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

Melanie Wasserman*

February 2018

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. Using data on the universe of U.S. medical school graduates from 1993 to 2010, I find that when a specialty reduces its weekly hours, more women enter the specialty, whereas there is little change in men's entry. To shed light on why women and men respond differently to the reform, I analyze physicians' family formation choices during residency. I link female resident physicians to administrative birth records from two large U.S. states and find that reducing a specialty's weekly hours increases its female fertility rate in California. I discuss these results in the context of a model in which individuals choose work and family investments during their early careers, trading off long-term incomes, investments in children, and leisure.

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

*UCLA Anderson School of Management. Address: 110 Westwood Plaza, Entrepreneurs Hall C521, Los Angeles, CA 90095. E-mail: melanie.wasserman@anderson.ucla.edu. I thank David Autor, Esther Duflo and Heidi Williams for their extensive guidance and support. I have benefited from discussions with Nikhil Agarwal, Josh Angrist, Martha Bailey, Alex Bartik, Manasi Deshpande, Joseph Doyle, Claudia Goldin, Jon Gruber, Sally Hudson, Erin Johnson, Danielle Li, 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, ACLEC, and Chicago Harris. 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 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 challenges, 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 1 provides a snapshot of the heterogeneity in male and female specialty outcomes for the 2002 cohort of U.S. medical school graduates. Panel A plots the share of a medical specialty that is female against the specialty's average hours worked per week during the second year of medical residency, while Panel B plots the female share against the specialty's hourly earnings 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 5 to 16 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. I then return to the hypothesis in Goldin (2014) and assess the implications of the rearrangement of women across specialties due to reducing early career time requirements for the physician gender wage gap. A back of the envelope calculation suggests that the entry of women into historically time-intensive and highly compensated specialties due to the reform will narrow the physician gender wage gap by 13 percent.

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.

In the next portion of the paper, I probe why women are more responsive than men to a reduction in a specialty's hours. Since 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, it is plausible that preferences over the timing of family formation could generate women's responsiveness to the reform. I develop a conceptual framework in which physicians jointly choose their medical specialty and whether to have children during residency, with women deriving more disutility from hours worked when they have children. The model demonstrates that a reduction in a specialty's hours can induce more women to enter the specialty, but yields an ambiguous prediction regarding the effect of an hours reduction on the specialty's female fertility rate. The ambiguity arises due to two potentially offsetting phenomena: the effect of the reduction in hours on the fertility of inframarginal women and the effect of the hours reduction on the composition – in terms of preferences regarding having children during residency – of women who enter time-intensive specialties. Depending on the relative magnitudes of these responses, a specialty's fertility rate can rise, fall, or stay the same in response to a reduction in hours.

I empirically investigate the effect of time requirements on the timing of family formation. To do so, I construct a novel linkage between administrative physician data and Vital Statistics birth records from two large states, California and Texas, and examine the effect of the duty hour reform on female fertility choices during the first three years of residency. I find evidence that the reform increases a specialty's fertility rate during residency in California, but no evidence of an increase in Texas. In California, a four hour per week reduction results in a 5 to 15 percent increase in a specialty's fertility rate. Since the effect of the reform on a specialty's fertility rate encompasses both the treatment effect – the effect of an hours reduction on the fertility of inframarginal women – and compositional changes from the new entrants, I propose an empirical strategy to separately identify these two effects. The results of this exercise suggest that a substantial portion of the estimated positive effect of the reform on fertility in California stems from the changing fertility choices of inframarginal women, that is, the reduction in hours relaxes a constraint on childbearing for women who would have chosen the specialty absent the reform. In addition, there is suggestive evidence that the marginal female who enters a specialty due to the reform is *less* likely to have a child during residency, perhaps due to existing childcare demands or the willingness to delay fertility until after residency training. I discuss how the contrasting findings in California and Texas are potentially driven by differences in the magnitude and composition of women induced to enter more time-intensive specialties in the two states. Considering the specialty entry and fertility findings in tandem, these results indicate that time requirements during early careers are deterministic of women's career paths as well as the timing of their childbearing decisions.

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 discusses the empirical framework for examining the effect of hours requirements on specialty choice, presents the main results on specialty entry and stated specialty preferences, and explores heterogeneity in the effect of the reform by residency program attributes. Section 5 presents theoretical and empirical evidence on the effect of hours requirements on fertility timing. Section 6 characterizes the implications of 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 first portion of the USMLE occur early in medical school, as depicted in the medical school timeline in Figure 2, students often plan years in advance in order to emerge a competitive applicant.

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

In addition to a specialty's monetary payoff, factors that have been cited as influential in specialty choice include the practice setting (hospital, solo practice, group practice), extent of interaction with patients, intellectual content, and lifestyle considerations such as the number of hours and the extent to which the

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

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 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

[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, 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.

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 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 3 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

[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.

after the introduction of the policy in 2003, there is a discrete drop of 4 hours per week, with the reduction sustained over the subsequent years. Since the reform restricted average hours per week to 80, we expect the upper end of the hours distribution to be primarily affected. Figure 3 Panel B plots the fraction of physicians who worked more than 80 hours in the previous week separately for resident and non-resident physicians. Consistent with the stipulations of the policy, right after its introduction in 2003, there is a precipitous fall of more than ten percentage points in the fraction of medical residents who worked more than 80 hours per week, whereas there is little change among non-resident physicians.[9]

While the duty hour cap was common across all medical specialties, it should have had disproportionate impacts on the most time-intensive specialties, such as General Surgery and Urology, where the typical resident pre-reform worked far in excess of 80 hours per week (Philibert et al., 2009). To confirm whether this hypothesized pattern is substantiated by trends in hours worked by specialty, I use reports of hours worked from three surveys of medical residents that were conducted either pre-reform by the ACGME or by independent researchers. For a measure of pre-policy hours, I use data from a 1999 nationally representative survey of 2,000 second year medical residents conducted by Baldwin Jr et al. (2003). To measure the change in hours before and after the implementation of the reform, I use two nationally representative surveys conducted by Landrigan et al. (2006), which collected reports of hours worked from approximately 2,700 first year residents in 2002 and 1,300 first year residents in 2003. Figure 4 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

[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).

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 year of graduation) and primary 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").[11] 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,[12] I crosswalk the more detailed specialty information in the AMA Masterfile to the coarser categories, using a classification scheme provided by the Dartmouth Atlas.[13] 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

[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.

[11] 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.

[12] 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.

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

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 school students. In survey years 1998-2006 and 2009-2010, students are asked the specialty category they are considering upon enrollment in medical school. Twenty-six specialties are represented in the MSQ and I again crosswalk these specialties to the 20 specialty categories described above.

I explore the allocation of new entrants across residency programs using the American Association of Medical Colleges' (AAMC) National Graduate Medical Education Census Track survey of residency program directors. The data set contains program-level attributes for every U.S. residency program 1996-2010, including average hours per week worked by first year residents, provision of parental leave, availability of onsite childcare, and gender composition of full-time faculty. The data set includes the gender composition of residents, which permits a residency program-level analysis of the effect of the reform and how the effect varies by baseline program attributes.

In order to analyze the effect of the reform on the family formation decisions of medical residents, I construct a new linkage between the CA/TX Masterfile and Vital Statistics birth records from California and Texas, for years 1993-2013. The birth records include parental identifiers (mother/father name, date and place of birth), child birthdate and demographic characteristics. Due to limited information on fathers, it is only possible to link female physicians. I conduct a probabilistic merge using female physician first name, last name (maiden and/or current), year of birth, and birthplace (U.S. state or country). Texas birth certificates additionally record mother occupation, which I use as a post-merge verification of the quality of the match. The result of this merge is the fertility outcomes of female physicians who started residency 1993-2010 in CA or TX. I limit the analysis to fertility outcomes during the first three years of residency, since residency lasts at least three years. After residency training, a substantial fraction of individuals move out of state, precluding observation of their fertility decisions in the state Vital Statistics birth records.

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 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.[14] 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.[15] For the fertility analysis, the sample is limited to female USMGs who started residency training between 1993 and 2010, and completed the first three years of residency in California or Texas.[16] The age and missing specialty restrictions are also implemented. The final sample for the fertility analysis is 12,580 female physicians in CA and 7,093 female physicians in TX.

Table 1 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. Appendix Table A.2 reports the pre-policy distribution of physicians across medical specialties, by gender. For both men and women, the two largest specialties are Internal Medicine and Family Practice.

[14] 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.

[15] 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.

[16] There are a few consequences of this data restriction. First, if an individual drops out of residency before the third year, then she is not included in the sample. Approximately four percent of individuals complete the first year of residency in California and have no information on the subsequent years of residency training, indicating that they did not proceed past the first year. Second, if an individual completes one or two years of training in California, and then moves to another state for the remainder of residency, then she will not be included in the sample. About 89 percent of individuals who complete their first year of residency in California go on to complete the second and third year of training in California.

After that, we observe a divergence in the specialty entry patterns of men and women. The next largest specialties for men are General Surgery, Pediatrics, and Emergency Medicine, while for women they are Pediatrics, Obstetrics/Gynecology, and Psychiatry.

The bottom half of Table 1 presents summary statistics for the CA and TX samples, respectively. In comparison to the female USMG sample, the CA sample is less likely to be born in the U.S., and is more likely to have attended a ranked medical school. The TX female sample is comparable to the female USMG population, with the exception that it is slightly younger. The final rows of the table report summary statistics for female fertility outcomes during the first three years of residency. Among CA female medical residents, the average number of children during the first three years of residency is 0.13, with 12 percent of women having at least one child. In comparison to their counterparts in CA, female residents in TX are much more likely to have children during residency, with 19 percent having at least one child and the average number of children 0.20.

4 The Effect of the Duty Hour Reform on Specialty Entry

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 4, average hours declined across all specialties with pre-policy hours near or above 80, with the steepest reductions among the specialties with the highest pre-policy hours. The second source of variation stems from the timing of an individual's residency training. Medical school graduates who started residency training before 2003 were not aware of the hours restrictions at the time they chose a specialty for medical residency. Medical school graduates who started residency from 2003 onward, however, had the capacity to select their specialty, taking into consideration the reduction in hours associated with the reform.

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.[17] 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 5 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. To probe the identifying assumption of parallel trends in more and less time intensive specialties over time, I test the joint significance of the pre-policy coefficients and find a p-value of 0.40. While the individual coefficients after the reform tend to be imprecisely estimated, they are jointly significant with a p-value of less than 0.001.[18] For men, there is a negative (but statistically insignificant) 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.

As observed in Figure 5 Panel A, the effect of the duty hour reform on female specialty entry appears to increase over time. There are three possible reasons for this upward trend. First, consistent with the discussion in Section 2.1 on human capital investments necessary for specialty selection, it would have been difficult for students who were in the final years of medical school when the reform was introduced to change their specialties. Second, in line with the discussion in Section 2.3, there may have been weaker effects of the reform immediately after the roll out due to lagged implementation or imperfect compliance from residency programs. As documented in Figure 3, while there was an immediate drop in the hours of medical residents

[17] 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.

[18] Restricting the test of joint significance to the positive coefficients also yields a p-value of <0.001.

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:

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

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 2 reports the effects of the reform on specialty entry, separately for men and women. Starting with Table 2 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.[19]

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 paths over time. The event study analysis above provides evidence in support of this

[19] 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.

assumption. I further test the parallel trends assumption through the inclusion of specialty-specific controls. Certain specialties and specialty groups exhibit strong trends prior to the duty hour reform. For example, as depicted in Figure 1, Obstetrics/Gynecology (Ob/Gyn) is an outlier specialty in that it is highly time-intensive but has experienced substantial female entry over the last 30 years. There has also been declining interest in the primary care specialties: Family Practice, Internal Medicine and Pediatrics. In Table 2 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 little.

In order to gauge whether compositional changes in the male or female medical school graduate population are potentially driving these results, column 4 controls for demographic characteristics of residents, including age at medical school graduation and quality of medical school attended. Column 5 incorporates the Ob/Gyn, primary care, and demographic controls. The main results are qualitatively the same with the inclusion of these individual controls. In column 6, I include all specialty-specific linear time trends. Even in this restrictive specification, 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. As expected, this specification has low statistical power since the specialty-specific trends absorb a substantial portion of the identifying variation. To further test the sensitivity of the results to specialty-specific trends, I re-estimate equation (3) 20 times, dropping one specialty each time. The results for both men and women are robust to this check.[20] The results are additionally robust to various parameterizations of pre-policy hours requirements during residency (Appendix Table A.4) and the inclusion foreign and osteopathic medical school graduates (Appendix Table A.5).[21]

I account for 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).[22] Appendix Table A.3 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

[20] These results are available upon request.

