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

Melanie Wasserman*

February 2017

Abstract

Are the long, inflexible work hours required by many high-paying professions a barrier to entry for women? I explore this question by studying the introduction of a policy in 2003 that capped the average workweek for medical residents at 80 hours. I leverage the fact that the reform was differentially binding for medical specialties due to pre-existing differences in medical specialties' average hours per week. Using administrative data on the universe of U.S. medical school graduates from 1993 to 2010, I find that when a medical specialty reduces its weekly hours, more women enter the specialty, whereas there is little change in men's entry. I present evidence that changes in medical school students' preferences for specialties – rather than residency programs' preferences for students – are the primary driver of the increase in women's entry. A back of the envelope calculation suggests that women's entry into historically time-intensive and highly compensated specialties due to a modest reduction in weekly hours would close the physician gender pay gap by at least 15 percent.

Keywords: occupational choice, non-monetary occupational attributes, long hours, gender

JEL codes: J16, J24, J44

*UCLA Anderson School of Management and University of Michigan Population Studies Center. Address: 426 Thompson St Room 2046, Ann Arbor, MI 48104. 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, Claudia Goldin, Jon Gruber, Sally Hudson, Erin Johnson, Brendan Price, Adam Sacarny, Ashish Shenoy, Nils Wernerfelt, and helpful comments from seminar participants at MIT Labor and Summer Applied Micro Lunches, UCLA Anderson GEM, LSE, Dartmouth, SUNY Binghamton, UIUC, University of Michigan, and ACLEC. 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, 2016). 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 professional 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, 2016a). As of yet, there is little evidence 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., 2015). 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 identify 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.

In order to address these empirical hurdles, this paper leverages a unique 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-1970's, 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 I 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 annual earnings for women 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 a job's time requirements on individuals' propensity to enter a job.

My 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 five 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

[1]There is a positive correlation (correlation coefficient of 0.81) between hours worked during residency and professional practice, based on data reported in Iserson (2006). Fully trained physicians, however, have substantially more discretion over their hours worked through choice of practice setting (solo, group practice, hospital, academic) and volume of patients.

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

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 various methods of statistical inference.

It is possible the reform also induced a labor demand response, that is, medical residency programs shifted their preferences or hiring practices in response to the reduction in hours. To shed light on whether the effects of the reform on female entry are generated by changes in medical residents' preferences for specialties (labor supply) or residency programs' preferences for female applicants (labor demand), I analyze survey data on the stated preferences of U.S. medical school students upon matriculation in medical school. The results reveal that female medical school matriculants shift their preferences for time-intensive specialties in response to the reform. The point estimates, while imprecisely estimated, are slightly larger than those from the specialty entry analysis but similarly differentiated by gender. This evidence supports the interpretation that reform-induced changes in medical residents' preferences for time-intensive specialties are the driving force behind the increased entry of women.

I discuss two theoretical mechanisms that could generate women's differential response to the reform. First, residency work hours could disproportionately affect women's specialty choices due to the anticipated or actual time demands of family formation. Medical residency coincides with prime childbearing years and women tend to experience a steeper trade-off between market and non-market time when they have children. Second, women could have a distaste for working in a male-dominated environment. If a decline in a specialty's hours spurs the entry of women due to expected fertility choices, it could lead to a further increase in entry from those women who prefer to work in settings with greater female representation. By examining heterogeneity in the female entry response to the reform, there is suggestive evidence that both mechanisms are operative, although I cannot rule out other hypotheses that could give rise to the specialty entry results.

As a final exercise, I use the estimated effects of the reform on specialty entry to assess the implications of reducing early career time requirements for the physician gender pay gap. Since cohorts of physicians affected by this policy are finishing residency training at present, I extrapolate the effect of the reform on earnings using the estimated changes in the allocation of women across specialties along with specialty-specific earnings. A back of the envelope calculation suggests that the entry of women into historically time-intensive and highly compensated specialties due to the reform will narrow the physician gender earnings gap upon labor market entry by at least 15 percent. Overall, it appears that a modest reduction in time requirements during early career years could have important implications for gender pay differentials.

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, 2016a). 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, 2016; Mas and Pallais, 2016). 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 presents the empirical framework for examining the effect of hours requirements on specialty choice and presents the main results. Section 5 discusses theoretical mechanisms that could account for the differential responsiveness of men and women to the reform and provides suggestive empirical evidence on which mechanisms are operative. In Section 6, I characterize the implications of the reallocation of women among medical specialties due to the reform for the physician gender wage gap. Section 7 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 USMLE occur early in medical school, students often plan years in advance in order to emerge a competitive applicant, particularly for oversubscribed specialties.

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 wage 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 – wield 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

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

[5]Agarwal (2015) estimates preferences for non-pecuniary attributes of residency programs within one specialty, Family Medicine, and finds that 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 are deterministic of specialty choice. As long as these attributes are stable over time, the empirical strategy in this paper will account for this variation.

[6]In his 2015 book "Let Me Heal", Ludmerer describes pre-WWII medical residency with the following passage: "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,

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 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).

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 the monitoring data collected by the ACGME, 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

which customarily started Saturdays at noon and continued through the following Monday at 8 a.m" (Ludmerer 2015, p. 104).

[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 both the House of Representatives and the Senate to regulate resident hours, and the Occupational Safety and Health Administration (OSHA) considered petitions along similar lines in 2001.

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 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.

Figure II Panel A uses individual reports of hours last 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.[8] 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 51 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 II 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.[9]

While the duty hours 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

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

[9]Appendix Figure A.1 plots the trends in hours worked separately for men and women. Before the policy was enacted, female residents worked, on average, six fewer hours per week than male residents, and were three percentage points less likely to work more than 80 hours per week, likely an artifact of gender differences in specialty choice, with women tending to cluster in less time-intensive specialties. Appendix Figure A.1 Panel A demonstrates the reduction in average hours after the reform is concentrated primarily among men. Men experience a reduction of approximately six hours per week, while women experience a more modest reduction of three hours per week. The fraction of women and men who work more than 80 hours per week declines to approximately the same level after the reform is enacted (see Appendix Figure A.1 Panel B).

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 III 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 :

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

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, derived 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.[10]

3 Data

3.1 Data Sources

In order to examine how physician specialty entry responds to the introduction of the ACGME duty hour reform, I utilize the American Medical Association (AMA) Physician Masterfile (henceforth "Full Masterfile"), which covers the universe of physicians in the United States. The Full 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 residency training institution) and pri-

[10] In addition to reducing the level of weekly hours, 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.

mary specialty. An individual's inclusion in the Full Masterfile is triggered by entry into a U.S. medical school or U.S. medical residency program. In addition to the training information provided in the Full Masterfile, for the subset of physicians who did any part of their residency training in California or Texas, I have detailed information on their residency training history, including the start/end date, and specialty (henceforth "CA/TX Masterfile"). The information on residency training in the CA/TX Masterfile comes from the American Association of Medical Colleges' (AAMC) National Graduate Medical Education Census, which collects data from residency program directors of programs accredited by the ACGME. I use the supplementary residency information to validate the less detailed training information in the Full Masterfile.

I classify specialties based on their pre-policy time intensity during residency training with use of the previously discussed 1999 nationally representative survey of second year medical residents, conducted by Baldwin Jr et al. (2003). Since information on hours is reported for 20 specialty categories,[11] I crosswalk the more detailed specialty information in the AMA Masterfile to the coarser categories, using a classification scheme provided by the Dartmouth Atlas.[12] 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.

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 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.

3.2 Sample Restrictions and Summary Statistics

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 duty hour 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 the exposure to the duty hour reform. In the Full 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

[11] 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 exclude Preventive Medicine from the analysis since the survey sample size is fewer than five individuals and Preventive Medicine residency programs accept no more than 10 individuals each year.

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

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.[13] 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 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 281,477 U.S. medical school graduates.[14]

Table I presents summary statistics for the Masterfile samples, with column 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 the specialty entry analysis. In each case, the sample is almost half female, with an average age at medical school graduation of approximately 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. 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.

4 The Effect of the Duty Hour Reform on Specialty Choice

4.1 Evidence on Specialty Entry

In order to identify the effect of a specialty's hours requirements on specialty entry, 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 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 III, average hours declined across all specialties with pre-policy hours near or above 80, with the steepest reductions among the specialties

[13] Using the more detailed information on residency training in CA/TX Masterfile, Appendix Figure A.2 plots the distribution of the gap between medical school graduate year and residency start year for U.S. medical school graduates and foreign medical school graduates, respectively.

[14] In Appendix Table A.1, I validate the sample for the specialty entry analysis with the official American Association of Medical Colleges' official data on U.S. medical school graduates. The sample restrictions I impose reduce the sample of U.S. medical school graduates by at most three percent, with no apparent trend over the analysis time period.

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.

I present initial evidence of the effect of the reform on male and female specialty entry by estimating the following event study specification:

$$\ln sh_{st} = \beta_0 + \sum_{k=1993}^{2010} \beta_k (\text{Hours}_{s,1999} \times \mathbb{1}\{t = k\}) + \alpha_s + \gamma_t + \epsilon_{st} \tag{2}$$

where the dependent variable is the natural logarithm of the share of individuals from medical school cohort t working in specialty s.[15] I use the natural logarithm of the share rather than the linear share due to the right-skewed distribution of specialty shares, depicted in Appendix Figure A.3. The independent variables are specialty fixed effects α_s, which control for time-invariant characteristics of specialties, and medical school cohort fixed effects γ_t, which control for overall trends in specialty entry. $\text{Hours}_{s,1999}$ represents a specialty's pre-policy hours worked, as measured by the 1999 survey of medical residents. The regression is estimated separately for men and women.

The coefficients of interest are β_k, the interactions of pre-policy specialty hours worked and the medical school cohort fixed effects ($\text{Hours}_{s,1999} \times \mathbb{1}\{t = k\}$). A positive β_k indicates that individuals are more likely to choose high hours relative to low hours specialties, relative to the reference year 2002. The event study framework permits visual inspection of pre-existing trends in entry into high versus low hours specialties as well as whether there is a mean shift or trend break after the duty hour reform goes into effect in 2003.

