	-0.011***	-0.005***	0.000	-0.148***	-0.015***
	(0.002)	(0.001)	(0.000)	(0.053)	(0.002)
P-value for test of equality of male/female coeff.					
Average Weekly Hours × Transition	0.011	0.072	0.000	0.003	0.000
Average Weekly Hours × Post	0.000	0.000	0.000	0.000	0.000
Specialty FE	X	X	X	X	X

Source: AMA Physician Masterfile, GME Census Track, Baldwin Jr et al. (2003). Note: This table reports the results of maximum likelihood estimation of a conditional logit model in which the baseline specification has specialty outcome as the dependent variable and the explanatory variables include specialty fixed effects and specialty pre-policy hours ($\text{Hours}_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). Average marginal effects associated with the coefficient on the interaction terms are reported. The average marginal effect is the average of individual specialty-specific marginal effects for all individuals in the sample. The results from specifications utilizing three alternative parameterizations of specialty pre-policy hours are reported. % above 80 hours is a measure of the fraction of residents who reported working more than 80 hours per week, reported in Baldwin Jr et al. (2003). Total residency hours is computed by multiplying average pre-policy hours per week by the number of years of residency for that specialty. I do not additionally multiply by the number of weeks in a year, since this is uniform across medical specialties and would only result in a rescaling of the estimated coefficients. Average weekly hours above 80 is an indicator for whether a specialty's pre-policy average weekly hours were in excess of 80. Heteroskedastic robust standard errors are reported in parentheses.

Table A.8: Conditional Logit Average Marginal Effects: Alternative Methods of Statistical Inference

Dependent Variable: Specialty Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			I: Medium-Run Specialty Outcome, AMA Masterfile				
Panel A: Female							
Average Weekly Hours × Transition	0.002	0.012	0.000	-0.007	0.000	0.002	-0.005
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	{0.003}	{0.004}	{0.002}	{0.003}	{0.003}	{0.003}	{0.002}
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Average Weekly Hours × Post	0.023	0.037	0.019	0.014	0.019	0.023	0.014
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	{0.003}	{0.004}	{0.004}	{0.004}	{0.003}	{0.003}	{0.003}
	[0.004]	[0.003]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Panel B: Male							
Average Weekly Hours × Transition	-0.004	-0.007	-0.003	-0.016	-0.031	-0.01	0.007
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
	{0.003}	{0.003}	{0.003}	{0.004}	{0.006}	{0.004}	{0.002}
	[0.003]	[0.003]	[0.004]	[0.003]	[0.004]	[0.003]	[0.003]
Average Weekly Hours × Post	0.000	-0.006	0.001	-0.012	-0.034	-0.009	0.018
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
	{0.003}	{0.003}	{0.002}	{0.004}	{0.006}	{0.003}	{0.003}
	[0.003]	[0.003]	[0.003]	[0.003]	[0.004]	[0.003]	[0.003]
			II: Initial Specialty Entry, GME Census Track				
Panel A: Female							
Average Weekly Hours × Transition	0.007	0.008	-	-0.002	0.002	0.006	0.004
	(0.003)	(0.003)		(0.003)	(0.003)	(0.003)	(0.003)
	{0.002}	{0.003}		{0.002}	{0.002}	{0.002}	{0.002}
Average Weekly Hours × Post	0.015	0.016	-	0.005	0.007	0.013	0.011
	(0.002)	(0.003)		(0.002)	(0.002)	(0.002)	(0.002)
	{0.003}	{0.003}		{0.003}	{0.003}	{0.003}	{0.003}
Panel B: Male							
Average Weekly Hours × Transition	-0.002	-0.007	-	-0.010	-0.029	-0.005	0.007
	(0.002)	(0.002)		(0.002)	(0.003)	(0.002)	(0.002)
	{0.004}	{0.005}		{0.005}	{0.008}	{0.004}	{0.002}
Average Weekly Hours × Post	-0.011	-0.02	-	-0.022	-0.049	-0.017	0.005
	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
	{0.003}	{0.004}		{0.004}	{0.007}	{0.003}	{0.002}
Specialty FE	X	X	X	X	X	X	X
Ob/Gyn., Primary Care time trends		X					
Controls for age, medical school rank			X				
Specialty characteristics X cohort FE				X			
Fraction female X cohort FE					X		
Competitiveness X cohort FE						X	
All specialty linear pre-trends							X

Source: AMA Physician Masterfile, GME Census Track, Baldwin Jr et al. (2003). Note: This table reports the results of maximum likelihood estimation of a conditional logit model in which the baseline specification has specialty outcome as the dependent variable and the explanatory variables include specialty fixed effects and specialty pre-policy hours ($\text{Hours}_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). The average marginal effects associated with the coefficient on the interaction terms are reported. The average marginal effect is the average of individual specialty-specific marginal effects for all individuals in the sample. Heteroskedastic robust standard errors are reported in parentheses. Standard errors clustered at the cohort level are reported in curly braces. Standard errors clustered at the medical school level are reported in square brackets. Medical school is not observed in the GME Census Track data.