[21] 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. 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.

[22] 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 2. 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.

at the specialty level, but the results remain statistically significant in the majority of specifications.[23]

How large is the effect of the duty hour reform on female specialty entry? To answer this question, I provide an estimate of the effect of an hours reduction on specialty entry by computing 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).[24] 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 2 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 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.[25] 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.[26]

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

[23] 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.

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

[25] 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.

[26] 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.

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.[27] Second, conditional on individuals submitting applications to programs of a given specialty, residency programs in more time-intensive specialties could shift their hiring practices post-reform. On one hand, the reform could improve the hiring prospects of female candidates, if work hours were previously considered an impediment to program success particularly for women. In this scenario, 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. On the other hand, if hours were known to serve as a constraint for women in particular and opting into high hours requirements was utilized by residency programs as a proxy for applicant quality, then eliminating this proxy could have detrimental effects on female applicants' prospects. [28]

To disentangle the labor supply and the labor demand channels, I assess whether how the reform affected stated preferences for specialties using data from the Association of American Medical Colleges (AAMC) Matriculating Student Questionnaire (MSQ) on the specialty category students are considering upon enrollment in medical school, for students entering in years 1998-2006 and 2009-2010.[29] 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.[30]

The results in Table 3 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

[27] 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.

[28] This is similar to the unintended consequences for African Americans of "ban the box" policies . In addition, affirmative action for female applicants could imply that similar labor supply responses among men and women translate into greater female entry. In contrast, discriminatory attitudes toward female applicants would have the opposite result.

[29] 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.

[30] I continue to allow the effects of the reform to evolve over time due to lagged information dissemination or reform implementation.

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 3 are larger than the main estimates of the effect of the reform on specialty entry in Table 2, which suggests that post-reform, the increased interest – for both women and men – could not be fully accommodated by available residency slots.

4.3 Implications for Talent Allocation

A growing literature suggests that the underrepresentation of women in certain occupations could arise from occupational frictions that inhibit women from pursuing their comparative advantage in human capital accumulation and occupational choice, leading to the misallocation of talent (Hsieh et al., 2016). If this is the case, we would expect specialties with low female representation to have the most talented women. Indeed, before the reform was introduced, there was a negative relationship between a specialty's female representation and the fraction of women in the specialty who attended a ranked medical school. In addition, women in time-intensive specialties were more likely to have attended a ranked medical schools than their male counterparts. The talent misallocation hypothesis also has implications for the marginal female entrant. If the hours cap relaxes an occupational friction, then the marginal female induced to enter a time-intensive specialty due to the reform should be of lower quality than the average female entrant.[31]

To test this prediction regarding the quality of the marginal female entrant, I estimate the following specification:

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

where $FractionRanked_{st}$ is the fraction of individuals in specialty s from medical school cohort t who attended a ranked medical school. The specification is estimated separately for men and women. The results are presented in Appendix Table A.6. In terms of quality differentiation, the new female entrants are slightly less likely to come from a ranked medical school, with the coefficients implying a four hour reduction reduces the fraction of female entrants from a ranked medical school by 0.86 to 2.8 percentage points. This is consistent with women selecting into time-intensive specialties in part based on aptitude,

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

with the marginal female entrant of lower aptitude than the inframarginal women. It is interesting to note that this compositional change is also present among male entrants, particularly in the period immediately after the reform's introduction, even absent an effect of the reform on men's specialty entry.

4.4 Residency Program Drivers of Female Entry

Recall that medical school students first select a specialty and then apply to programs within that specialty. Given the positive effect of the reform on women's specialty entry and null effect on men's entry, we would expect that after the reform, female representation in residency programs in more time-intensive specialties will rise relative to programs in less time-intensive specialties. In this section, I confirm this positive effect of the reform on programs' female representation and additionally investigate heterogeneity in the effects of the reform by baseline attributes of residency programs. Building on the literature on the effects of same-gender role models on career path choice, the first set of attributes is related to programs' baseline female representation, including female representation among full-time physician faculty who supervise residents and among first year residents. The second set of attributes captures whether programs have official policies to accommodate having children during residency, including whether the program has a maternity leave policy (paid or unpaid) and provides onsite childcare. The heterogeneity analysis will provide suggestive evidence of the characteristics of residency programs that absorb new female entrants, either due to programs with those attributes being more amenable to hiring women or due to female residency applicants preferring to work in programs with such attributes.

For this analysis, I employ residency-program level data from AAMC Graduate Medical Education Census Track survey of residency program directors for years 1996-2010.[32] Summary statistics are reported in Appendix Table A.7 for the overall GME Census Track sample and by specialty. The sample consists of 2,567 programs and 34,253 program-year observations, spanning 19 of the 20 broad specialties from the above analysis.[33] As expected, there is substantial variation across specialties in the female representation among residents and faculty. In addition, there is a positive correlation between the fraction of the faculty and the fraction of the residents who are female, a relationship that holds even within specialty. There is far less variation in family-friendly policies across specialties. Before the reform, there is no relationship between the female representation in a program and the program's provision of maternity leave and onsite

[32] On a yearly basis, residency program directors are asked to update program attributes, the answers of which are listed in a web-based directory known as Fellowship and Residency Electronic Interactive Database Access. (FREIDA) Online and utilized for informational purposes by prospective residency program applicants.

[33] I exclude fellowship programs in sub-specialties (e.g. Cardiology, Vascular Surgery) that require completion of a previous residency program before entry. I also exclude the broad specialty Internal Medicine-Pediatrics, since it is inconsistently defined across program years. I include residency programs that require a transitional year or preliminary years of training in other specialties (e.g. Urology, Diagnostic Radiology). Residency programs with one resident are excluded, but their inclusion does not qualitatively change the results. An unbalanced panel of program is used in the main specifications, with 95% of programs that meet the other sample restrictions observed 14 or 15 times throughout the 1996-2010 time period. Dropping the 5% of programs with fewer than 14 observations yields similar results.

childcare.[34]

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

$$FractionFemale_{pst} = \beta_0 + \beta_1(\text{Hours}_{ps,1996} \times \text{Transition}_t) + \beta_2(\text{Hours}_{ps,1996} \times \text{Post}_t) + \alpha_p + \gamma_t + \epsilon_{pst} \quad (5)$$

where $FractionFemale_{pst}$ is the fraction of first year residents in program p in specialty s in year t who are female, $Hours_{ps,1996}$ are the average hours per week, reported by the residency program director in 1996, and α_p are program fixed effects. Since the survey asks residency program directors the average hours per week of first year residents, I use the 1996 measure for pre-policy time intensity at the program level. Specifications using the specialty-specific hours from Baldwin Jr et al. (2003) yield similar results.[35] Using this alternative unit of analysis and measure of pre-policy time intensity, the results in Appendix Table A.8 demonstrate that the reform has a positive and statistically significant effect on the fraction of residents in a program who are female. Consistent with the specialty entry analysis, the effects are larger in the 2006-2010 (Post) period than in the 2003-2005 (Transition) period; an additional hour pre-policy causes a 0.05 percentage point increase in the fraction of residents in a program who are female in the Transition period and a 0.07 percentage point increase in the Post period.

Next I test whether the effects of the reform are larger among programs that have higher baseline female representation or family-friendly policies. I estimate equation (5) separately for programs above/below the specialty-specific average for female representation and for programs with/without family-friendly policies, where each of these classifications stems from the program's 1996 survey response. The results are presented in Table 4. In panel A, the effect of the reform on the fraction of residents who are female is larger among programs that have higher representation of women among their full-time physician faculty and residents. This result is indicative of new female entrants selecting programs with an established female community or, alternatively, programs with established female communities being more amenable to hiring new female entrants after the reform.[36] While this result is not causal, it is consistent with the literature on the positive effects of female role models on women's propensity to enter STEM careers and points to a potential externality of increased female representation (Kahn and Ginther, 2018; Carrell et al., 2010; Canes and

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

[35] Over 40 percent of the variation in hours across programs is accounted for by specialty alone, lending support to the specialty-level analysis above.

[36] A third possibility is that programs with initially higher female representation experienced larger reductions in hours and therefore attracted more women. This could be due to such programs having higher hours initially or exhibiting greater compliance with the provisions of the reform. While I cannot investigate this possibility directly, I implement two tests. First, I test whether pre-reform program female representation is associated with hours. Conditional on specialty, there is no relationship between pre-reform female faculty/resident representation and hours. Second, I re-estimate the heterogeneity specifications using the specialty-specific measure of hours, rather than the program-specific measure, and find similar patterns.

Rosen, 1995). This result also suggests that female role models could be complements to – rather than substitutes for – work hour requirements. Turning to the family-friendly policies, there is little heterogeneity in the effect of the reform based on whether a program has official policies in place that make residency more compatible with having children. This null result could arise because women do not value or anticipate using family-friendly policies. Alternatively, the result could be due to mismeasurement of such policies, that is, the existence of a policy does not capture the generosity of the policy or norms surrounding policy take-up.[37]

5 Investigating Fertility Timing as a Mechanism

If men and women have 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. The 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? In this section, I explore whether women's greater response to early career hours requirements in their specialty choices stems from work hours constraining their actual or anticipated fertility decisions.

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 residency hours requirements in their career path choices within medicine. Although residency comprises a small minority of physicians' expected working years, it tends to coincide with physicians' late 20's and early 30's, which is a prime childbearing period for women. During residency individuals have access to only limited or costly means to adjust their labor supply on the extensive and intensive margins to accommodate their fertility choices.[38] Due to biological constraints, it could be costly in terms of fecundity for women to delay having children until after their training, particularly if they choose a specialty with a lengthy residency. 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, or social norms that induce women to

[37] After the reform was implemented, 41 percent of female surgeons who had children during residency reported having seriously considered dropping out of their program, with the 80% citing inadequate maternity leave (of no more than 6 weeks) and lack of childcare support as contributing factors. Almost a third stated that they would discourage female medical students from choosing a career in surgery.

[38] Due to the natural progression from medical school to residency training, individuals have limited scope to adjust their labor supply on the extensive margin through delaying their career investments. 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. Taking additional time off from residency requires special permission. On the intensive margin, at the time the reform was implemented, 10 percent of residency programs offered part-time positions. In contrast, fully-trained physicians have more discretion over their work hours through choice of practice setting and number of patients/procedures.

spend more time on parental duties than men.[39]

5.1 Conceptual Framework

To guide our understanding of how a reduction in hours can potentially affect specialty and fertility decisions, I model a physician's choice of medical specialty and whether to have children during residency.[40] The full model is presented in Appendix B. Individuals have preferences over early career work hours (the complement of leisure), whether to have children during residency, and the post-residency wage. Specialties are characterized as bundles of attributes: residency hours and post-residency wages, where the two attributes are positively correlated. Men and women have the same preferences over the timing of children, residency hours, and wages, with the exception that when women have children they incur an additional disutility from work hours.

The framework generates the following predictions regarding specialty choice. Consistent with the empirical evidence presented above, due to the additional disutility from hours worked when women have children, women are less likely than men to enter high hours specialties. Furthermore, if the disutility of hours worked when women have children is convex, then after the reform reduces the hours of high hours specialties, women may be more likely than men to enter these specialties. This model prediction aligns with the findings of the specialty entry analysis in Section 4. The framework also has the following predictions regarding female fertility choices during residency. Women in high hours specialties are less likely than women in low hours specialties and less likely than all men to have children during residency. This pattern is due to women sorting into specialties based on their fertility preferences as well as, conditional on specialty, a subset of women being constrained in their fertility choices by their work hours.[41] Next, the reform should weakly increase childbearing among inframarginal women in high hours specialties, that is, the women who would have chosen a high hours specialty absent the reform. I will refer to this below as the "treatment effect" of the reform on female fertility. Last, depending on distributional assumptions regarding model parameters, the reform can cause the female fertility rate of high hours specialties to rise, fall, or stay the same. The direction of the change in a specialty's female fertility rate depends on the magnitude and composition of the new entrants into high hours specialties and the magnitude of the effect of the reform on childbearing among inframarginal women in high hours specialties. I will refer to this below as the "combined effect" of the reform on female fertility, which incorporates the treatment and composition effects.

[39] 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).

[40] This model abstracts from the decision to become a physician.