Figure IV provides an illustration of female and male specialty entry by plotting the β_k coefficients from regressions estimated separately for men and women. Starting with the estimates from the female sample in Panel A, in the years before the duty hour reform, there is little change in the entry of women into more time-intensive specialties relative to less time-intensive specialties. Almost immediately after the reform was introduced in 2003, however, there is an upward trend in the coefficients, which is indicative of a shift toward more time-intensive specialties. For men, there is a slight negative pre-trend in specialty entry in the years preceding the duty hour 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 change in men's propensity to enter more time-intensive specialties.

[15]This specification yields the same estimates as Berry logit, a discrete choice approach used in industrial organization, where the dependent variable is the log share in a specialty, normalized by the log share in an outside option specialty (Berry, 1994). The results are also similar if I estimate a conditional logit model using maximum likelihood.

As observed in Figure IV Panel A, the effect of the duty hour reform on female specialty entry appears to increase over time. There are three potential 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. As documented in Figure II, 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. 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.

The event study analysis provides graphical evidence that the reform shifted women into more time-intensive specialties, but has low statistical power due to the estimation of year-by-year coefficients. I therefore pool the years pre- and post-reform in my main regression specification. Motivated by the patterns in the event study analysis and following other studies of this reform, 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). I estimate the following specification:

$$\ln sh_{st} = \beta_0 + \beta_1(\text{Hours}_{s,1999} \times \text{Transition}_t) + \beta_2(\text{Hours}_{s,1999} \times \text{Post}_t) + \alpha_s + \gamma_t + \epsilon_{st} \qquad (3)$$

where Transition_t is an indicator for medical school cohorts 2003-2005, and Post_t is an indicator for medical school cohorts 2006-2010, and all other variables are as defined above in equation (2). The coefficients β_1 on the interaction term ($\text{Hours}_{s,1999} \times \text{Transition}_t$) and β_2 on the interaction term ($\text{Hours}_{s,1999} \times \text{Post}_t$) capture the effect of the reform on the log share working in a specialty for medical school cohorts 2003-2005 and 2006-2010, respectively. A positive β_1 or β_2 indicates that individuals are more likely to choose time-intensive specialties after the duty hour reform is introduced.

Table II reports the effects of the reform on specialty entry, separately for men and women. Starting with Table II Panel A column 1, the results from the baseline model, the effect of the reform during the transition period for women is positive but small in magnitude and statistically insignificant. For the period post-transition, the coefficient for women is positive and statistically significant, indicating that women are more likely to choose more time-intensive medical specialties relative to less time-intensive specialties after the reform goes into effect. For each additional hour pre-policy, there is a 0.67 percent increase in the share of women that enter a specialty in the post-transition period from 2006 to 2010. Turning to Panel B, the coefficients for men reveal there is little change in the specialty entry of men due to the duty hour reform; if anything, there is a negative response, which could be a direct consequence of the increased entry of women displacing men.[16]

I account for potential serial correlation in specialty entry by clustering standard errors at the specialty level (Bertrand et al., 2004). Since estimation relies on 20 specialties, which is below the suggested number for reliable statistical inference using standard errors clustered at the specialty level, I also compute the p-values associated with wild cluster bootstrapped t-statistics and permutation tests that non-parametrically approximate the distribution of treatment effects (Cameron et al., 2008).[17] Appendix Table A.2 reports the results of alternative methods of statistical inference. The p-values from wild cluster bootstrapped t-statistics and randomization inference are generally larger than the p-values from standard errors clustered at the specialty level, but the results remain statistically significant in the majority of specifications.[18]

To provide an estimate of the effect of an hours reduction on specialty entry, I 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).[19] 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. 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 entry estimated in Table II column 1 of 0.67 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 hours during residency causes a 5 to 16 percent increase in the share of women who enter the specialty.

4.2 Robustness

The results are robust to the inclusion of time-varying specialty controls, various parameterizations of pre-policy hours requirements during residency, and the expansion of the sample to include foreign and osteo-

[16] Due to the limited growth in residency slots over this time period and the fact that pre-reform some slots in time-intensive specialties went unfilled, there isn't a mechanical decline in male entry when there is an increase in female entry.

[17] I take the specialty share profile from 1993 to 2010 for each of the 20 specialties and randomly assign, without replacement, a value of average hours per week from the observed set. With the new data, I re-estimate the specifications from Table II. I repeat this procedure 999 times and compute p-values based on the distribution of coefficients. Bootstrapped standard errors additionally address the potential statistical inference issues associated with the fact that the pre-policy average hours per week measure used to classify the historical time intensity of specialties is a generated regressor.

[18] I have also implemented two-way clustering of standard errors (at the specialty and cohort levels) to account for the fact that increased entry into one specialty implies decreased entry into another specialty. The standard errors are nearly identical to those from one-way clustering.

[19] I cannot estimate two stage least squares since I only have data on hours worked after the reform for seven specialties.

pathic medical school graduates. I discuss each of these in turn below.

Time-varying specialty controls: Identification of the causal effect of the reform on specialty entry hinges on the assumption that absent the duty hour reform, the share of individuals entering more and less time-intensive specialties would have followed the same relative paths over time. Certain specialties and specialty groups exhibit strong trends prior to the duty hour reform. For example, as depicted in Figure I, Obstetrics/Gynecology (Ob/Gyn) is an outlier specialty in that it is highly time-intensive yet women have entered en masse over the last thirty years. There has also been declining interest in the primary care specialties: Family Practice, Internal Medicine and Pediatrics. In Table II columns 2 and 3, I include controls for a linear time trend for Ob/Gyn and primary care specialties, respectively. With the inclusion of the time trends, both the female and male coefficients change very little. In order to gauge whether compositional changes in the male or female medical school graduate population are potentially driving these results, in column 4 I include the demographic controls, age at medical school graduation, and quality of medical school attended. The main results are qualitatively the same with the inclusion of these individual controls. Column 5 incorporates the Ob/Gyn, primary care, and demographic controls, with again a negligible change in the estimated effects. Finally, in column 6, I include all specialty-specific linear time trends. As expected, this specification has low statistical power since the specialty-specific linear trends absorb a substantial portion of the identifying variation. Even in this restrictive specification, however, there is a positive (and imprecisely estimated) effect of the duty hour reform on women's entry into more time-intensive specialties and the female coefficient is larger than the male coefficient.

Alternative parameterizations of pre-policy hours: I re-estimate the main specification with alternative parameterizations of the pre-reform time intensity of specialties, the results of which are presented in Appendix Table A.3. The first parameterization classifies specialties based on their total time investment. I compute the pre-policy total hours worked during residency for each specialty by multiplying the average pre-policy hours per week by the number of years of training each specialty requires. The second parameterization is a binary classification of specialties based on whether their pre-policy average weekly hours exceeded 80. As reported in Table A.3, both parameterizations yield positive and statistically significant results for women in the post period and are substantially larger in magnitude and statistically distinguishable from the coefficients for men. Appendix Figures A.4 and A.5 confirm that the qualitative patterns in the event study analysis remain similar across various parameterizations of pre-policy hours.

Inclusion of foreign and osteopathic medical school graduates: I repeat the main analysis on the sample that includes foreign and osteopathic medical school graduates. The results are reported in Appendix Table A.4. The estimated coefficients are slightly larger than those from the USMG sample but the qualitative conclusions remain the same. I hypothesize that the larger estimates stem from the fact that I observe foreign medical school students in the data only once they start residency training in the U.S. Given that over 80 percent of foreign medical school graduates do not proceed straight from medical school to residency training in the U.S., those with shorter gaps are disproportionately represented among the more recent cohorts of medical school graduates. The increase in the estimated effect of the reform may arise from the fact that foreign graduates who have shorter gaps are also more likely pursue time-intensive specialties.

4.3 Evidence on Stated Preferences for Specialties

The presence of capacity constraints for residency slots potentially alters the interpretation of the above analysis. It is possible that men's and women's specialty preferences responded similarly to the introduction of the reform, but the rationing of residency slots produced the disparate specialty outcomes by gender. Due to capacity constraints for residency positions that vary by specialty, increased interest in a medical specialty does not necessarily translate to increased participation. For example, in 2004, there were 1,230 US medical school graduate applicants and 2,004 total applicants for 1,044 General Surgery residency positions.[20] Over the 2004 to 2010 period, the number of positions in General Surgery residency programs barely rose from 1,044 to 1,077. Thus, even limited increased interest in General Surgery is unlikely to be accommodated by growth in available residency positions. Instead, greater interest in General Surgery after the duty hour reform would result in stiffer competition for residency slots.[21]

There are two stages at which the presence of capacity constraints could result in a disconnect between specialty preferences and specialty outcomes. First, 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 duty hour reform but cannot due to capacity constraints, the resulting specialty entry changes will be downward biased relative to the unconstrained case.[22] Second,

[20]Since individuals are permitted to rank multiple specialties through NRMP, perhaps a better measure is the number of applicants who ranked General Surgery first. In 2004, there were 1,726 total applicants who ranked General Surgery as their most preferred choice, still far exceeding the 1,044 available positions.

[21]Using NRMP data on the number of residency slots for each specialty during 1993 to 2010, I find that the duty hour reform does not affect the provision of residency slots. Results are available upon request.

[22]Nicholson (2002) investigates how medical students' elasticity of specialty choice with respect to specialty income is biased by omitting consideration of specialty rationing. He shows that by taking into consideration a medical student's subjective probability of being accepted to a specialty in their most preferred specialty in the NRMP, the income elasticity estimates are substantially smaller than under a scenario in which rationing is not considered. The presence of specialty rationing among the highest hours specialties may introduce downward bias in the effect of a reduction in hours on specialty choice for similar reasons.

conditional on individuals submitting applications to programs of a given specialty, if residency programs in more time-intensive specialties shift their hiring practices to prefer female to male applicants post-reform, then an equivalent increase in the applications of men and women could result in a greater share of the new female applicants obtaining positions, and thus a greater entry response among women.