Table A.9: The Effect of the Duty Hour Reform on Specialty Entry:
Conditional Logit Average Marginal Effects
Inclusion of Foreign and Osteopathic Medical School Graduates

Dependent Variable: Specialty Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			I: Medium-Run Specialty Outcome, AMA Masterfile				
Panel A: Female							
Average Weekly Hours × Transition	0.009***	0.018***	0.009***	0.002	0.007***	0.009***	0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Average Weekly Hours × Post	0.032***	0.044***	0.032***	0.024***	0.029***	0.032***	0.023***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Panel B: Male							
Average Weekly Hours × Transition	-0.004*	-0.005**	-0.003	-0.013***	-0.021***	-0.007***	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Average Weekly Hours × Post	-0.001	-0.004**	0.000	-0.012***	-0.029***	-0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
P-value for test of equality of male/female coeff.							
Average Weekly Hours × Transition	0.000	0.000	0.000	0.000	0.000	0.000	0.684
Average Weekly Hours × Post	0.000	0.000	0.000	0.000	0.000	0.000	0.000
			II: Initial Specialty Entry, GME Census Track				
Panel A: Female							
Average Weekly Hours × Transition	0.009***	0.010***	-	0.007***	0.008***	0.009***	0.006**
	(0.002)	(0.003)		(0.003)	(0.002)	(0.002)	(0.002)
Average Weekly Hours × Post	0.012***	0.013***	-	0.009***	0.008***	0.011***	0.005**
	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
Panel B: Male							
Average Weekly Hours × Transition	-0.003*	-0.003*	-	-0.005***	-0.013***	-0.004**	-0.002
	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
Average Weekly Hours × Post	-0.012***	-0.012***	-	-0.015***	-0.032***	-0.015***	-0.010***
	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)	(0.002)
P-value for test of equality of male/female coeff.							
Average Weekly Hours × Transition	0.000	0.000		0.000	0.000	0.000	0.013
Average Weekly Hours × Post	0.000	0.000		0.000	0.000	0.000	0.000
Specialty FE	X	X	X	X	X	X	X
Ob/Gyn., Primary Care time trends		X					
Controls for age, medical school rank			X				
Specialty characteristics X cohort FE				X			
Fraction female X cohort FE					X		
Competitiveness X cohort FE						X	
All specialty linear pre-trends							X

Source: AMA Physician Masterfile, GME Census Track, Baldwin Jr et al. (2003). Note: This table reports the results of maximum likelihood estimation of a conditional logit model in which the baseline specification has specialty outcome as the dependent variable and the explanatory variables include specialty fixed effects and specialty pre-policy hours ($\text{Hours}_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). The average marginal effects associated with the coefficient on the interaction terms are reported. The average marginal effect is the average of individual specialty-specific marginal effects for all individuals in the sample. The sample includes foreign and osteopathic medical school graduates. Heteroskedastic robust standard errors are reported in parentheses.

Table A.10: The Effect of the Duty Hour Reform on Specialty Entry: Nested Logit

Dependent Variable: ln($Share_{st}$)*100	(1) Berry Logit	(2) Nest 1	(3) Nest 2	(4) Nest 3
Panel A: Female				
Average Weekly Hours × Transition	0.19	0.58***	0.35***	0.25**
	(0.21)	(0.06)	(0.08)	(0.11)
Average Weekly Hours × Post	0.79***	0.89***	0.59***	0.49***
	(0.18)	(0.05)	(0.07)	(0.09)
Panel B: Male				
Average Weekly Hours × Transition	-0.33	0.21***	-0.03	-0.04
	(0.21)	(0.02)	(0.08)	(0.09)
Average Weekly Hours × Post	-0.24	0.27***	0.01	-0.00
	(0.17)	(0.02)	(0.07)	(0.08)
P-value for test of equality of male/female coeff.				
Average Weekly Hours × Transition	0.078	0.000	0.001	0.040
Average Weekly Hours × Post	0.000	0.000	0.000	0.000
N	342	342	342	342

Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This table reports the results of Berry logit version of a nested logit model in which the baseline specification has specialty outcome as the dependent variable and the explanatory variables include specialty fixed effects, medical school cohort fixed effects, and specialty pre-policy hours ($Hours_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). All variables have been normalized by a specialty outside option, Pathology, which is then dropped from the sample. Column 1 reports the results of the baseline Berry logit specification. In order to allow more flexible substitution patterns across specialties, columns 2-4 implement the Berry logit version of nested logit, which entails controlling for the natural logarithm of a specialty's share in its designated nest. In column 2, the nests are (1) surgical and (2) non-surgical specialties. In column 3, the nests are (1) surgical, (2) primary care, and (3) the rest. In column 4, the nests are (1) surgical, (2) primary care, (3) E-ROAD, and (4) the rest. The coefficients on the interaction terms are reported, with standard errors clustered at the specialty level in parentheses.