[41] Since individuals choose specialties – rather than hours worked during residency – it is possible that the hours within their chosen specialty constrain their fertility choices. Some individuals will work more than their optimal hours, absent the minimum time requirements to enter a specialty.

5.2 Empirical Evidence

The fertility patterns of early career physicians lend support to the baseline predictions of the model. Even though residency occurs during women's main reproductive years, a lower fraction of female than male residents have children.[42] Moreover, female physicians time their childbearing relative to their residency training, with the years immediately after residency the most common time to have children (Hamilton et al., 2012; Turner et al., 2012).[43] Using the linked physician-Vital Statistics data for CA/TX female residents, Figure 6 plots the relationship pre-policy between a specialty's female fertility rate during the first three years of residency and average pre-policy weekly hours.[44] Consistent with the conjectures that work hours constrain women's fertility choices or women choose specialties based in part on preferences regarding the timing of their fertility, there is a negative relationship in both CA and TX.

5.2.1 Strategy for Estimating the Effect of the Reform on Female Fertility

I propose an empirical strategy to estimate the two components of the effect of the reform on a specialty's female fertility rate: (1) the treatment effect of a reduction in hours on female fertility, holding constant the composition of women who are in a high hours specialty, and (2) the effect of the reduction in hours on a specialty's fertility rate though compositional changes from new female entrants into time-intensive specialties. The former effect tells us whether women are constrained by hours in their fertility choices. The latter effect characterizes the fertility choices of the marginal entrant and sheds light on a potential difference between the new entrants into time intensive specialties. I contrast the fertility outcomes of individuals in more and less time-intensive specialties, before and after the reform, and leverage the magnitude and timing of the specialty sorting results from Section 4. I estimate an individual-level specification analogous to the specialty entry specification:

$$Y_{ist} = \beta_0 + \beta_1(\text{Hours}_{s,1999} \times \text{Transition}_t) + \beta_2(\text{Hours}_{s,1999} \times \text{Post}_t) + \alpha_s + \gamma_t + X'_{ist}\delta + \epsilon_{ist} \tag{6}$$

where Y_{ist} is the fertility outcome of individual i from residency cohort t who entered specialty s, $\text{Hours}_{s,1999}$ are the average pre-policy hours of specialty s, Transition_t is an indicator for starting residency training 2001-2005, Post_t is an indicator for starting residency training 2006-2010, α_s are specialty fixed effects, γ_t are residency start year fixed effects, and X'_{ist} is a vector of individual-level controls, including age at medical school graduation and medical school fixed effects. The omitted residency cohorts are 1993-2000. I estimate

[42] This stands in contrast to childbearing patterns among every age, race/ethnic, and education group in the U.S. population, where women's average age at first birth falls below men's and women are more likely than men to have had at least one child (Matthews and Hamilton, 2014).

[43] Park and Rim (2017) document a similar pattern among partner-track lawyers.

[44] Due to confidentiality restrictions regarding small cells, the fertility rates for two specialties could not be reported.

equation (6) using the linked Masterfile-Vital Statistics data with the main outcome the number of children an individual has during the first three years of residency training. The sample includes female U.S. medical school graduates who started residency training in California or Texas between 1993 and 2010.

In order to disentangle the treatment and compositional effects of the reform, I leverage the timing of the specialty entry effects relative to the timing of the reform. Recall from Section 4 that there is a small effect of the reform on women's specialty entry among the Transition_t cohorts. This is particularly the case when the specialty entry analysis is re-estimated restricting the sample to physicians who completed their residency in California (Appendix Table A.9 columns 5 and 6). While the state-specific results are overall imprecisely estimated, to the extent there are positive effects of the reform on female specialty entry in California, they appear to be concentrated among the Post_t cohorts.[45] To estimate the treatment effect of the reform on female fertility, I assume the Transition_t cohorts in CA do not experience compositional changes from the entry of new women, but still experience the reduction in hours during their residency training. Note that the definition of the transition period in equation (6) also incorporates cohorts that started residency in 2001 and 2002. Since these cohorts had already chosen their medical specialty by the time the reform was enacted, they were likely impervious to compositional changes from the hours reduction, but experienced the reduction in hours during their residency training that could alter their residency fertility choices[46]The coefficient β_1 identifies the treatment effect, by contrasting the fertility outcomes of Transition_t and control cohorts, across high and low hours specialties.

Based on the national analysis in Section 4, the Post_t female cohorts altered their specialty choices in response to the reduction in hours and also experienced the reduction in hours due to the reform. The coefficient β_2 identifies the combined effect of the reform on a specialty's fertility rate, incorporating the entry of new women with potentially different fertility choices.[47] Note that, in Texas, even among the Transition_t cohorts there is a positive effect of the reform on female specialty entry (Appendix Table A.9 columns 7 and 8), implying the Texas setting is not suitable for the estimation of the treatment effect of the reform on female fertility. The identifying assumption for the analysis to yield unbiased estimates of the treatment and combined effects of the reform on a specialty's fertility rate, is that absent the reform, the fertility outcomes of individuals in more and less time-intensive specialties would have evolved similarly over time.

[45]Given the weaker specialty entry results in CA, one might ask whether the reform was properly implemented. Pre-policy residency hours in CA, as reported in the residency program survey, mirror those at the national level. Prior to the reform, the representation of women across all specialties but particularly among time intensive specialties, was higher in CA than the national average, which in turn was higher than the female representation in TX. It is possible there was less scope for growth in female share in CA, attenuating the effects.

[46]The inclusion of these additional cohorts does not substantively alter the results.

[47]Individuals who started residency in 2001 or 2002 were partially exposed to the reform, since they were in residency at the time the reform was enacted. In equation (6), I combine this group with the control cohorts, who all finished their first three years of residency before the reform was enacted. The results are robust to separately analyzing cohorts 2001 and 2002.

I characterize the fertility choices of the marginal female entrant by contrasting the β_1 and β_2 coefficients. Specifically, I modify the approach outlined in Gruber et al. (1999), based on methods proposed by Berndt (1991) for estimating average and marginal cost curves. In Appendix C, I formalize this empirical strategy. In order to back out a characterization of the marginal female entrant's fertility from the estimates of β_1 and β_2, I require the stability of the treatment effect across the 2003-2005 and 2006-2010 cohorts. This rules out inframarginal high hours specialty women responding more in the latter period than in the former period to a given reduction in hours. If individuals do not internalize the duty hour reductions immediately in their decision-making or require planning in order to implement their fertility choices, then this assumption of stability across cohorts would be violated. Embedded in this assumption is the stability of the reduction in hours due to the reform across the two cohorts. From Figure 3, it appears there was rapid adjustment in hours within the first couple years of the reform, but accounts of lagged implementation make it possible that the reduction in hours was larger for the second cohort. These forces serve, if anything, to attenuate the estimate of the treatment effect.

5.2.2 Results

The CA fertility results are presented in Table 5 Panel A. The coefficient β_1 on the interaction ($\text{Hours}_{s,1999} \times \text{Transition}_t$) is reported in the first row, and the coefficient β_2 on the interaction of ($\text{Hours}_{s,1999} \times \text{Post}_t$) is reported in the second row. Starting with column 1 and β_1, holding the composition of a specialty constant, there is a positive and statistically significant effect of the reform on a specialty's fertility rate in CA. The results change little with the inclusion of additional individual control variables or when the sample is restricted to individuals without common last names to minimize the incidence of false positive matches. In order to assess whether the identifying assumption is plausible, I examine the trends in fertility among women in more and less time-intensive specialties in the years leading up to the reform. Appendix Figure A.6 plots the year-by-year coefficients from an event study specification. In CA (Panel A), there appears to be, if anything, a slightly negative trend in the years preceding the reform. Immediately after the reform is implemented, the coefficients increase substantially and are positive, but the positive effects are not sustained post-2007. This pattern is consistent with the decline in the regression coefficients between the transition and post periods. I additionally include specialty-specific linear time trends. The precision of the estimates drastically decreases, but the treatment effect remains positive and rises in magnitude.

To provide a sense for the magnitude of these effects, the β_1 coefficient from column 3 implies that the reform increases the number of children inframarginal female residents have during their first three years of residency by 0.00097. Scaling this coefficient by the bounds on the first stage relationship, a four hour per week reduction due to the reform corresponds to an increase of 0.007 to 0.02 children, amounting to a 5 to 15 percent rise over the average pre-reform level. Turning to the estimates of β_2, the combined effect is statistically indistinguishable from the treatment effect of the reform on a specialty's female fertility rate. If anything, the combined effect is smaller than the treatment effect, suggesting that the marginal female entrant is less likely to have children during residency than the average entrant.

The results for Texas are reported in Table 5 Panel B and are imprecisely estimated, but overall negative. The results are also more sensitive to the inclusion of specialty-specific time trends (not reported here). For Texas, their inclusion flips the sign of the coefficient. In addition, the event study reveals little movement in the relative fertility rates of more and less time-intensive specialties. Given the imprecise estimates, we should proceed with caution in our interpretation of the results. Recall that in the Transition_t cohorts in Texas, there is a positive effect of the reform on female specialty entry (though imprecisely estimated). This implies that β_1 is an estimate of the combined effect of the reform. In both CA and TX, the estimates of the treatment effect, β_1, are larger in magnitude than β_2, the combined effect of the reform (but both coefficients are positive in CA and negative in TX, and I cannot reject that the β_1 and β_2 are equal in CA and TX, respectively). If compositional changes from new entrants are indeed exerting a negative effect on a specialty's female fertility rate, and female specialty entry responds more in TX than in CA, then the negative composition effect could outweigh a positive treatment effect in TX.

5.3 Discussion

While there is a clear negative relationship between a specialty's pre-policy hours and the specialty's female fertility rate, the causal evidence on the effect of the reform on female fertility is more mixed. Using the CA estimates, it appears that relaxing work hours indeed increases the propensity of female residents to have children. This result aligns with the literature on women delaying the timing of family formation to accommodate career investments (Goldin and Katz, 2002; Bailey et al., 2012). Both the CA and the TX results suggest the new female entrants into time-intensive specialties are less likely than the average female entrant to have children during residency. Why might the marginal female entrant have lower fertility than the average female entrant? It is possible that the women induced to enter due to the reform are willing to forego having children during residency in order to access a higher wage specialty. Second, the new female entrants could have valued the additional time the reduced hours permitted to find a partner during residency and eventually have children.[48] A third explanation is the first three years of residency do not

[48] Qualitative surveys provide further insight on this decision-making process. 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 participating in a qualitative survey 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 and Bosk, 2012).

adequately capture all residency fertility, particularly in the more time-intensive specialties where residency lasts between five and seven years.[49]

6 Implications of the Reform for the Physician Gender Wage Gap

As discussed in the Introduction, Goldin (2014) posits that a key contributor to the remaining gender wage gap is the presence of convex returns to working long hours, and women's lower likelihood of reaping these returns, either through their choice of job or choice of hours within a job. 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 wage associated with each specialty in order to quantify the contribution of a reduction in hours during residency to the physician gender wage gap.

For this exercise, I use specialty-specific average hourly earnings from the Community Tracking Study, as reported in Leigh et al. (2010). One limitation of these data is the specialties covered are those that require direct patient care and therefore exclude Anesthesiology, Pathology, and Radiology. I simulate the gender pay gap before the introduction of the policy by computing the weighted average of specialty-specific pay, w_s, where the weights are the shares of men and women in each specialty pre-reform, $sh_{sg} = \frac{n_{sg}}{N_g}$: $\bar{w}_g = \sum_s w_s sh_{sg}$. According to this calculation, the pre-reform gender hourly earnings gap is \$5.18 (\$76.85 for men and \$71.67 for women, all in 2004 dollars) or 6.7 percent. This disparity is smaller than the gap in average salaries for male and female physicians reported elsewhere, likely due to the other estimates incorporating gender differences in choice of specialty, practice setting, and hours worked (Lo Sasso et al., 2011; Sasser, 2005). This measure, however, isolates the portion of physician gender pay gap that is due to hourly pay differences across specialties alone.[50]

Quantifying the contribution of the reform to average female (male) hourly earnings entails a sum of specialty-specific female (male) share changes due to the reform, weighted by specialty earnings: $\Delta\bar{w}_g = \sum_s w_s \Delta sh_{sg}$.[51] The back of the envelope calculation suggests that, through specialty selection, the reform

[49]Another possibility is the new female entrants could have preferences for early childbearing such that they had children during medical school and already have childcare responsibilities but do not plan to have children during residency.