I assess whether the duty hour reform differentially increased interest in time-intensive specialties among women by using data from the Association of American Medical Colleges (AAMC) Matriculating Student Questionnaire (MSQ) on the specialty category students are considering upon enrollment in medical school, for students entering in years 1998-2006 and 2009-2010.[23] 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 equation (3), where the dependent variable is now the 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.

The results in Table III 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 large, positive, and statistically significant effect of the reform on women's stated preferences for more time-intensive specialties. Analogous to the specialty entry results, the female coefficients are consistently larger than the male coefficients, although due to large standard errors I cannot reject the null hypothesis that the effects for women and men are the same. 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. Notably, the male and female coefficients in Table III are larger than the main estimates of the effect of the reform on specialty entry in Table II, which is consistent with the notion that post-reform, increased interest – for both women and men – could not be fully accommodated by available residency slots.

5 Mechanisms for Differential Responsiveness by Gender

If men and women had the same preferences over hours worked during their early careers, then the reduction of hours in certain medical specialties should theoretically make those specialties more attractive for all young physicians. My results, however, demonstrate substantial heterogeneity in the response of male and female physicians to the 2003 ACGME duty hour reform. What accounts for women's greater responsiveness? I discuss and provide suggestive evidence of two mechanisms that could contribute to the greater

[23] To test whether men and women were differentially responsive to the duty hour reform in their residency program application behavior, the ideal data would contain the most preferred specialty listed in the NRMP rankings, by gender. I have not been able to obtain these data.

responsiveness of women to the reform: (1) women differentially value hours requirements due to expected childbearing during residency and (2) women prefer workplaces with higher female representation.

Expected fertility during residency: Differential demands on women's time once they have children combined with the limited flexibility of the residency training period position women to be more responsive to specialties' residency hours requirements than men. Although residency comprises a small minority of physicians' expected working years, it tends to coincide with one's late 20's and early 30's, which are prime childbearing years for women. During residency individuals have access to limited or costly means to adjust their labor supply and the timing of their career investments relative to their fertility decisions.[24] Due to biological constraints, it could be costly in terms of fecundity for women to delay having children until after residency training is complete, particularly if they choose a specialty with a lengthy residency or if they choose to sub-specialize. When women have children, they may face a steeper tradeoff between market and non-market time (relative to men with children and individuals without children). This steeper tradeoff could arise from an increase in the productivity of home production, taste-based preferences for spending more time at home, and social norms that induce women to spend more time on parental duties than men.[25]

Descriptive evidence supports the idea that women are more constrained than men in their fertility choices during residency and these constraints could have been relaxed by the duty hour reform.[26] Even though residency occurs during women's main reproductive years, a lower fraction of female than male physicians have children during residency. Moreover, women appear to time their childbearing relative to their residency training, with the years immediately after residency the most common time for female physicians to have children (Hamilton et al., 2012; Turner et al., 2012). In qualitative surveys before the reform, female surgery residents reported reluctance to have children during residency and perceived that men do not grapple with these concerns to a similar extent (Kellogg, 2011, p. 80). After the reform, a female surgery resident commented that "[O]n a personal level, the 80-hour workweek for me opened up surgery as an option...as a female and being a little bit older, I didn't come straight out of college into med school, I'm at a point in my life where I would have never even considered a specialty where I was here 120 hours a week" (Brooks

[24] Part-time residency positions remain rare, limiting the capacity to adjust one's labor supply on the intensive margin. Furthermore, taking time off requires special accommodation. In order 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. In contrast, fully-trained physicians wield substantially more discretion over their work hours through choice of practice setting and number of patients/procedures.

[25] Among young academic physicians, women engage in 8.5 more hours per week of domestic duties than their male counterparts and are more likely to take time off of work due to childcare disruptions (Jolly et al., 2014). It is also possible women have a greater preference for leisure, regardless of marital and family status.

[26] *The Ultimate Guide to Choosing a Medical Specialty*, published in 2007, gives the following advice to women: "So if you are planning to have kids (or already have a family), keep in mind whether or not your dream specialty will permit time for them. Take a closer look at whether physicians in your chosen specialty might penalize female physicians for maternity leaves or even actively discourage their pregnancies" (Freeman, 2007, p. 61).

and Bosk, 2012).

Did the reform relax an hours constraint on female fertility choices during residency? I explore this question by examining whether the effects of the reform on specialty entry are stronger among women with a preference for earlier childbearing. Drawing on the literature on the cultural determinants of fertility, I proxy an individual's preference for fertility timing by assigning the average female age at first birth in her state of birth from the 1990 National Vital Statistics Report (Fernandez and Fogli, 2009). Then I repeat the main analysis on the two subsamples of individuals with age at first birth above and below the median, respectively. Table IV columns 1 and 2 report the results. The coefficients on the interaction term ($\text{Hours}_{s,1999} \times \text{Post}_t$) reveal that women from both groups respond positively to the reform in their specialty entry, while there is no effect for any male subgroup. Women from states with early childbearing are the most responsive, indicating that the reduction in hours potentially reconciled their competing desires for work and family.

Preferences over gender composition of the workplace: A complementary hypothesis relates to preferences over the gender composition of the workplace. Suppose that there are two types of women: those who incur additional disutility from working long hours when they have children and those who have a distaste for working in a environment with very few other women. The distaste for working in a male-dominated environment implies a model with multiple equilbiria, where one equilibrium represents high hours and low female representation, and the other represents low hours and high female representation. As depicted in Figure I Panel A, prior to the duty hour reform, the most time-intensive specialties were often those with the lowest female representation. The duty hour reform provides the necessary catalyst to attract women whose desired fertility is now compatible with the new work hours. With the knowledge that more women are willing to enter time-intensive specialties, women with a distaste for working in an environment with a dearth of other women will also be willing to enter, spurring a move from the low to the high female representation equilibrium.

My empirical results are consistent with this mechanism. The event study analysis suggests the effect of the reform on female entry into time-intensive specialties increases over time, possibly due to a virtuous cycle of increases in female representation begetting further increases. I also explore whether the effects of the reform are stronger among women who attend medical schools with a history of low female representation in time-intensive specialties. The results of this analysis are presented in Table IV columns 3 through 6. The women who are most responsive to the reform are those who attend medical schools where, prior to the introduction of the reform, a low share of women chose high hours (>80) specialties and a low fraction of those who chose high hours specialties were female. While this evidence is not dispositive, it suggests that women from low representation environments are more likely to respond to the reduction in hours, potentially due to the knowledge of increased female representation in their future work environments.

6 Specialty Entry and the Physician Gender Pay 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. In this section, I assess the implications of the entry of women into high hours, high compensation specialties due to the duty hour reform for physician gender wage gap. There is evidence that the physician gender wage gap narrowed throughout the 1990s, widened during the early 2000s, and stagnated in the late 2000s (Ly et al., 2016; Seabury et al., 2013; Lo Sasso et al., 2011; Modestino, 2012). It is still too soon after the reform to be able to measure the earnings of recent cohorts of physicians during professional practice (those who started residency in 2010 are just entering the labor market at the time of writing). Instead, I use the estimates of the effect of the reform on specialty entry and the average pay associated with each specialty in order to quantify the contribution of a reduction in hours during residency to the physician gender pay gap.

For this exercise, I use specialty- and gender-specific average annual earnings upon labor market entry from the New York State Survey of Residents Completing Training 1999-2008, as reported in Lo Sasso et al. (2011). In addition to capturing differences in pay across specialties, this metric incorporates within specialty differences in remuneration for men and women due to choice of subspecialty, hours, practice setting, etc. I simulate the gender pay gap before the introduction of the policy by computing the weighted average of specialty- and gender-specific pay, w_{sg}, where the weights are the shares of men and women in each specialty pre-reform, $sh_{sg} = \frac{n_{sg}}{N_g}$: $\bar{w}_g = \sum_s w_{sg} sh_{sg}$. According to this calculation, the pre-reform gender earnings gap is \$28,924 (\$185,703 for men and \$156,779 for women, all in 2008 dollars) or 16 percent, which is quite similar to the average salaries for men and women reported in Lo Sasso et al. (2011).[27] Consistent with previous studies that document a substantial role for specialty choice in the physician gender pay gap, if women maintained their specialty-specific salaries but had the same pre-reform specialty distribution as men, their average earnings would be \$174,021, closing the gender pay gap by 60 percent (Sasser, 2005).

Quantifying the contribution of the reform to average female (male) earnings entails a sum of specialty-specific female (male) share changes due to the reform, weighted by specialty earnings: $\Delta\bar{w}_g = \sum_s w_{sg} \Delta sh_{sg}$.[28]

[27] Disparities in annual earnings across specialties are only partly accounted for by differences in hours worked. There is a positive relationship between specialty hourly wages and residency work hours, indicative of convex returns to residency work hours (Leigh et al., 2010).

[28] Specialty-specific share changes due to the reform are the difference between the predicted specialty shares before and after the reform, with specialty shares in each period normalized to add up to one. I assume that specialty-specific pay is time invariant.

My back of the envelope calculation suggests that, through specialty selection, the reform will increase women's average annual earnings by $624. The change in male average earnings implied by this reform is zero, since men are unresponsive in their specialty entry decisions. Thus, the rearrangement of women among medical specialties due to the duty hour reform will close the physician gender earnings gap by two percent. If I scale this estimate by the lower and upper bounds on the first stage relationship, a reduction of four hours per week among the most time-intensive specialties relative to the least time-intensive specialties would reduce the physician gender pay gap by $4,540 to $14,680, or 16 to 51 percent.[29]

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 (2016) 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 (2016b) 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 (2016) 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 a modest reduction in early career time requirements could be instrumental in narrowing the gender pay gap.

7 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 that 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 work hours induces more women to enter, but has little effect on men's entry. Furthermore, the entry of women appears to be due to changes in physician preferences for specialties, not a shift in residency program preferences for

[29] 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.

hiring women. I estimate that the entry of women into historically time-intensive and high paying specialties due to a modest four hour per week reduction could close the physician gender pay gap by at least 15 percent.