Table A.11: The Effect of the Duty Hour Reform on Residency Program Size

Dependent Variable: $ProgramSize_{pst}$	(1)	(2)	(3)
Panel A: Program-level hours			
Average Weekly Hours × Transition	-0.002	-0.003	-0.002
	(0.002)	(0.002)	(0.002)
Average Weekly Hours × Post	-0.006**	-0.007**	-0.005*
	(0.003)	(0.003)	(0.003)
N	45,833	45,833	45,833
Panel B: Specialty-level hours			
Average Weekly Hours × Transition	-0.002	-0.001	-0.003
	(0.006)	(0.005)	(0.004)
Average Weekly Hours × Post	-0.010	-0.008	-0.011
	(0.012)	(0.011)	(0.009)
N	45,833	45,833	45,833
Ob/Gyn. and Primary Care specialty time trends		X	
All specialty linear pre-trends			X

Source: AAMC GME Census Track, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-differences specification estimating the effect of the reform on residency program size. The dependent variable is the number of first year residents in a residency program in a given calendar year. In Panel A, the explanatory variables include program fixed effects, calendar year fixed effects and program pre-policy hours ($Hours_{sp,1996}$) interacted with an indicator for entering residency in 2003-2005 (Transition) and 2006-2010 (Post). The year 2000 is excluded from the analysis due to a low response rate to the survey in that year. In Panel B, programs are assigned the pre-policy hours associated with their specialty from the Baldwin Jr et al. (2003) survey. Column 1 reports the results of the baseline specification with no additional controls. Columns 2-3 include specialty- and program-specific controls. The sample imposes no size restrictions on residency programs. Standard errors, reported in parentheses, are clustered at the program level in Panel A and specialty level in Panel B.

Table A.12: Characterizing the Quality of the Marginal Entrant

Dependent variable: Indicator for whether an individual attended a ranked medical school*100	(1)	(2)	(3)	(4)	(5)	(6)
				A. All		
Average Weekly Hours × Transition	-0.107***	-0.114***	-0.113***	-0.122***	-0.116***	-0.085**
	(0.025)	(0.024)	(0.027)	(0.026)	(0.029)	(0.033)
Average Weekly Hours × Post	-0.048*	-0.058**	-0.061**	-0.056*	-0.056*	-0.012
	(0.026)	(0.024)	(0.025)	(0.029)	(0.030)	(0.035)
Mean Dependent Variable	48.06					
				B. Men		
Average Weekly Hours × Transition	-0.137***	-0.145***	-0.145***	-0.146***	-0.153***	-0.114***
	(0.027)	(0.027)	(0.029)	(0.029)	(0.033)	(0.033)
Average Weekly Hours × Post	-0.039	-0.050*	-0.054*	-0.032	-0.049	-0.000
	(0.025)	(0.026)	(0.026)	(0.030)	(0.032)	(0.037)
Mean Dependent Variable	47.86					
				C. Women		
Average Weekly Hours × Transition	-0.084**	-0.076**	-0.087***	-0.090***	-0.082**	-0.069
	(0.030)	(0.028)	(0.027)	(0.024)	(0.031)	(0.040)
Average Weekly Hours × Post	-0.095**	-0.084**	-0.102***	-0.097***	-0.093**	-0.069
	(0.035)	(0.033)	(0.027)	(0.033)	(0.036)	(0.047)
Mean Dependent Variable	48.30					
Ob/Gyn. and Primary Care specialty time trends		X				
Specialty characteristics X cohort FE			X			
Fraction female X cohort FE				X		
Competitiveness X cohort FE					X	
All specialty linear pre-trends						X

Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-differences specification, estimated separately for men and women, on the sample of U.S. medical school graduates. The dependent variable is an indicator for whether an individual in a given specialty from a given medical school cohort attended a ranked medical school, multiplied by 100. The explanatory variables include specialty fixed effects, medical school graduation cohort fixed effects and specialty pre-policy hours ($\text{Hours}_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). Standard errors clustered at the specialty level are reported in parentheses.