[50]The gap remains a conservative estimate of hourly pay differences due to specialty choice, if hourly wages, conditional on specialty, vary by gender.

[51]Specialty-specific share changes due to the reform are the difference between the predicted specialty shares before and

will increase women's average hourly earnings by \$0.68. The change in male average hourly 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 wage gap by 13 percent.[52] 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 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 13 percent.

As the training and ramp-up periods for professional occupations tend to coincide with women's prime

after the reform, with specialty shares in each period normalized to add up to one. I assume that specialty-specific pay is time invariant.

[52] 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.

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 examining fertility choices during residency, descriptive and causal evidence are consistent with early career hours requirements posing a constraint on the timing of female fertility timing. The analysis characterizing the marginal female entrant, however, suggests that the women induced to enter a specialty due to the reduction in hours are are no more likely to have children during the first three years of residency training than the previous entrants.

Long, inflexible hours may serve a productive purpose, such as fostering skill acquisition, producing gains from the continuity of work, or ameliorating the imperfect substitutability of workers. This paper demonstrates, however, that the hours requirements ubiquitous among professional occupations have important implications for occupational segregation by gender. From 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), it is possible that long hours are an occupational friction restricting the optimal allocation of talent in the labor market.

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.

Bailey, Martha J, Brad Hershbein, and Amalia R Miller, "The Opt-In Revolution? Contraception and the Gender Gap in Wages," *American Economic Journal: Applied Economics*, jul 2012, *4* (3), 225–254.

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.

Berndt, Ernst, *The Practice of Econometrics: Classic and Contemporary*, Addison Wesley Publishing Company, 1991.

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.

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

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

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

Chen, Michelle M, Heather L Yeo, Sanziana a Roman, Richard H Bell, and Julie a Sosa, "Life Events During Surgical Residency Have Different Effects on Women and Men Over Time," *Surgery*, aug 2013, *154* (2), 162–70.

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.

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.

_ **and Lawrence F. Katz**, "The Power of the Pill: Oral Contraceptives and Womens Career and Marriage Decisions," *Journal of Political Economy*, 2002, *110* (4), 730–770.

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.

Gruber, Jonathan, Phillip Levine, and Douglas Staiger, "Abortion Legalization and Child Living Circumstances: Who is the "Marginal Child"?," *The Quarterly Journal of Economics*, 1999, *114* (1), 263–291.

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.

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

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.

Matthews, T J and Brady E Hamilton, "First Births to Older Women Continue to Rise," *NCHS Data Brief*, 2014, (152), 1–8.

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.

Park, Kyung and Nayoung Rim, "The Gendered Effects of Career Concerns on Fertility," *Working Paper*, 2017.

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.

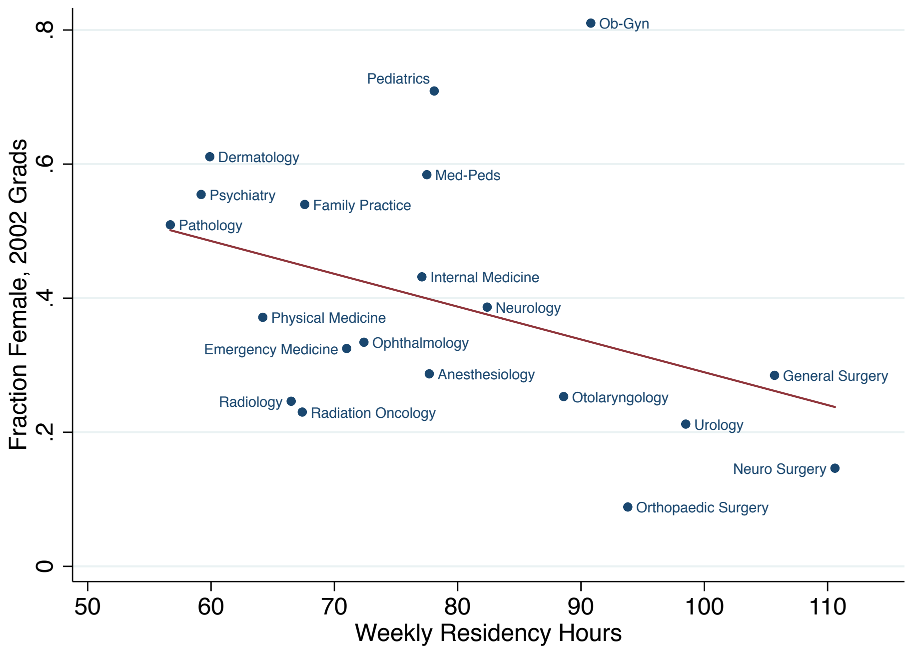
Volpp, Kevin G., Dylan S. Small, Patrick S. Romano, Kamal M F Itani, Amy K. Rosen, Orit Even-Shoshan, Yanli Wang, Lisa Bellini, Michael J. Halenar, Sophia Korovaichuk, Jingsan Zhu, and Jeffrey H. Silber, “Teaching Hospital Five-year Mortality Trends in the Wake of Duty Hour Reforms,” *Journal of General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

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

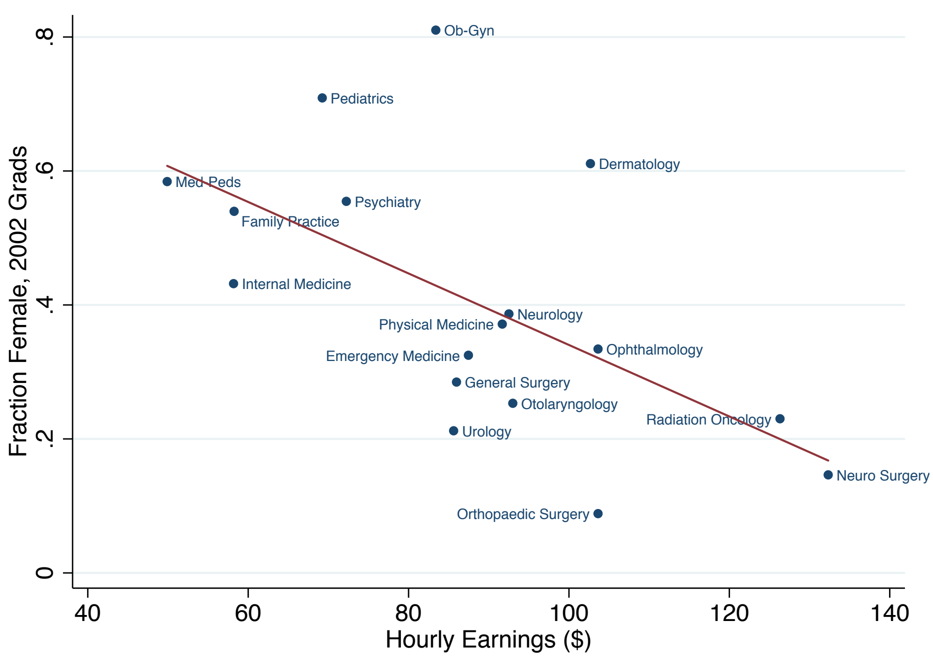
Yoon, Harry H., “Adapting to Duty-Hour Limits: Four Years On,” *New England Journal of Medicine*, 2007, *356*, 2668–2670.

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

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

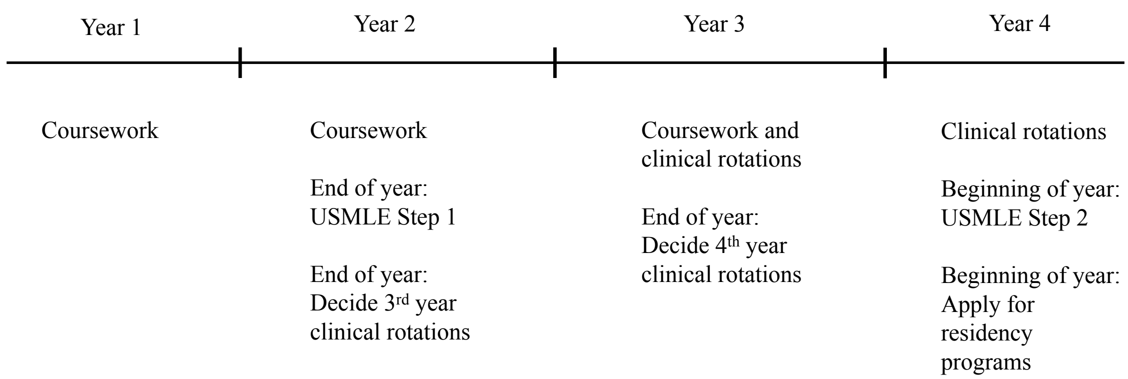


B. Female Share of Specialties and Average Earnings Post-residency



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

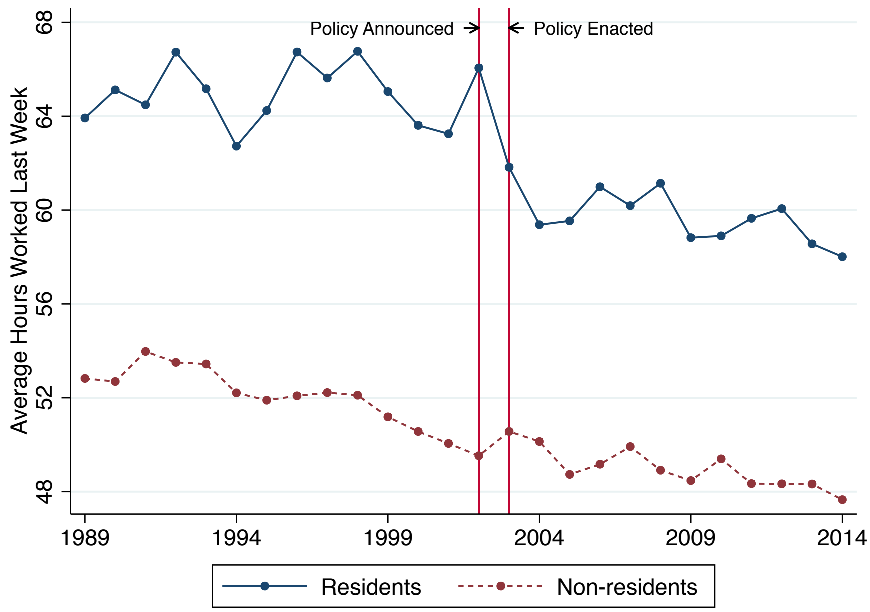
Figure 2: Medical School Timeline



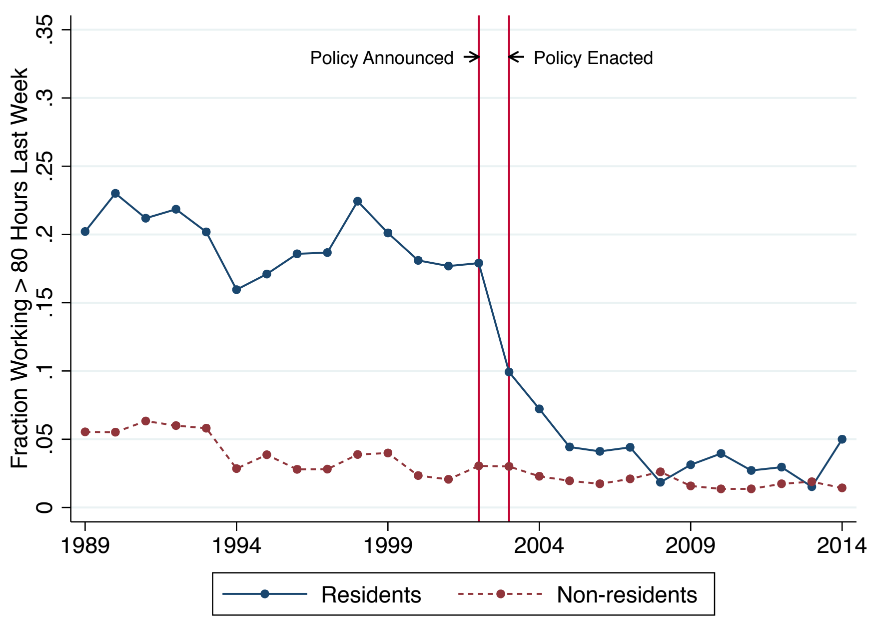
Source: Association of American Medical Colleges *Medical School Admission Requirements United States and Canada 2001-2002.*