As the training and ramp-up periods for professional occupations tend to coincide with women's prime childbearing years, it has been widely postulated that women's underrepresentation in highly compensated tracks can be partially attributed to anticipated or actual conflicting demands of work and family. By exploring heterogeneity in the response of women to the duty hour reform, I find suggestive evidence for the hypothesis that women are more responsive to a reduction in hours requirements due to a relaxation of time constraints during prime childbearing years that allows them to pursue their desired fertility choices. In addition, there is evidence that the entry response is the strongest among women from medical school settings with low female representation in time-intensive specialties, potentially due to the prospect of additional women entering specialties that were historically majority male.

This paper demonstrates that the long, inflexible work hours ubiquitous among professional occupations have important implications for occupational segregation by gender. From a policy perspective, reducing hours requirements could be considered an effective tool alongside the many gender diversity initiatives enacted by employers. From an economic perspective, as discussed by Hsieh et al. (2016), there could be additional societal benefits from removing occupational frictions that may inhibit certain demographic groups from pursuing their comparative advantage in human capital accumulation and occupational choice.

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," 2015.

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, Esther Duflo, and Sendhil Mullainathan, "How Much Should We Trust Difference-in-Difference Estimates?," *Quarterly Journal of Economics*, 2004, *119* (1), 249–275.

Blau, Francine D and Lawrence M Kahn, "The Gender Wage Gap: Extent, Trends, and Explanations," *NBER Working Paper #21913*, 2016.

Blau, Francine D., 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.

Brooks, Joanna Veazey and Charles L. Bosk, "Remaking Surgical Socialization: Work Hour Restrictions, Rites of Passage, and Occupational Identity," *Social Science and Medicine*, 2012, *75* (9), 1625–1632.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90* (3), 414–427.

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," 2016.

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.

Fernandez, Raquel and Alessandra Fogli, "Culture: An Empirical Investigation of Beliefs, Work, and Fertility," *American Economic Journal: Macroeconomics*, 2009, *1* (1), 146–177.

Freeman, Brian, *The Ultimate Guide to Choosing a Medical Specialty*, 2 ed., McGraw Hill, 2007.

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.

Hamilton, Abigail R, Mark D Tyson, Julie A Braga, and Lori B Lerner, "Childbearing and Pregnancy Characteristics of Female Orthopaedic Surgeons," *The Journal of Bone and Joint Surgery*, 2012, *94* (11), 1–20.

Hsieh, Chang-Tai, Erik Hurst, Charles Jones, and Peter Klenow, "The Allocation of Talent and U.S. Economic Growth," 2016.

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.

Jolly, Shruti, Kent a Griffith, Rochelle DeCastro, Abigail Stewart, Peter Ubel, and Reshma Jagsi, "Academia and the Profession Annals of Internal Medicine Gender Differences in Time Spent on Parenting and Domestic," *Annals of Internal Medicine*, 2014, *160*, 344–353.

Kellogg, Katherine C, *Challenging Operations: Medical Reform and Resistance in Surgery*, University of Chicago Press, 2011.

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," 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.

Ly, Dan P, Seth A Seabury, and Anupam B Jena, "Differences in Incomes of Physicians in the United States by Race and Sex: Observational Study," *BMJ*, 2016, *353* (2923), 1–8.

Mas, Alexandre and Amanda Pallais, "Valuing Alternative Work Arrangements," *NBER Working Paper #22708*, 2016.

Modestino, Alicia Sasser, "The Impact of Managed Care on the Gender Earnings Gap among Physicians," *Federal Reserve Bank of Boston Working Paper #13-1*, 2012.

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.

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.

Seabury, Seth A, Amitabh Chandra, and Anupam B Jena, "Trends in the Earnings of Male and Female Health Care Professionals in the United States, 1987 to 2010," *JAMA internal medicine*, 2013, *173* (18), 1748–50.

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.

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.

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.

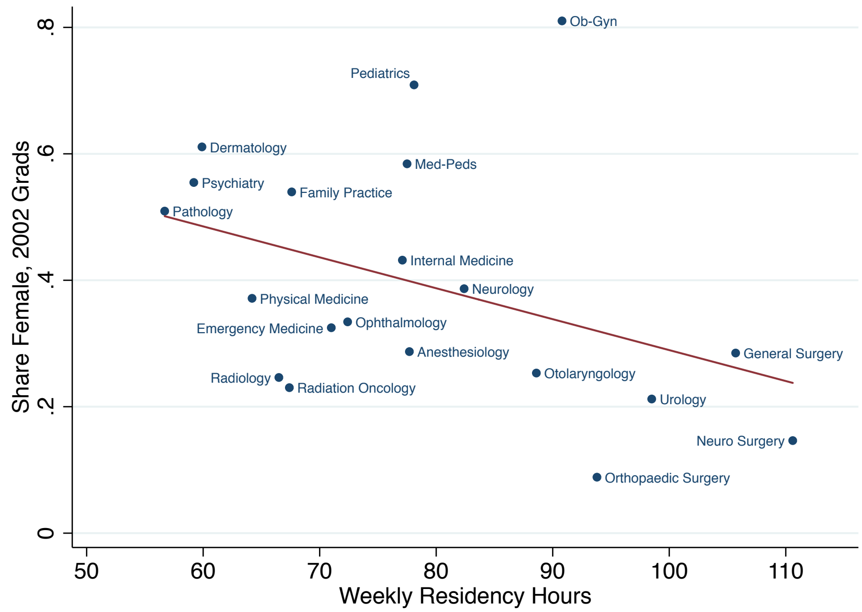
Volpp, Kevin G., Dylan S. Small, Patrick S. Romano, Kamal M F Itani, Amy K. Rosen, Orit Even-Shoshan, Yanli Wang, Lisa Bellini, Michael J. Halenar, Sophia Korovaichuk, Jingsan Zhu, and Jeffrey H. Silber, "Teaching Hospital Five-year Mortality Trends in the Wake of Duty Hour Reforms," *Journal of General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

Wiswall, Matthew and Basit Zafar, "Preference for the Workplace, Human Capital, and Gender," 2016.

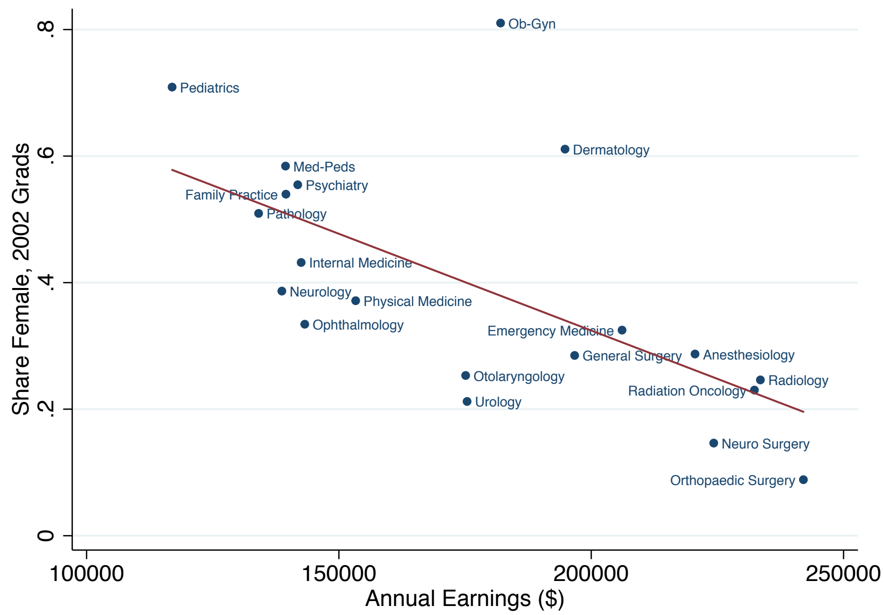
Yoon, Harry H., "Adapting to Duty-Hour Limits: Four Years On," *New England Journal of Medicine*, 2007, *356*, 2668–2670.

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

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



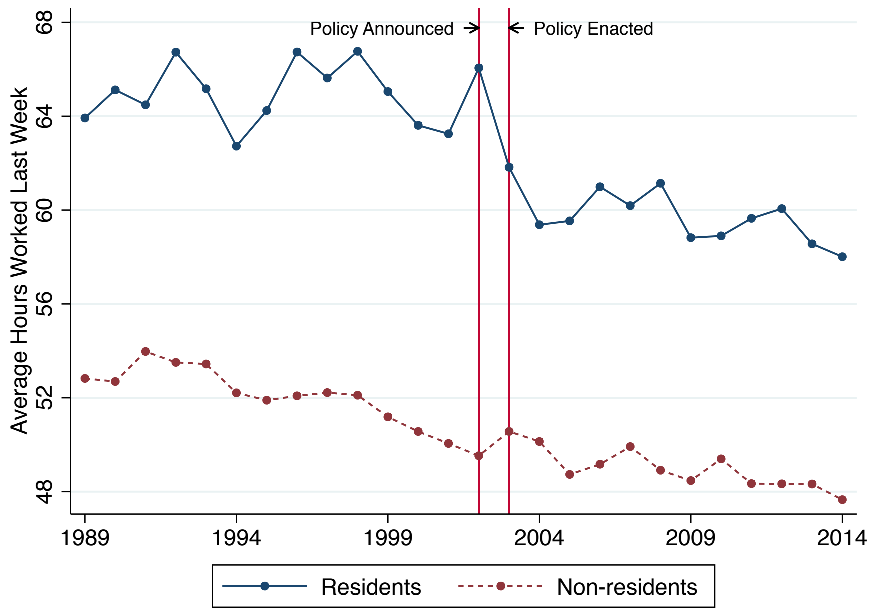
B. Female Share of Specialties and Average Earnings Post-residency



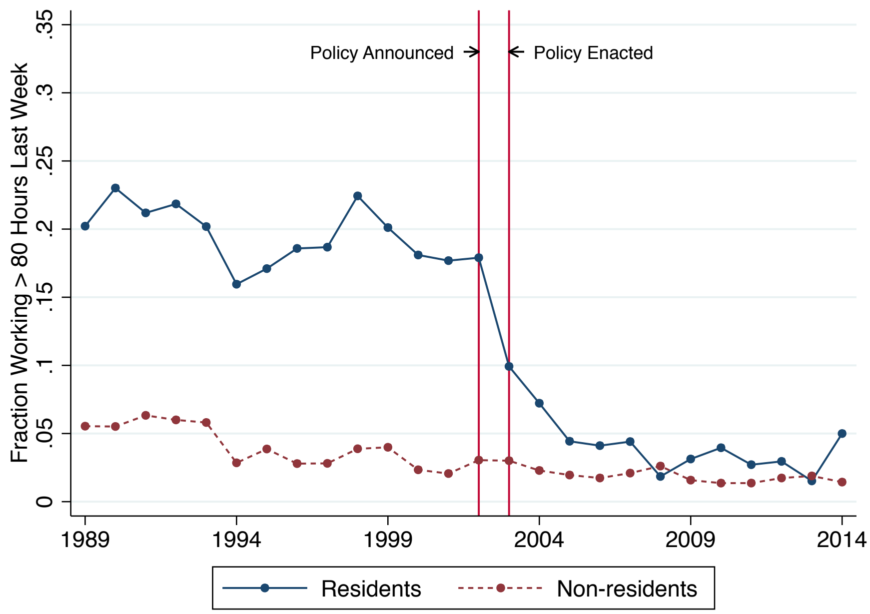
Source: AMA Physician Masterfile, Baldwin Jr et al. (2003), Lo Sasso et al. (2011). Note: This figure plots the fraction of each specialty that is female, using the 2002 U.S. medical school graduation cohort, against: in Panel A, the average hours per week worked during the second year of medical residency; and in Panel B, average post-residency annual earnings for women from the New York State Survey of Residents Completing Training 1999-2008, as reported in Lo Sasso et al. (2011). The solid line in Panel A (B) represents the line of best fit from a regression of female share on hours (earnings).

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

A. Average Hours

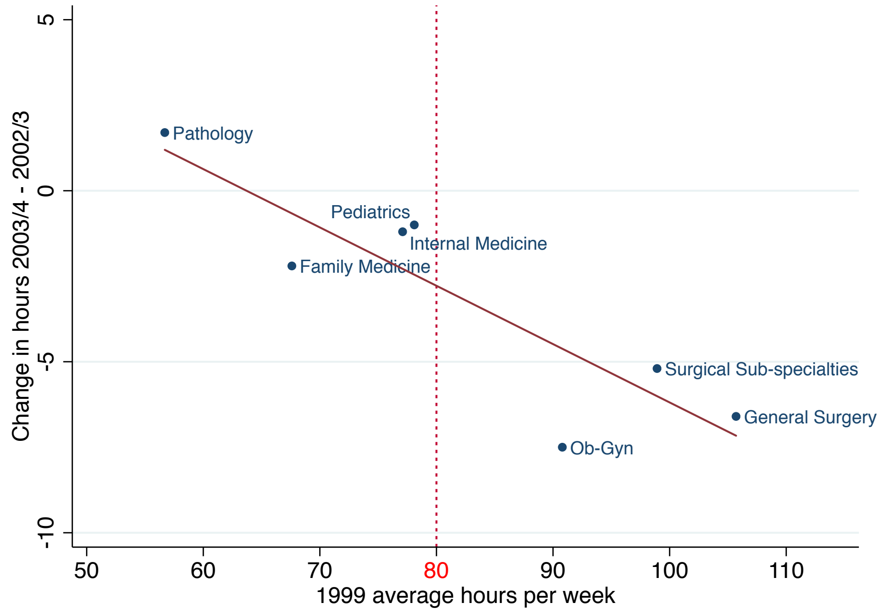


B. Fraction Working > 80 Hours



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.

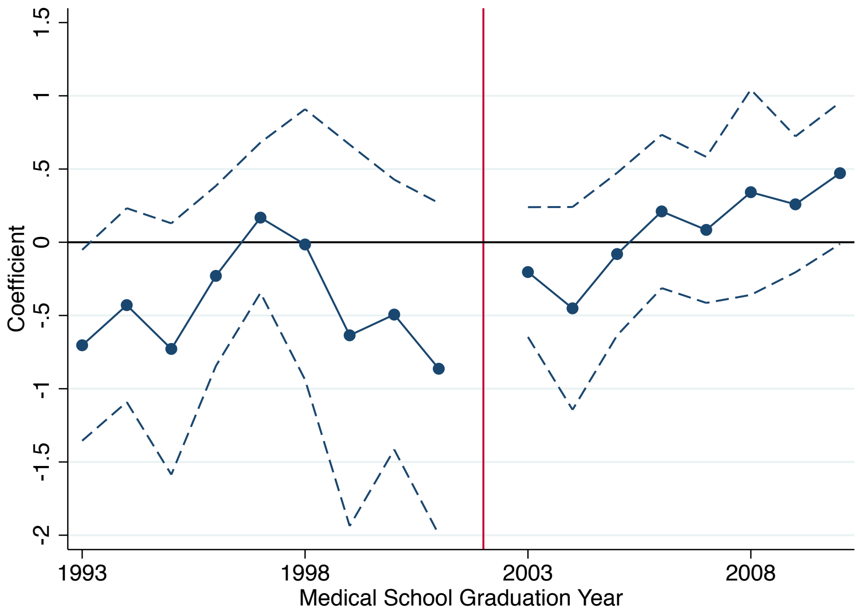
Figure III: Relationship between 1999 Hours and the Change in Hours 2002/3-2003/4, by Specialty



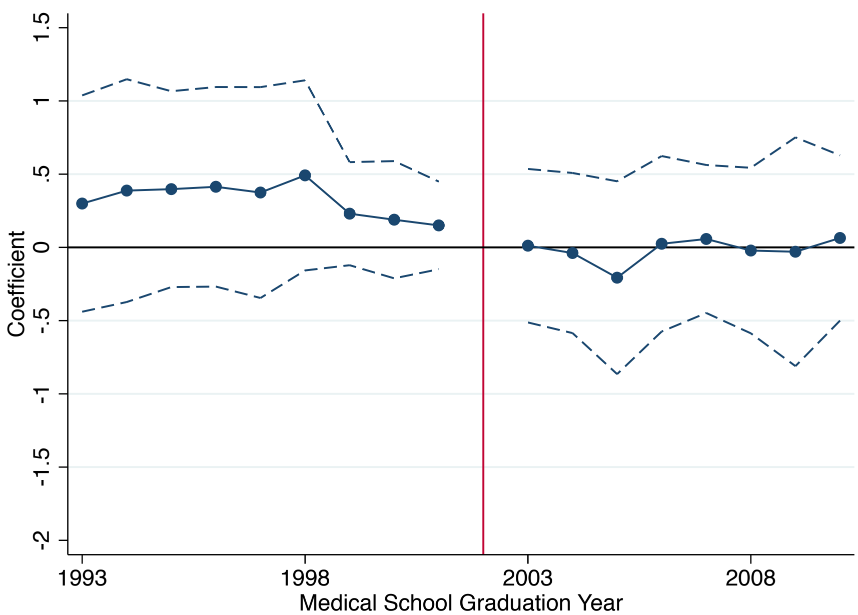
Source: Baldwin Jr et al. (2003), Landrigan et al. (2006) and personal correspondence with author. Note: This figure plots the average number of hours worked per week for second year medical resident physicians in 1999 on the x-axis. The change in average hours per week between residency years 2003/4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

Figure IV: The Effect of the Duty Hour Reform on Specialty Entry: Event Study Analysis

A. Women



B. Men



Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This figure plots the estimated coefficients from the 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 an 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.

Table I: 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
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	414,075	181,861	232,214	281,477	124,817	156,660

Source AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: The Full Sample includes all medical school graduates from years 1993 to 2010, including foreign medical school graduates and osteopaths. The USMG Sample includes only U.S. medical school graduates from years 1993 through 2010. 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 II: The Effect of the Duty Hour Reform on Specialty Entry

Dependent Variable: ln($Share_{st}$)*100	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Female						
Average Weekly Hours × Transition	0.15	0.22	0.08	0.26	0.25	-0.01
	(0.29)	(0.27)	(0.25)	(0.29)	(0.23)	(0.37)
Average Weekly Hours × Post	0.67**	0.78***	0.56**	0.78**	0.77***	0.41
	(0.30)	(0.27)	(0.23)	(0.29)	(0.20)	(0.57)
Panel B: Male						
Average Weekly Hours × Transition	-0.37	-0.22	-0.43	-0.24	-0.19	-0.16
	(0.36)	(0.28)	(0.32)	(0.36)	(0.23)	(0.36)
Average Weekly Hours × Post	-0.27	-0.04	-0.38	-0.15	-0.05	0.07
	(0.41)	(0.31)	(0.35)	(0.42)	(0.25)	(0.42)
P-value for test of equality of male/female coeff.						
Average Weekly Hours × Transition	0.231	0.271	0.238	0.226	0.271	0.643
Average Weekly Hours × Post	0.011	0.011	0.013	0.012	0.015	0.571
Ob/Gyn time trend		X			X	
Primary care specialty time trend			X		X	
Age/medical school controls				X	X	
All specialty time trends						X

Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-difference 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 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 ($Hours_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). All specifications have 360 observations stemming from the analysis of 20 specialties over 18 years. Column 1 reports the results of the baseline specification with no additional controls. Columns 2-6 progressively include specialty-specific controls. Standard errors clustered at the specialty level are reported in parentheses. The p-values at the bottom of the table are from a Wald test of the null hypothesis that the male and female coefficients are equal.