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

A. Average Hours

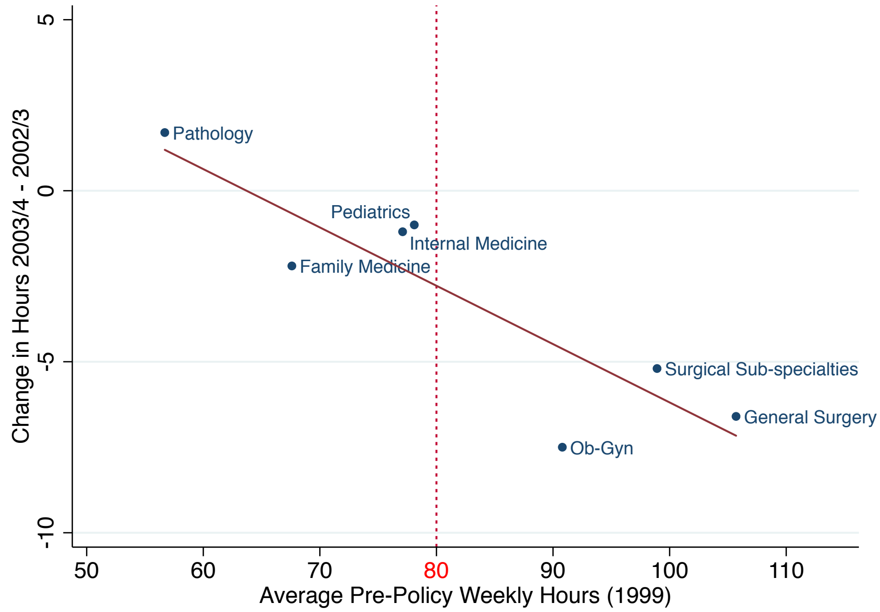


B. Fraction Working > 80 Hours



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

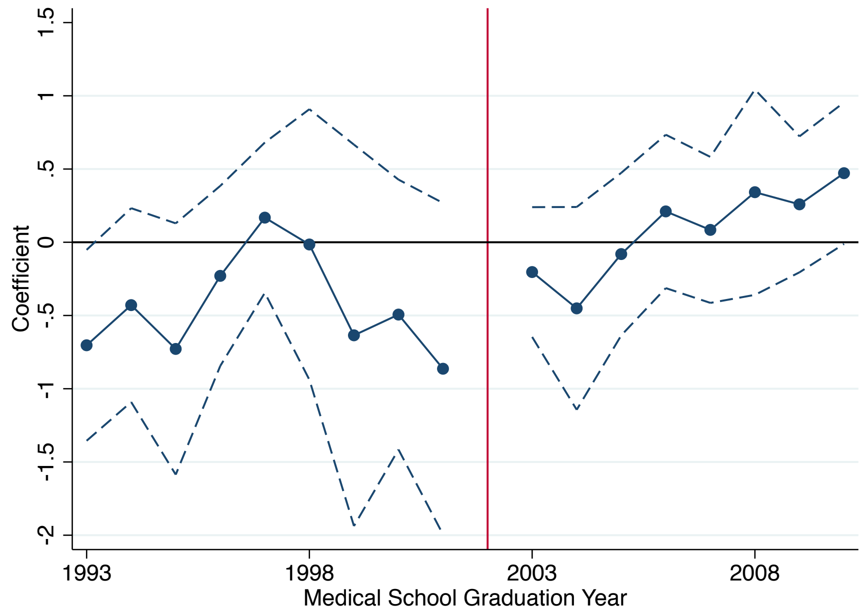
Figure 4: Relationship between 1999 Hours and the Change in Hours 2002/3-2003/4, by Specialty



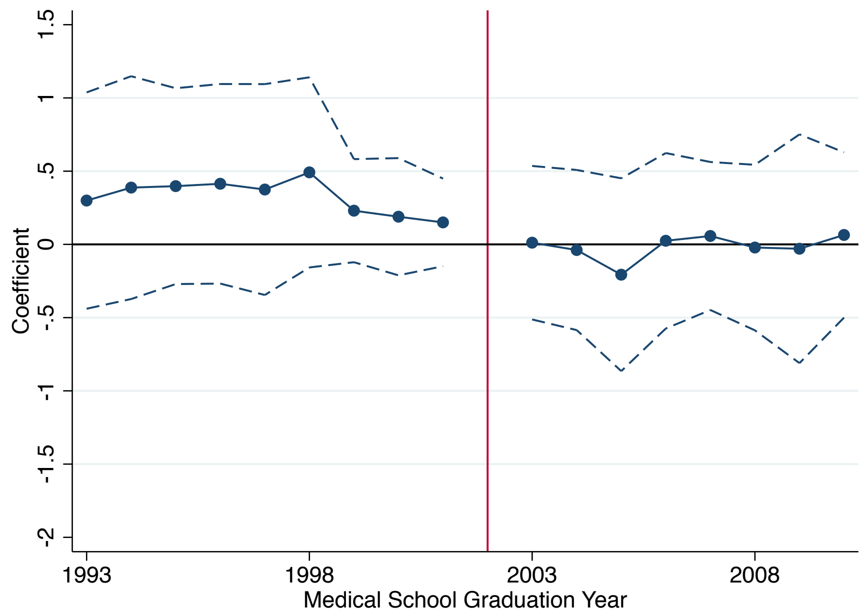
Source: Baldwin Jr et al. (2003), Landrigan et al. (2006) and personal correspondence with author. Note: This figure plots the average number of hours worked per week for second year medical resident physicians in 1999 on the x-axis. The change in average hours per week between residency years 2003/4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

Figure 5: The Effect of the Duty Hour Reform on Specialty Entry: Event Study Analysis

A. Women



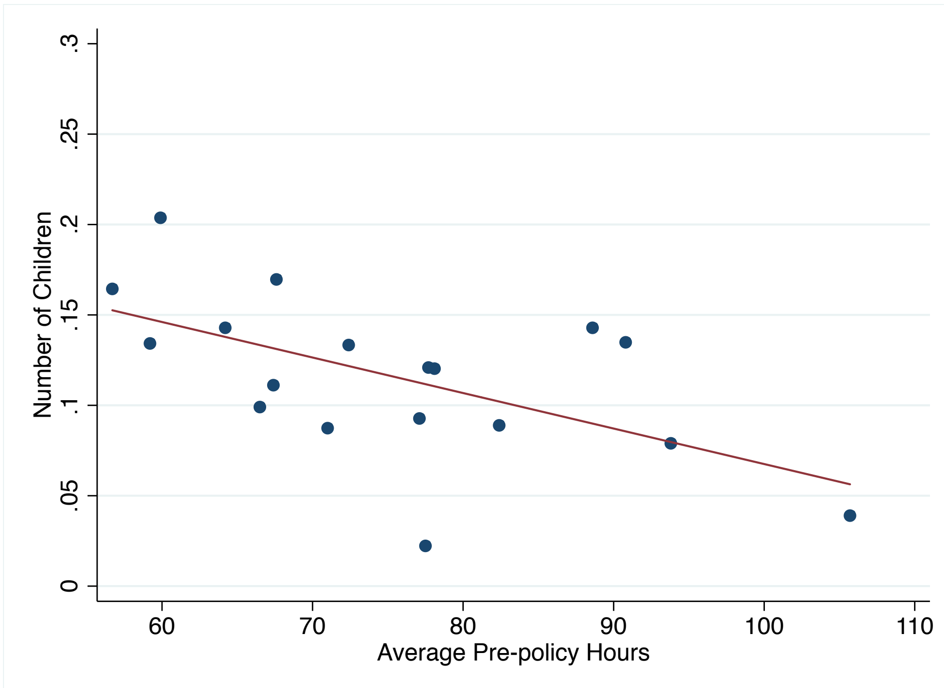
B. Men



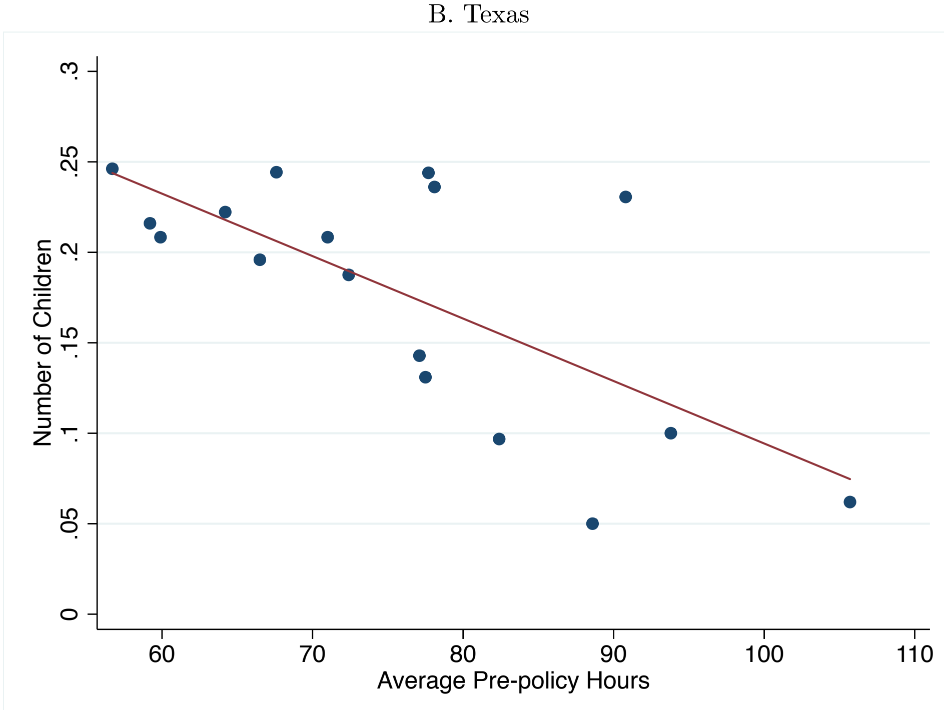
Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This figure plots the estimated coefficients from the event study model. The dependent variable is the natural logarithm of the share of women (men) from a given medical school graduation cohort in a specialty, multiplied by 100. The explanatory 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.

Figure 6: The Relationship between Pre-Policy Fertility and Pre-Policy Hours Requirements

A. California



B. Texas



Source: Source: AMA Physician Masterfile, California and Texas Vital Statistics birth records, Baldwin Jr et al. (2003) Note: This figure plots the mean number of children during the first three years of residency against the average pre-policy hours, by specialty. The CA (TX) sample includes female U.S. medical school graduates from years 1993 through 2010 who completed their first three years of residency training in CA (TX). Fertility during the first three years of residency is computed according to the typical residency year: July-June. For example, if an individual starts residency in 2001, then fertility during the first three years of residency is determined based on July 2001 - June 2004.

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

	(1)	(2)	(3)	(4)	(5)	(6)
	Full Sample			USMG Sample		
Panel A: National Samples	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
	California			Texas		
Panel B: State Samples	All	Female	Male	All	Female	Male
Female	0.47	-	-	0.43	-	-
Age at Medical School Graduation	28.33	28.26	28.39	28.10	27.86	28.29
	(3.05)	(3.11)	(2.99)	(3.30)	(3.23)	(3.34)
U.S. Born	0.73	0.74	0.72	0.82	0.81	0.82
Attended Ranked Medical School	0.68	0.69	0.67	0.47	0.47	0.47
Fertility During Residency						
Number of Children	-	0.13	-	-	0.20	-
Any Children	-	0.12	-	-	0.19	-
N	26,592	12,580	14,012	16,385	7,093	9,292

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. The California and Texas samples include USMGs who completed their first three years of residency in CA or TX, respectively. Number of children denotes the number of children individuals had during the first three years of residency. Medical school rank is determined by the inclusion of the medical school in U.S. News and World Report 2014 rankings. Standard deviations are reported in parentheses.