Table III: The Effect of the Duty Hour Reform on Stated Specialty Preference

Dependent Variable: ln($Share_{st}$)*100	(1)	(2)	(3)	(4)	(5)
Panel A: Female					
Average Weekly Hours × Transition	0.32	0.40	0.21	0.30	-0.08
	(0.43)	(0.40)	(0.36)	(0.33)	(0.41)
Average Weekly Hours × Post	0.88*	1.06**	0.66	0.85*	0.06
	(0.50)	(0.49)	(0.42)	(0.44)	(0.56)
Panel B: Male					
Average Weekly Hours × Transition	0.27	0.36	0.17	0.26	0.34
	(0.44)	(0.41)	(0.39)	(0.35)	(0.37)
Average Weekly Hours × Post	0.32	0.50	0.10	0.31	0.47
	(0.52)	(0.49)	(0.43)	(0.41)	(0.54)
P-value for test of equality of male/female coeff.					
Average Weekly Hours × Transition	0.888	0.899	0.902	0.910	0.457
Average Weekly Hours × Post	0.214	0.226	0.242	0.250	0.602
Ob/Gyn time trend		X		X	
Primary care specialty time trend			X	X	
All specialty time trends					X

Source: AAMC Matriculating Student Questionnaire 1998-2007, 2010, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-difference 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 progressively include time-varying specialty controls. Standard errors clustered at the specialty level are reported in parentheses.

Table IV: Exploring Mechanisms for the Differential Responsiveness of Women to the Reform

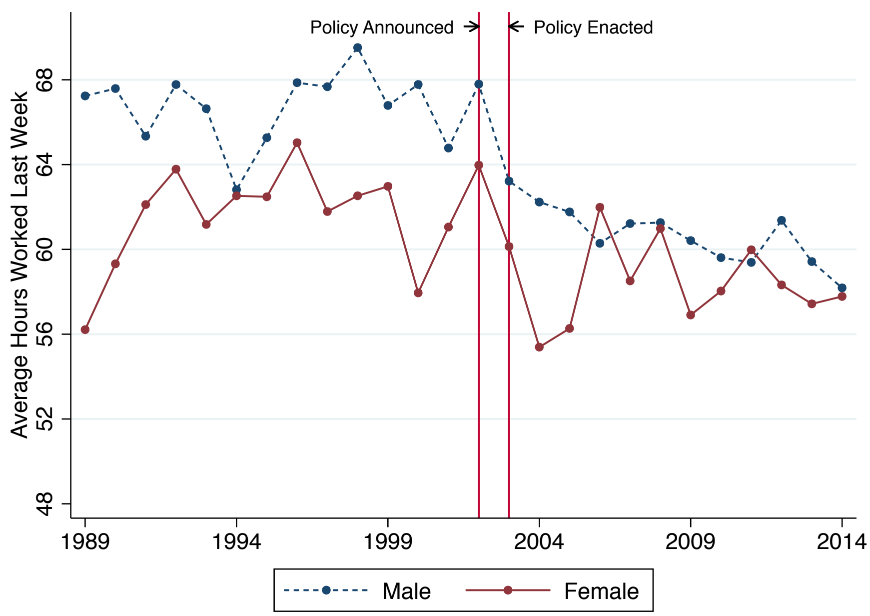
Dependent Variable: ln(Share$_{st}$)*100	(1)	(2)	(3)	(4)	(5)	(6)
	Female Average Age at First Birth in Birth State		Med. School Share of Women Choosing High Hours Specialties		Med. School Fraction Female among High Hours Specialties	
	High	Low	High	Low	High	Low
Panel A: Female						
Average Weekly Hours × Transition	-0.29	0.55*	-0.09	0.69*	0.13	0.15
	(0.45)	(0.27)	(0.29)	(0.38)	(0.28)	(0.41)
Average Weekly Hours × Post	0.54*	0.92***	0.40	1.21***	0.55*	0.82**
	(0.30)	(0.31)	(0.32)	(0.37)	(0.26)	(0.36)
Panel B: Male						
Average Weekly Hours × Transition	-0.43	-0.17	-0.42	-0.30	-0.23	-0.47
	(0.37)	(0.37)	(0.40)	(0.34)	(0.33)	(0.41)
Average Weekly Hours × Post	-0.40	-0.10	-0.38	-0.14	-0.16	-0.37
	(0.40)	(0.44)	(0.46)	(0.37)	(0.37)	(0.45)

Source: AMA Physician Masterfile, Baldwin Jr et al. (2003), CDC. This table reports the results of the difference-in-difference 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 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 ($\text{Hours}_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). All specifications have 360 observations stemming from the analysis of 20 specialties over 18 years. Columns 1 and 2 divide the sample into individuals whose birth state has an average age at first birth in the year 1990 that is above or below the median state, respectively. Columns 3 and 4 divide the sample into individuals who attended medical schools where the share of women choosing high hours specialties ($>$ 80 hours) pre-reform is above or below the median medical school, respectively. Columns 5 and 6 divide the sample into individuals who attended medical schools where the fraction female among those choosing high hours specialties pre-reform is above or below the median medical school, respectively. Standard errors clustered at the specialty level are reported in parentheses.

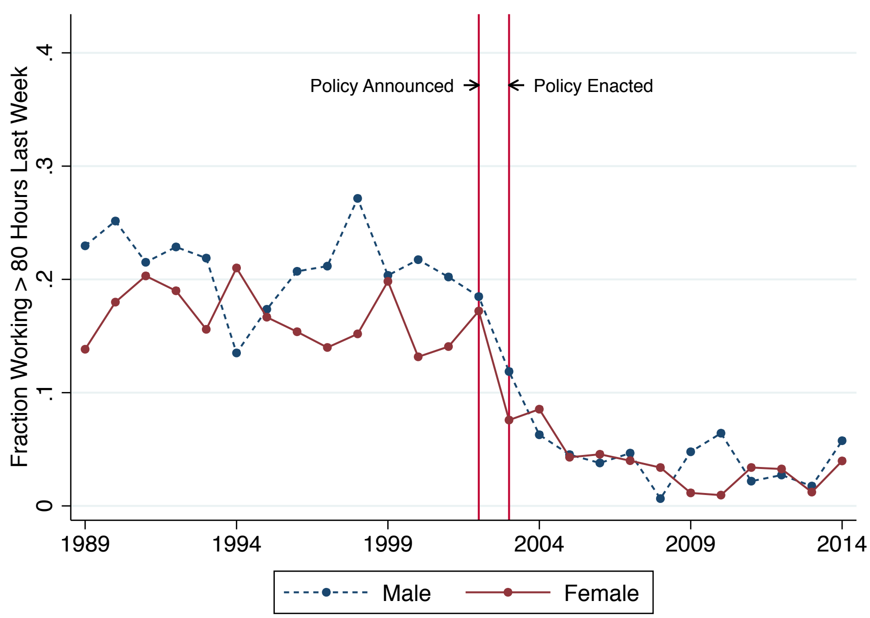
Appendix Figures

Figure A.1: Hours Worked in the Prior Week among Male and Female Resident Physicians, 1989-2014

A. Average Hours



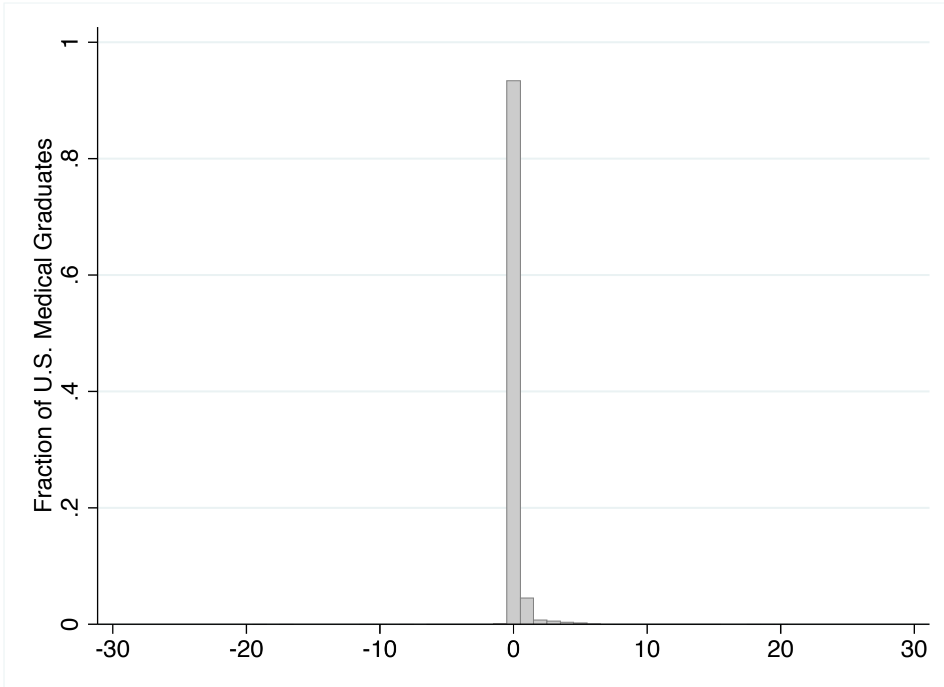
B. Fraction Working > 80 Hours



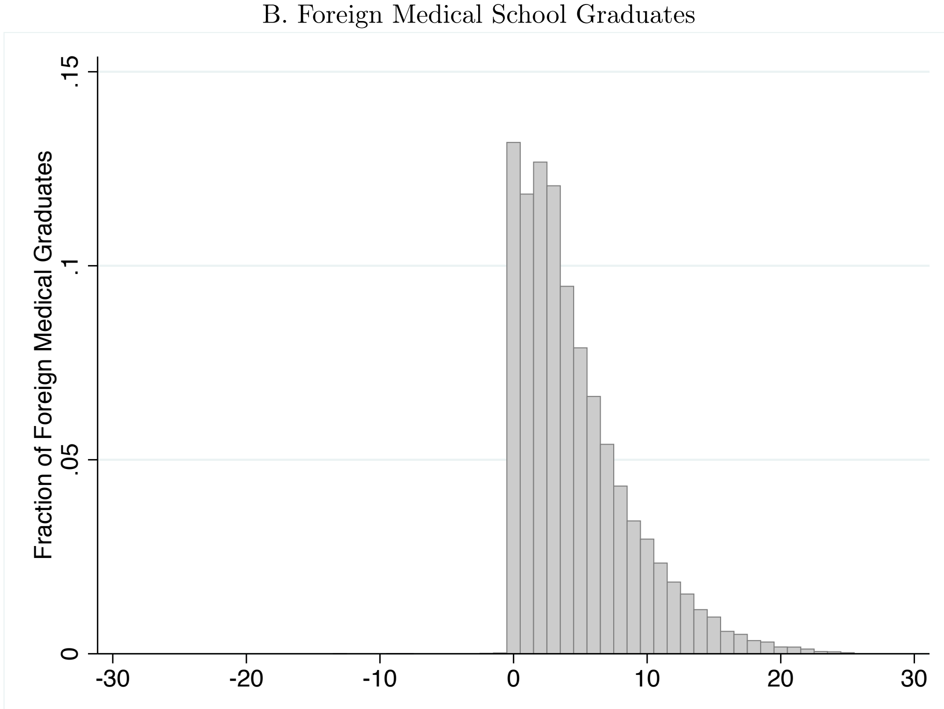
Source: Current Population Survey, monthly files January 1989-December 2014. Panel A plots the average number of hours worked last week for resident physicians, separately for men and women. Panel B plots the fraction of physicians who worked more than 80 hours last week, separately for men and women.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

Figure A.2: Distribution of Difference between Residency Start Year and Medical School Graduation Year

A. U.S. Medical School Graduates



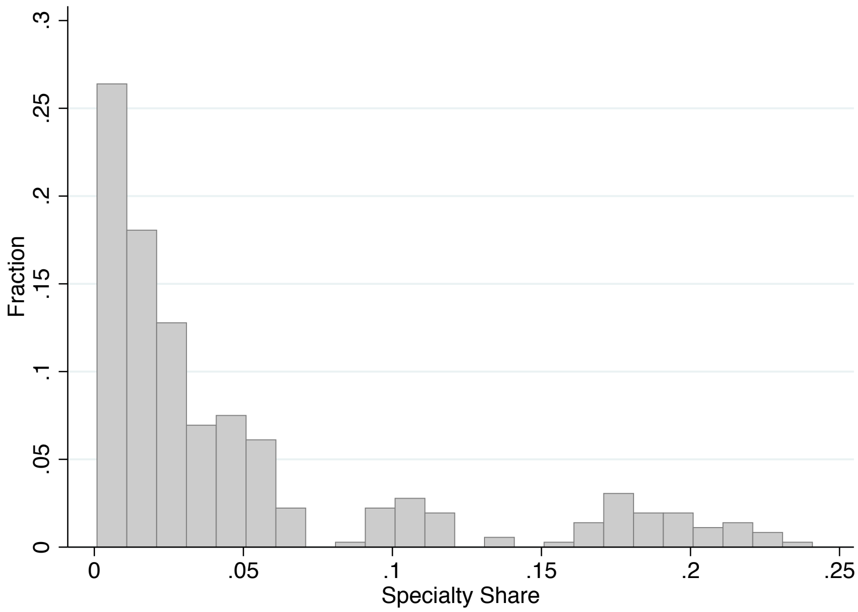
B. Foreign Medical School Graduates



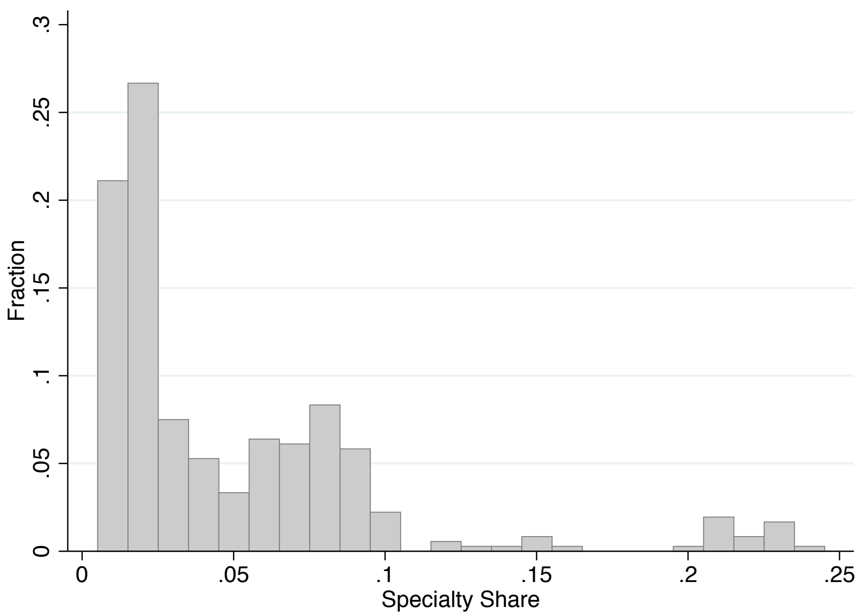
Source: AMA Physician Masterfile Note: This figure plots distribution of the difference between medical school graduation year and residency start year for individuals who graduated from U.S. medical schools and foreign medical schools and attended residency training in California or Texas.

Figure A.3: Distribution of Specialty Shares

A. Female



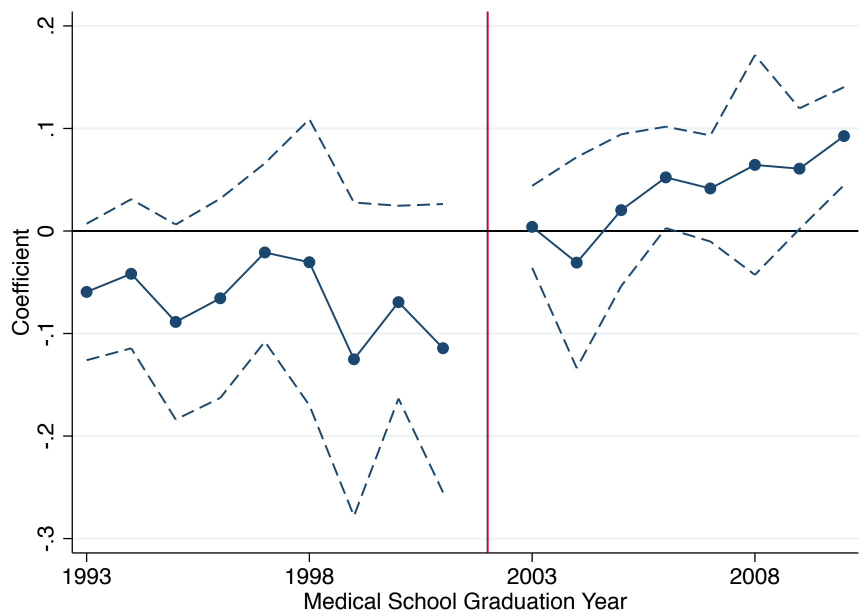
B. Male



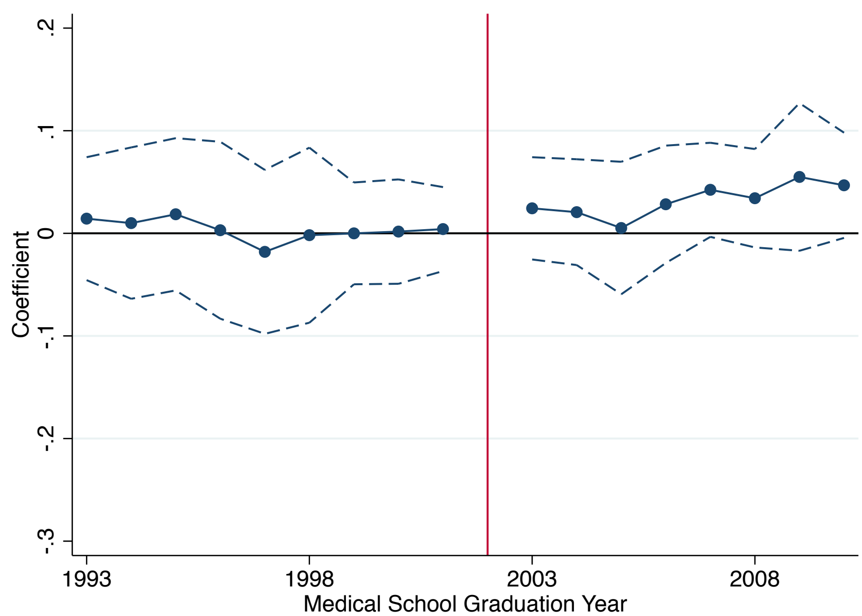
Source: AMA Physician Masterfile Note: This figure plots distribution of specialty shares for male and female U.S. medical school graduates who graduated from medical school 1993 to 2010.

Figure A.4: Event Study Analysis with Alternative Hours Parameterization
Pre-policy Total Hours during Residency