Table 2: 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 3: The Effect of the Duty Hour Reform on Stated Specialty Preference

Dependent Variable: $\ln(\text{Share}_{st})$*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 4: Residency Program Drivers of Female Entry

Dependent Variable: $FractionFemale_{pst}$	(1)	(2)	(3)	(4)
	Panel A: Female Representation			
	Faculty High	Faculty Low	Residents High	Residents Low
Average Weekly Hours × Transition	0.093***	0.019	0.059*	0.048
	(0.030)	(0.032)	(0.033)	(0.029)
Average Weekly Hours × Post	0.103***	0.046*	0.103***	0.044
	(0.030)	(0.027)	(0.029)	(0.027)
N	14,694	18,156	17,064	16,149
	Panel B: Family-Friendly Policies			
	Maternity Leave Policy	No Maternity Leave Policy	Onsite Childcare	No Onsite Childcare
Average Weekly Hours × Transition	0.060**	0.066	0.048	0.055*
	(0.025)	(0.044)	(0.039)	(0.028)
Average Weekly Hours × Post	0.076***	0.068*	0.076**	0.076***
	(0.024)	(0.039)	(0.035)	(0.026)
N	22,461	10,265	12,504	19,152

Source: AAMC GME Census Track, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-difference specification estimating the effect of the duty hour reform on residency program gender composition. The dependent variable is the fraction of first year residents who are female in a given residency program for a given start year, multiplied by 100. The explanatory variables include program fixed effects, residency entry year fixed effects and program pre-policy hours ($Hours_{sp,1996}$) interacted with an indicator for entering residency in 2003-2005 (Transition) and 2006-2010 (Post). Standard errors clustered at the program level are reported in parentheses.

Table 5: The Effect of the Duty Hour Reform on a Specialty's Female Fertility Rate During Residency

Dependent Variable: # children born during first three years of residency × 1000	(1)	(2)	(3)	(4)
Panel A: California				
Avg Weekly Hours × Transition	0.84*	0.90**	0.97**	1.08**
	(0.46)	(0.43)	(0.44)	(0.39)
Avg Weekly Hours × Post	0.64	0.78	0.77	0.94
	(0.99)	(0.95)	(0.89)	(0.93)
N	12,580	12,580	12,580	10,788
Panel B: Texas				
Avg Weekly Hours × Transition	-1.03	-0.98	-1.22	-0.79
	(1.04)	(1.04)	(1.10)	(1.07)
Avg Weekly Hours × Post	-1.48	-1.34	-1.58	-1.21
	(1.43)	(1.39)	(1.40)	(1.36)
N	7,093	7,093	7,093	6,468
Age at medical school graduation FE		X	X	X
Medical school FE			X	X
Exclude common names >50				X

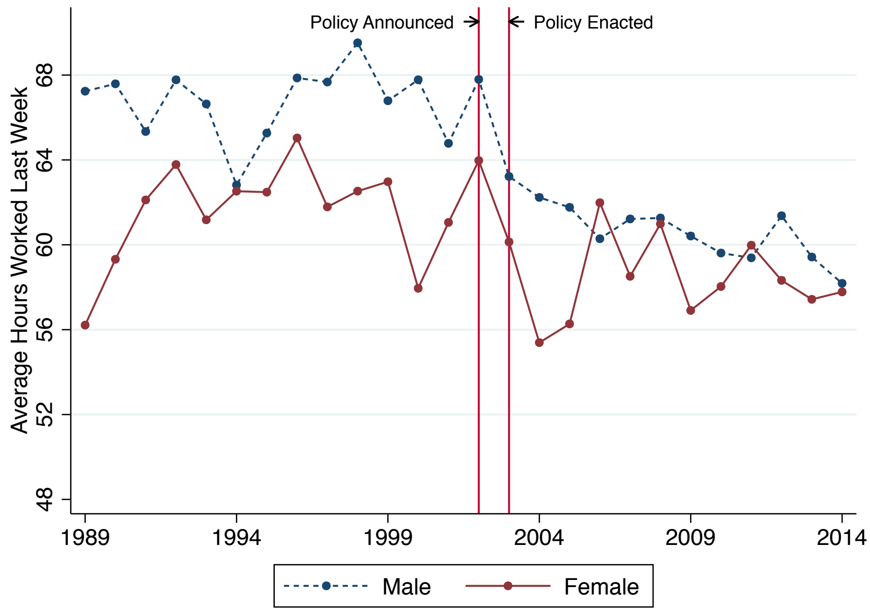
Source: AMA Physician Masterfile, California and Texas Vital Statistics birth records, Baldwin Jr et al. (2003). Note: This table reports the coefficients from a difference-in-difference regression of the number of children during the first three years of residency training on residency cohort fixed effects, specialty fixed effects, and the interaction of a specialty's hours during the second year of residency and an indicator variables for whether an individual started residency during the transition period after the reform (Transition: 2003-2005) or the after the transition period (Post: 2006-2010). The reference cohorts are individuals who started residency 1993-2002. Standard errors clustered at the specialty level are in parentheses. The sample includes individuals who did their first three years of residency in CA (TX) and who started residency training between 1993 and 2010. Omitted years are 1993-2000. Fertility during the first three years of residency is computed according to the typical residency year: July-June. For example, if an individual starts residency in 2001, then fertility during the first three years of residency is determined based on July 2001 - June 2004. Columns 1 through 3 report results from regressions with the progressive inclusion of covariates. In order to gauge the sensitivity of the results to the potential incidence of false positive matches between the CA (TX) Sample and CA (TX) Vital Statistics data, column 4 excludes from sample individuals with common last names, as defined by a frequency of 50+ in the CA (TX) Sample.

A Supplemental Figures and Tables

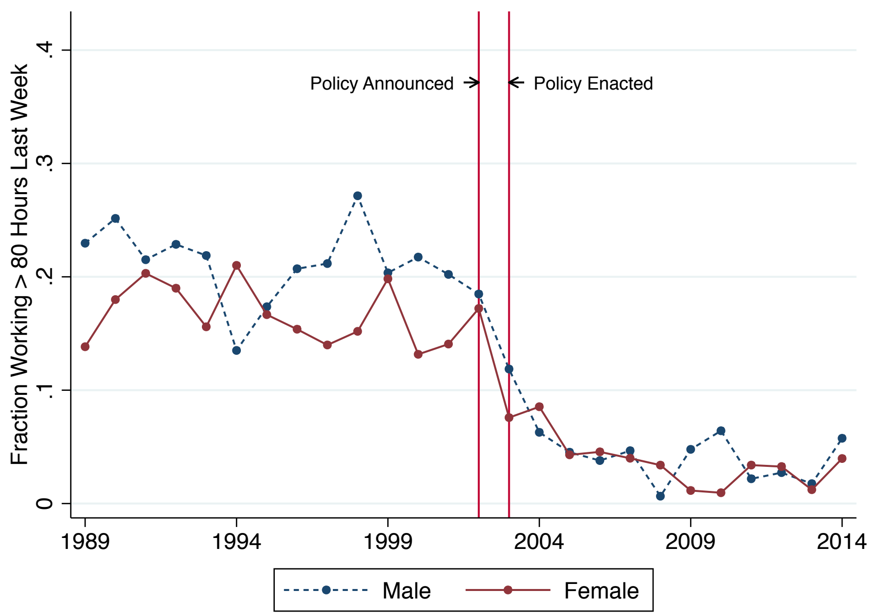
Figures

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

A. Average Hours



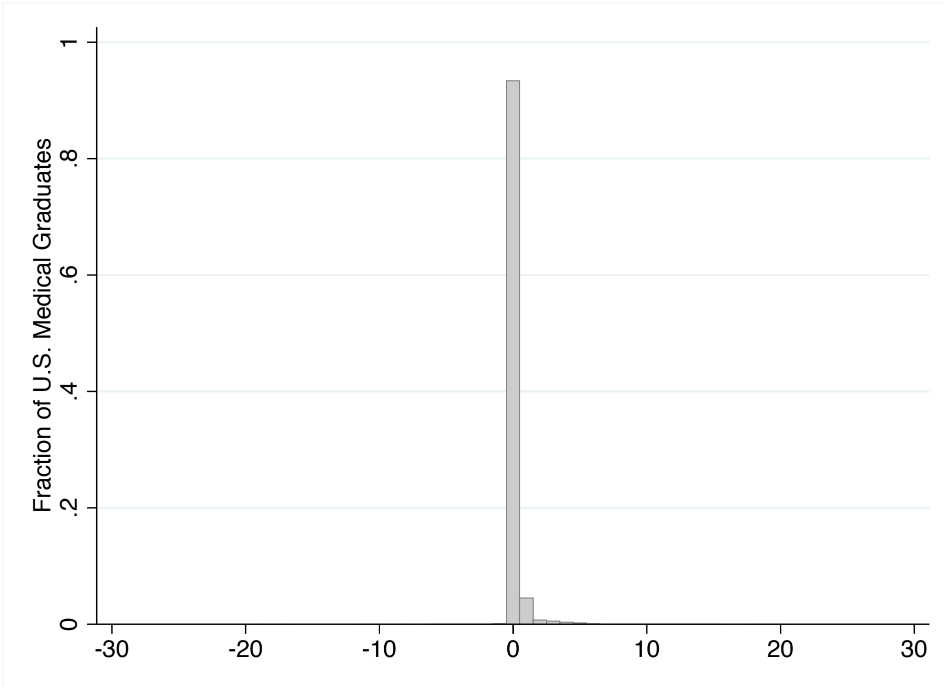
B. Fraction Working > 80 Hours



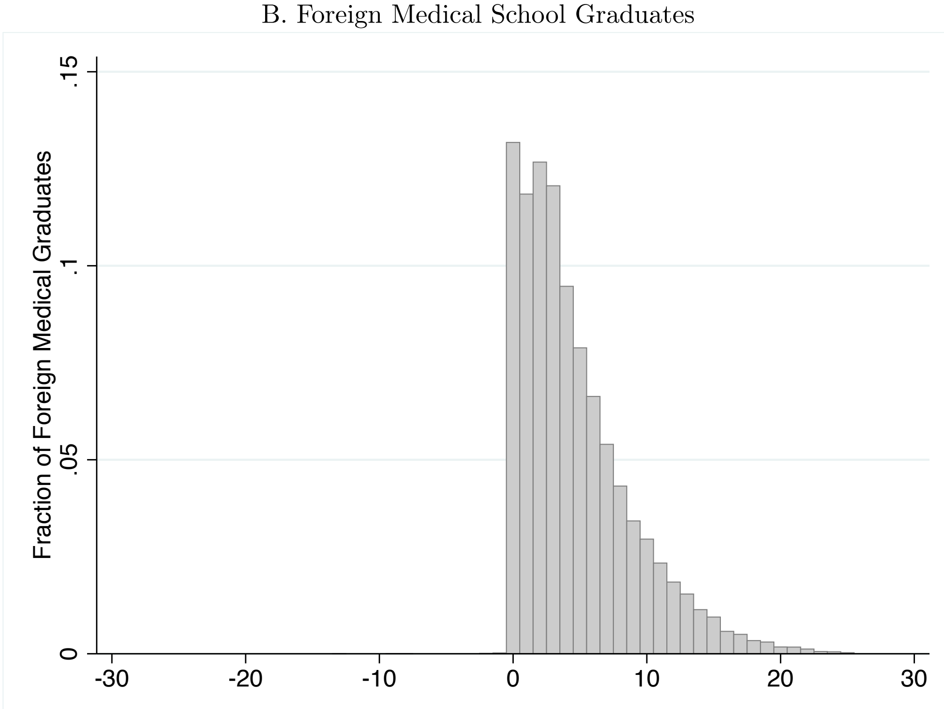
Source: Current Population Survey, monthly files January 1989-December 2014. Panel A plots the average number of hours worked last week for resident physicians, separately for men and women. Panel B plots the fraction of physicians who worked more than 80 hours last week, separately for men and women. 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



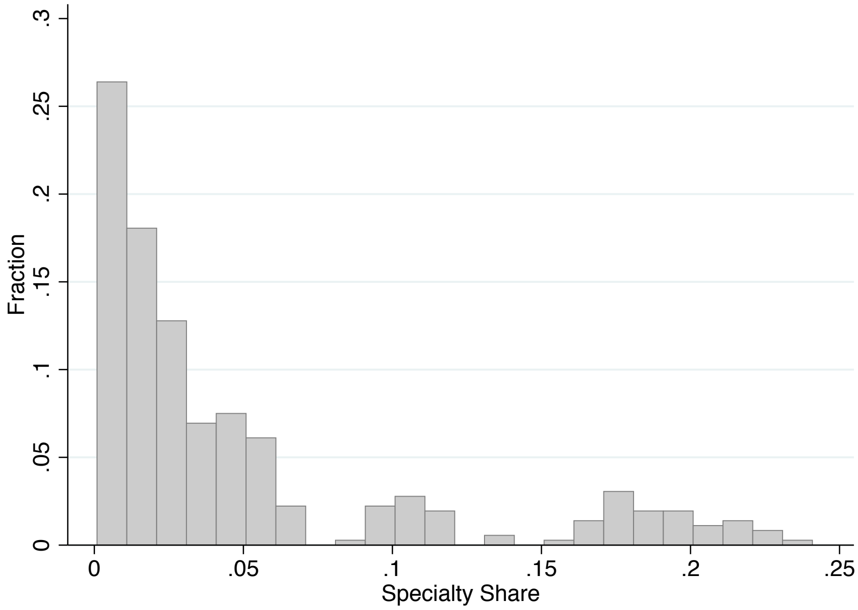
B. Foreign Medical School Graduates



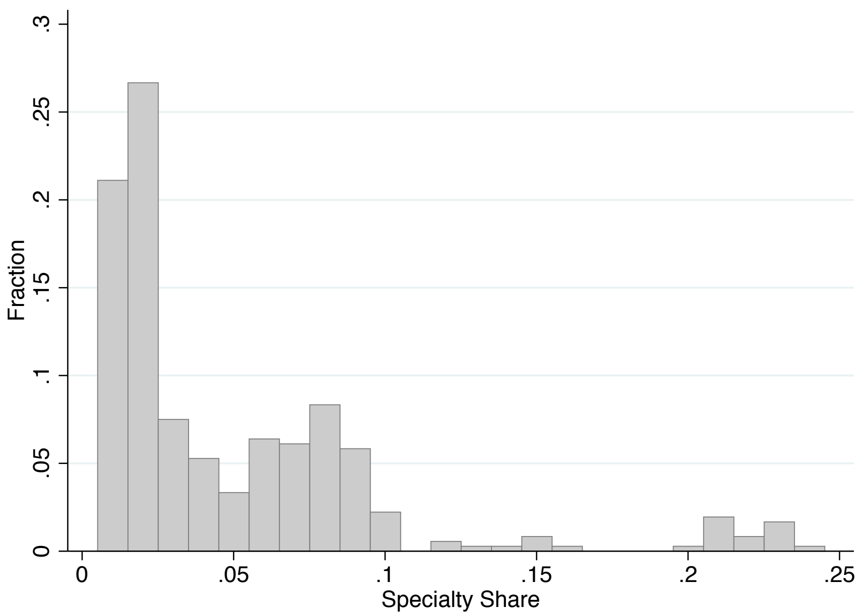
Source: AMA Physician Masterfile Note: This figure plots distribution of the difference between medical school graduation year and residency start year for individuals who graduated from U.S. medical schools and foreign medical schools and attended residency training in California or Texas.

Figure A.3: Distribution of Specialty Shares

A. Female



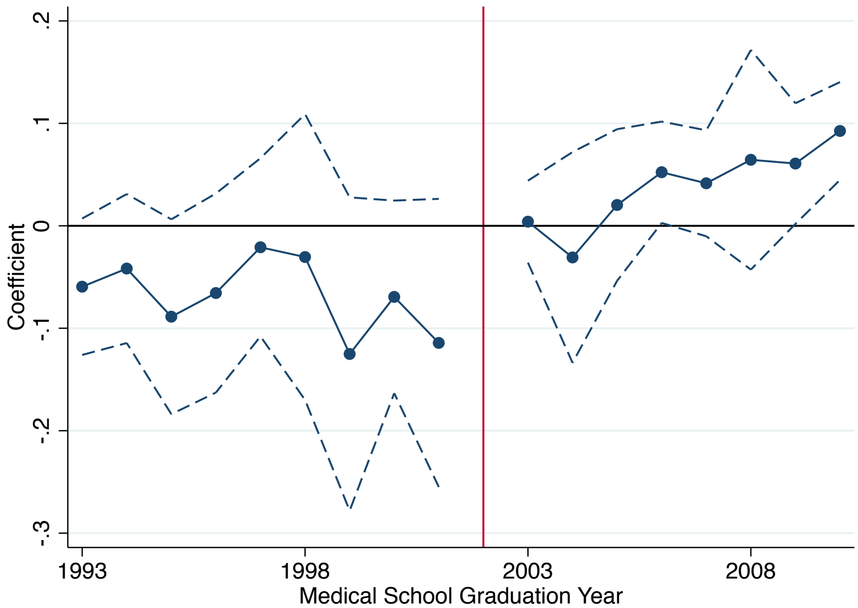
B. Male



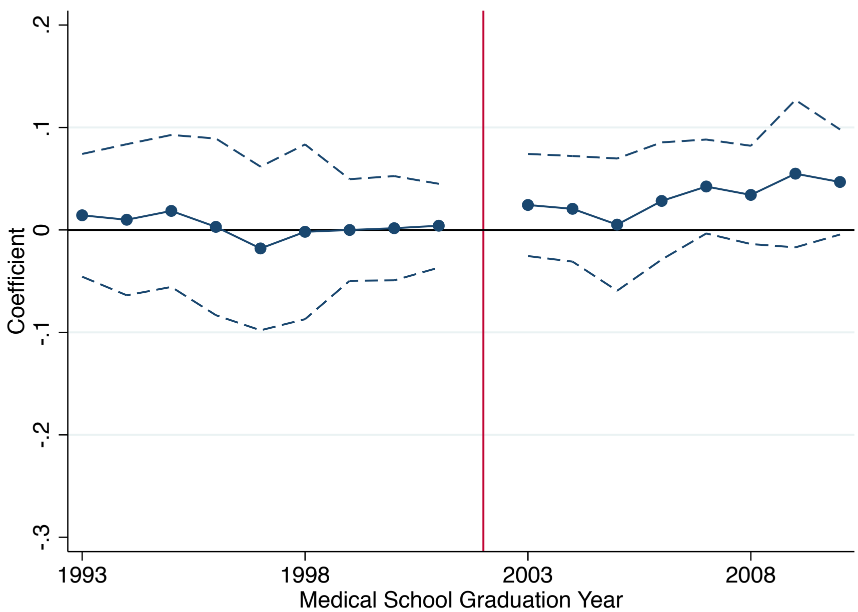
Source: AMA Physician Masterfile Note: This figure plots distribution of specialty shares for male and female U.S. medical school graduates who graduated from medical school 1993 to 2010.

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

A. Women



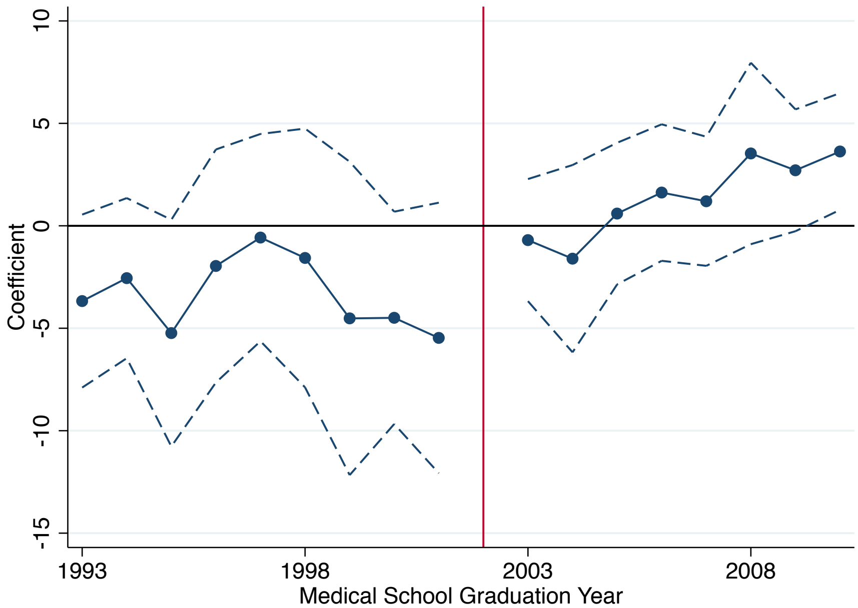
B. Men



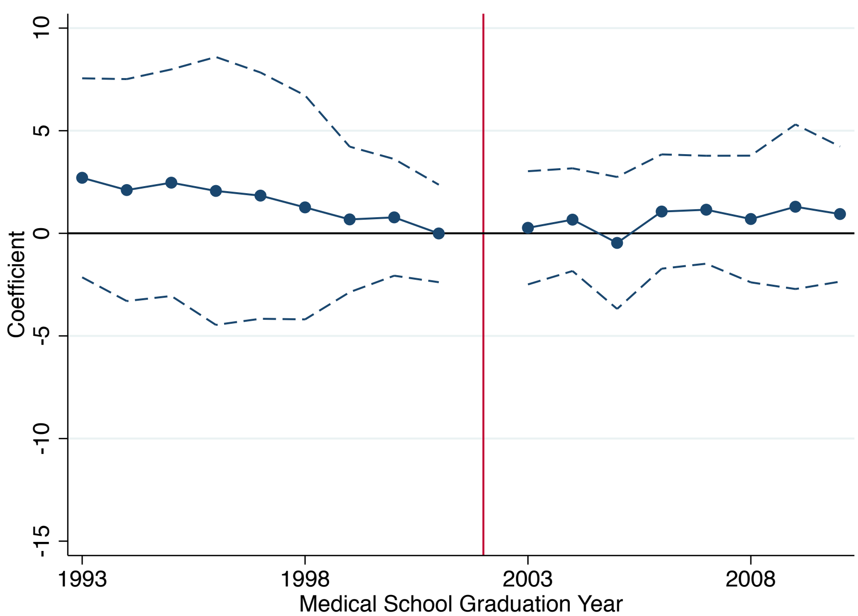
Source: AMA Physician Masterfile, Baldwin Jr et al. (2003), Freeman (2007). Note: This figure plots the estimated coefficients from the event study model with an alternative parameterization of pre-policy hours: the pre-policy total number of hours required by a specialty over the course of residency, 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.

Figure A.6: The Effect of the Duty Hour Reform on Female Fertility During Residency
Event Study

Source: AMA Physician Masterfile, California and Texas Vital Statistics birth records, Baldwin Jr et al. (2003). Note: This figure plots the results of an event study analysis of the effect of the duty hour reform on specialty fertility rates. The dependent variable is an individual's fertility during the first three years of residency, and the independent variables are residency start year fixed effects, specialty fixed effects and the interaction of specialty pre-policy hours with residency start year fixed effects. Residency cohort 2000 is omitted as the reference year. Standard errors are clustered at the specialty level. The solid line plots coefficients on the interaction of average pre-policy hours and residency cohort fixed effects. The dashed lines plot the 95% confidence intervals. The sample includes individuals who did their first three years of residency in CA (TX) and who started residency training between 1993 and 2010. Fertility during the first three years of residency is computed according to the typical residency year: July-June. For example, if an individual starts residency in 2001, then fertility during the first three years of residency is determined based on July 2001 - June 2004.

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: Pre-Policy Distribution of Men and Women across Specialties

Specialty	(1) Avg Pre-Policy Weekly Hours	(2) Male Share (%)	(3) Female Share (%)
Pathology	56.7	1.8	2.4
Psychiatry	59.2	3.9	5.5
Dermatology	59.9	1.0	1.8
Physical medicine/rehab	64.2	1.5	1.1
Radiology	66.5	5.4	2.8
Radiation oncology	67.4	0.6	0.4
Family practice	67.6	14.5	18.3
Emergency medicine	71.0	6.1	3.3
Ophthalmology	72.4	2.3	1.5
Internal medicine	77.1	29.4	27.5
Internal medicine/peds	77.5	1.0	1.4
Anesthesiology	77.7	5.4	2.9
Pediatrics	78.1	6.1	16.2
Neurology	82.4	1.8	1.6
Otolaryngology	88.6	1.5	0.5
Obstetrics/Gynecology	90.8	2.9	8.6
Orthopedic surgery	93.8	4.3	0.5
Urology	98.5	1.3	0.3
General surgery	105.7	8.2	3.3
Neurological surgery	110.6	1.0	0.2

Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This table reports the distribution of men and women across medical specialties for the pre-policy (1993-2002) cohorts of medical school graduates.

Table A.3: 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.4: 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.5: The Effect of the Duty Hour Reform on Specialty Entry: Inclusion of Foreign and Osteopathic Medical School Graduates

Dependent Variable: $\ln(\text{Share}_{st})$*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.