A. Women



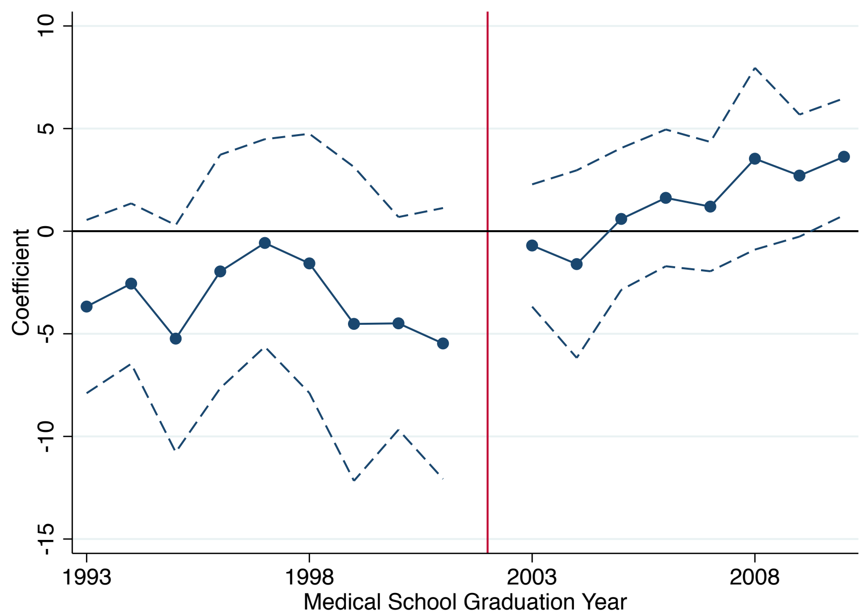
B. Men



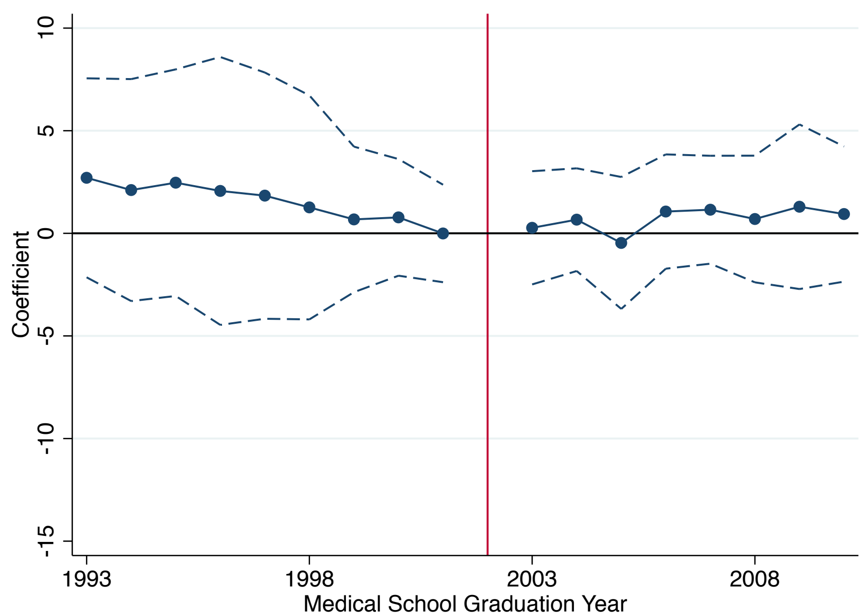
Source: AMA Physician Masterfile, Baldwin Jr et al. (2003), Freeman (2007). Note: This figure plots the estimated coefficients from the event study model with an alternative parameterization of pre-policy hours: the pre-policy total number of hours required by a specialty over the course of residency, which 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 simply result in a rescaling of the estimated coefficients. 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 an 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.5: Event Study Analysis with Alternative Hours Parameterization
Pre-policy Hours above 80 per week

A. Women



B. Men



Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This figure plots the estimated coefficients from the event study model with an alternative parameterization of pre-policy hours: an indicator variable for pre-policy hours above 80 per week. 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 an indicator variables for medical school cohorts. Cohort 2002 serves as the reference year. The solid line plots the coefficients on the interaction term ($\text{Hours}_{1999,s} \times \text{Year}$). The dashed lines plot the 95% confidence intervals based on standard errors clustered at the specialty level.

Appendix Tables

Table A.1: Comparison of Masterfile Sample and AAMC Data on Medical School Graduates

	(1) USMG Sample 1993-2010	(2) AAMC Data	(3) % difference
Medical School Graduation Year			
1993	15,236	15,474	1.54
1994	15,238	15,504	1.72
1995	15,645	15,883	1.50
1996	15,612	15,895	1.78
1997	15,686	15,894	1.31
1998	15,679	15,972	1.83
1999	15,648	16,006	2.24
2000	15,393	15,716	2.06
2001	15,585	15,796	1.34
2002	15,373	15,676	1.93
2003	15,340	15,531	1.23
2004	15,779	15,829	0.32
2005	15,369	15,760	2.48
2006	15,681	15,927	1.54
2007	15,799	16,140	2.11
2008	15,937	16,168	1.43
2009	16,149	16,467	1.93
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 Full Masterfile sample of graduates of U.S. medical schools, by medical school graduation year.

Table A.2: The Effect of the Duty Hour Reform on Specialty Entry: Alternative Methods of Statistical Inference

Dependent Variable: ln($Share_{st}$)*100	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Female						
Average Weekly Hours × Transition	0.15	0.22	0.08	0.26	0.25	-0.01
	[0.616]	[0.431]	[0.752]	[0.387]	[0.307]	[0.971]
	{0.635}	{0.428}	{0.749}	{0.392}	{0.300}	{0.955}
	(0.659)	(0.586)	(0.809)	(0.484)	(0.368)	(0.930)
Average Weekly Hours × Post	0.67	0.78	0.56	0.78	0.77	0.41
	[0.038]	[0.009]	[0.027]	[0.015]	[0.001]	[0.488]
	{0.072}	{0.002}	{0.036}	{0.022}	{0.000}	{0.555}
	(0.104)	(0.050)	(0.128)	(0.064)	(0.018)	(0.420)
Panel B: Male						
Average Weekly Hours × Transition	-0.37	-0.22	-0.43	-0.24	-0.19	-0.16
	[0.317]	[0.440]	[0.191]	[0.512]	[0.411]	[0.673]
	{0.306}	{0.422}	{0.164}	{0.531}	{0.452}	{0.639}
	(0.384)	(0.572)	(0.286)	(0.538)	(0.496)	(0.664)
Average Weekly Hours × Post	-0.27	-0.04	-0.38	-0.15	-0.05	0.07
	[0.508]	[0.908]	[0.293]	[0.719]	[0.840]	[0.862]
	{0.533}	{0.919}	{0.286}	{0.757}	{0.817}	{0.855}
	(0.563)	(0.952)	(0.375)	(0.746)	(0.843)	(0.870)
Ob/Gyn time trend		X			X	
Primary care specialty time trend			X		X	
Age/medical school controls				X	X	
All specialty time trends						X

Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-difference 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 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 ($Hours_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). All specifications have 360 observations stemming from the analysis of 20 specialties over 18 years. Column 1 reports the results of the baseline specification with no additional controls. Columns 2-6 progressively include specialty-specific controls. P-values from three alternative methods of statistical inference are presented: (1) p-values from standard errors clustered at the specialty level are reported in brackets; (2) p-values from wild cluster bootstrapped t-statistics are reported in braces; and (3) p-values from permutation tests are reported in parentheses.

Table A.3: The Effect of the Duty Hour Reform on Specialty Entry: Alternative Hours Parameterizations

Dependent Variable: ln($Share_{st}$)*100	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Female						
Total Residency Hours × Transition	0.06	0.06	0.03	0.07	0.04	0.05
	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Total Residency Hours × Post	0.12***	0.13***	0.08***	0.13***	0.09***	0.11**
	(0.04)	(0.04)	(0.02)	(0.04)	(0.02)	(0.05)
Avg Weekly Hours > 80 × Transition	10.93	15.50	5.36	13.79	12.26	6.48
	(11.25)	(9.24)	(10.44)	(10.89)	(8.19)	(10.68)
Avg Weekly Hours > 80 × Post	25.21*	32.60***	16.22	28.50**	26.32***	18.03
	(12.30)	(9.13)	(11.50)	(11.70)	(7.48)	(14.08)
Panel B: Male						
Total Residency Hours × Transition	0.01	0.02	-0.01	0.03	0.00	0.02
	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Total Residency Hours × Post	0.04	0.05	-0.01	0.06	0.01	0.05
	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.04)
Avg Weekly Hours > 80 × Transition	-8.65	-0.58	-14.65	-3.99	-3.16	2.45
	(13.22)	(9.11)	(12.29)	(12.83)	(7.17)	(8.74)
Avg Weekly Hours > 80 × Post	-4.98	8.05	-14.67	-1.77	0.76	12.94
	(15.13)	(10.10)	(13.91)	(15.11)	(8.07)	(11.42)
Ob/Gyn time trend		X			X	
Primary care specialty time trend			X		X	
Age/medical school controls				X	X	
All specialty time trends						X

Source: AMA Physician Masterfile, Baldwin Jr et al. (2003), Freeman (2007). Note: This table reports the results of the difference-in-difference 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 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 an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). The results from specifications utilizing two alternative parameterizations of specialty pre-policy hours are reported: $\text{TotalResidencyHours}_{s,1999}$ 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 simply result in a rescaling of the estimated coefficients. $\text{Hours}_{s,1999} > 80$ is an indicator for whether a specialty's pre-policy weekly hours were in excess of 80. All specifications have 360 observations stemming from the analysis of 20 specialties over 18 years. Column 1 reports the results of the baseline specification with no additional controls. Columns 2-6 progressively include specialty-specific controls. Standard errors clustered at the specialty level are reported in parentheses.

Table A.4: The Effect of the Duty Hour Reform on Specialty Entry: Inclusion of Foreign and Osteopathic Medical School Graduates

Dependent Variable: ln($Share_{st}$)*100	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Female						
Average Weekly Hours × Transition	0.30	0.35	0.25	0.38	0.36	-0.00
	(0.27)	(0.26)	(0.25)	(0.29)	(0.27)	(0.27)
Average Weekly Hours × Post	0.86***	0.95***	0.78***	0.90***	0.89***	0.38
	(0.23)	(0.21)	(0.20)	(0.25)	(0.20)	(0.35)
Panel B: Male						
Average Weekly Hours × Transition	-0.31	-0.19	-0.35	-0.12	-0.13	-0.11
	(0.24)	(0.19)	(0.22)	(0.21)	(0.17)	(0.28)
Average Weekly Hours × Post	-0.24	-0.05	-0.30	-0.15	-0.07	0.08
	(0.32)	(0.23)	(0.29)	(0.29)	(0.20)	(0.33)
Ob/Gyn time trend		X			X	
Primary care specialty time trend			X		X	
Age/medical school controls				X	X	
All specialty time trends						X

Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-difference specification for specialty entry, estimated separately for men and women, on the sample of U.S. medical school graduates, foreign medical school graduates, and osteopathic medical school graduates. 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 ($\text{Hours}_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). All specifications have 360 observations stemming from the analysis of 20 specialties over 18 years. Column 1 reports the results of the baseline specification with no additional controls. Columns 2-6 progressively include specialty-specific controls. Standard errors clustered at the specialty level are reported in parentheses.