Table A.6: Characterizing the Quality of the Marginal Entrant

Dependent variable: Share of individuals in a specialty who attended ranked medical school	(1) All	(2) Men	(3) Women
Average Weekly Hours × Transition	-0.117**	-0.152***	-0.078
	(0.048)	(0.051)	(0.083)
Average Weekly Hours × Post	-0.084*	-0.079	-0.119**
	(0.040)	(0.058)	(0.050)
Mean Dependent Variable	49.46	48.98	50.66

Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-difference specification, estimated separately for men and women, on the sample of U.S. medical school graduates. The dependent variable fraction of individuals in specialty s from medical school cohort t who attended a ranked medical school, multiplied by 100. The explanatory variables include specialty fixed effects, medical school graduation cohort fixed effects and specialty pre-policy hours ($\text{Hours}_{s,1999}$) interacted with an indicator for graduating medical school 2003-2005 (Transition) and 2006-2010 (Post). All specifications have 360 observations stemming from the analysis of 20 specialties over 18 years. Column 1 reports the results for the entire sample. Columns 2 and 3 report the results for the samples of men and women, respectively. Standard errors clustered at the specialty level are reported in parentheses.

Table A.7: Residency Program Summary Statistics

	1996-2010		1996				
	# Programs	Fraction Female among First Year Residents	# Programs	Fraction Female among First Year Residents	Fraction Female among Full-time Faculty	Any Maternity Leave Policy	Onsite Childcare
All Programs	34,253	0.44 (0.27)	2,469	0.38 (0.26)	0.22 (0.16)	0.69	0.39
By Specialty							
Anesthesiology	1,521	0.31 (0.18)	112	0.28 (0.20)	0.23 (0.10)	0.64	0.43
Dermatology	809	0.59 (0.28)	54	0.51 (0.28)	0.28 (0.18)	0.71	0.43
Emergency medicine	1,547	0.34 (0.17)	105	0.28 (0.14)	0.20 (0.12)	0.64	0.37
Family practice	5,281	0.50 (0.22)	387	0.45 (0.21)	0.26 (0.18)	0.66	0.34
General surgery	3,143	0.27 (0.19)	236	0.19 (0.17)	0.09 (0.12)	0.68	0.41
Internal medicine	5,159	0.41 (0.15)	364	0.36 (0.16)	0.21 (0.12)	0.78	0.39
Neurological surgery	155	0.11 (0.22)	12	0.03 (0.07)	0.02 (0.05)	0.60	0.27
Neurology	783	0.42 (0.26)	57	0.31 (0.26)	0.16 (0.11)	0.66	0.45
Obstetrics/Gynecology	3,316	0.74 (0.23)	235	0.65 (0.25)	0.27 (0.15)	0.66	0.36
Ophthalmology	1,127	0.35 (0.26)	88	0.30 (0.26)	0.15 (0.13)	0.79	0.30
Orthopedic surgery	1,658	0.10 (0.16)	110	0.07 (0.13)	0.05 (0.11)	0.56	0.40
Otolaryngology	600	0.25 (0.26)	43	0.22 (0.24)	0.09 (0.09)	0.62	0.40
Pathology	1,177	0.50 (0.25)	89	0.40 (0.24)	0.26 (0.13)	0.67	0.48
Pediatrics	2,736	0.68 (0.17)	193	0.61 (0.17)	0.37 (0.12)	0.68	0.42
Physical medicine/rehab	538	0.39 (0.25)	39	0.36 (0.24)	0.34 (0.13)	0.71	0.40
Psychiatry	2,059	0.52 (0.21)	145	0.47 (0.22)	0.27 (0.11)	0.73	0.41
Radiation oncology	137	0.32 (0.29)	14	0.29 (0.22)	0.24 (0.14)	0.58	0.50
Radiology	2,042	0.28 (0.21)	149	0.25 (0.22)	0.21 (0.12)	0.74	0.43
Urology	465	0.18 (0.25)	37	0.09 (0.15)	0.02 (0.05)	0.51	0.38
N	34,253	34,253	2,469	2,469	2,439	2,364	2,292

Source: AAMC GME Census Track. Note: This table reports summary statistics for the sample of residency programs used for the analysis of the effect of the duty hour reform on program gender composition. The data are comprised of an unbalanced panel of residency programs 1996-2010 for 20 broad specialties. Fellowship programs and small programs (those with one or fewer residents in any survey year) are dropped.

Table A.8: The Effect of the Duty Hour Reform on Residency Program Gender Composition

Dependent Variable: $\text{FractionFemale}_{pst}$	(1)	(2)	(3)	(4)
Panel A: Program-level hours				
Average Weekly Hours × Transition	0.048**	0.046**	0.050**	0.031
	(0.022)	(0.022)	(0.023)	(0.023)
Average Weekly Hours × Post	0.068***	0.064***	0.073***	0.037
	(0.020)	(0.021)	(0.021)	(0.024)
N	34,253	34,253	31,322	34,253
Panel B: Specialty-level hours				
Average Weekly Hours × Transition	0.045	0.033	0.045	0.010
	(0.026)	(0.028)	(0.029)	(0.043)
Average Weekly Hours × Post	0.077**	0.057	0.083**	0.014
	(0.034)	(0.042)	(0.032)	(0.059)
N	34,253	34,253	31,322	34,253
Ob/Gyn time trend		X		
Primary care specialty time trend		X		
Program baseline characteristics time trend			X	
All specialty time trends				X

Source: AAMC GME Census Track, Baldwin Jr et al. (2003). Note: This table reports the results of the difference-in-difference specification estimating the effect of the duty hour reform on residency program gender composition. The dependent variable is the fraction of first year residents who are female in a given residency program for a given start year, multiplied by 100. In Panel A, the explanatory variables include program fixed effects, residency entry year fixed effects and program pre-policy hours ($\text{Hours}_{sp,1996}$) interacted with an indicator for entering residency in 2003-2005 (Transition) and 2006-2010 (Post). In Panel B, programs are assigned the pre-policy hours associated with their specialty from the Baldwin Jr et al. (2003) survey. Column 1 reports the results of the baseline specification with no additional controls. Columns 2-6 progressively include specialty- and program-specific controls. Standard errors, reported in parentheses, are clustered at the program level in Panel A and specialty level in Panel B.

Table A.9: The Effect of the Duty Hour Reform on Specialty Entry: CA and TX

Dependent Variable: ln($Share_{st}$)*100	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All		CA and TX		CA		TX	
Panel A: Female								
Average Weekly Hours × Transition	0.15	0.25	-0.07	-0.10	-0.20	-0.26	0.25	0.27
	(0.29)	(0.23)	(0.55)	(0.54)	(0.46)	(0.43)	(0.66)	(0.64)
Average Weekly Hours × Post	0.67**	0.77***	0.54	0.57	0.35	0.38	0.97	0.97
	(0.30)	(0.20)	(0.41)	(0.39)	(0.43)	(0.40)	(0.68)	(0.66)
Panel B: Male								
Average Weekly Hours × Transition	-0.37	-0.19	-0.16	-0.07	0.18	0.23	-0.60*	-0.55**
	(0.36)	(0.23)	(0.32)	(0.20)	(0.38)	(0.29)	(0.33)	(0.25)
Average Weekly Hours × Post	-0.27	-0.05	-0.01	0.13	0.28	0.39	-0.49	-0.37
	(0.41)	(0.25)	(0.48)	(0.37)	(0.46)	(0.33)	(0.61)	(0.48)
Ob/Gyn time trend		X		X		X		X
Primary care specialty time trend		X		X		X		X
Age/medical school controls		X		X		X		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 various samples of U.S. medical school graduates (USMGs): all USMGs, the CA/TX sample, the CA sample, and the TX sample. 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. For each sample, the first column reports the results of the baseline specification with no additional controls. The second column includes time trends for primary care specialties and Ob/Gyn, and controls for the age and medical school composition of specialties. Standard errors clustered at the specialty level are reported in parentheses.

B Conceptual Framework

Suppose that the utility of physician i of gender g in specialty s with child c depends on hours worked during residency, wages post residency, and the presence of children. I formally represent the decision as a static, unconstrained utility maximization problem. The functional form is as follows:

$$u_{igsk} = \begin{cases} -\beta_i h_s + w_s & \text{if } c = 0 \\ -\beta_i h_s + w_s + \pi_i - \mathbb{1}\{g = f\}\phi(h_s) & \text{if } c = 1 \end{cases}$$

Specialties are considered bundles of attributes, the focus of which are hours worked during residency and wages post-residency: (h_s, w_s).[53] This utility specification embeds a few key components. First, there is individual heterogeneity in the relative valuation of non-market time and wages, captured by β_i, with $\beta_i \sim [0, b]$. Second, there is individual heterogeneity in valuation of children, π_i, where $\pi_i \sim [-p, p]$, if they have a child. Since π_i can take on negative values, some individuals derive disutility from having a child during residency and will not have children during this period. Third, there is an additional disutility of hours worked in the event a woman has a child: $\phi(h_s)$, where $\phi(h_s) > 0$, $\phi'(h_s) > 0$ and $\phi''(h_s) > 0$. I focus on the choice of two specialties: H and L, where H is a high hours, high wage specialty (h_H, w_H) and L is a low hours, low wage specialty (h_L, w_L). Assume $w_H > w_L$ and $h_L < h_H$.[54]

An individual maximizes her utility, i.e. chooses the specialty H or L and makes the choice whether to have children during residency, that is associated with the highest utility level. There are four options: $\{HC, HN, LC, LN\}$. For men, specialty choice is independent of the decision to have children. The choice to enter a high hours specialty is determined by whether β_i is greater than the cutoff $\beta^* = \frac{w_H - w_L}{h_H - h_L}$. Men will have children as long as it is utility enhancing, i.e. their value of π_i is greater than zero. For women, conditional on the choice to have children, specialty choice is determined by where an individual's β_i falls relative to two cutoffs, one associated with no children β^*_N and one associated with having children β^*_C.[55] Higher values of β reflect a greater disutility (dislike) of hours, and all else equal, make individuals weakly more likely to enter the low hours specialty. The choice to have a child, conditional on specialty, is dictated by where an individual's π_i falls relative to two cutoffs, one for the high hours specialty π^*_H and one for the low hours specialty π^*_L.[56] Higher values of π reflect a higher valuation of children and, all else equal, make individuals weakly more likely to have children.

[53] To simplify the exposition, I focus on one occupational attribute of a specialty post-residency, its wages, but w_s could also encompass a specialty's prestige, practice style, etc.

[54] The functional form for utility omits income effects in fertility choices.

[55] The β^*_N cutoff is determined by: $\beta^*_N = \frac{w_H - w_L}{h_H - h_L}$ The β^*_C cutoff is determined by: $\beta^*_C = \frac{w_H - w_L}{h_H - h_L} + \frac{\phi(h_L) - \phi(h_H)}{h_H - h_L}$, where $\beta^*_C < \beta^*_N$.

[56] The π^*_L cutoff for the low hours specialty case is determined by: $\pi^*_L = \phi(h_L)$ The π^*_H cutoff for the high hours specialty case is determined by: $\pi^*_H = \phi(h_H)$, where $\pi^*_L < \pi^*_H$.

For women, the joint choice of specialty and to have children during residency depends on their parameter values relative to the four cutoffs and is graphically summarized in Figure B.1 Panel A. There are three scenarios for β_i. For individuals with $\beta_i > \beta_N^*$, their disutility of hours is sufficiently high that they will always choose L. Depending on the value of π_i, an individual chooses to have a child or not during residency. For individuals with $\beta_i < \beta_C^*$, their disutility of hours is so low that they always choose the high hours specialty. Depending on the value of π_i an individual chooses whether to have children. For individuals with $\beta_C^* < \beta_i < \beta_N^*$, the disutility of hours is such that she either chooses L and has children, or chooses H and doesn't have children. Conditional on β and π, men are more likely to have children during residency than women, due to $\phi(h_s)$, the additional disutility of hours when a woman has children (Chen et al., 2013). If men and women have the same distributions of parameter values, then among those who choose the high hours specialty, a greater fraction of men than women will choose to have children during residency.

What happens when the duty hour reform goes into effect? This manifests as a decrease in hours in the high hours specialty, h_H. For men, the reform causes a decline in β^*, which induces entry into H. The decline in hours does not change male fertility choices, as these choices are determined independently of hours worked. The implications of the reform for women are graphically depicted in Figure B.1 Panel B. The decrease in h_H causes in a decline in π_H^*, the cutoff for having children in H. Thus, there is an expansion in the HC region, as individuals with lower child valuations now want to have children in the high hours specialty. I denote the individuals who change their fertility within H "fertility compliers." The reduction in h_H additionally causes an increase in the β cutoffs but leaves the π_L^* cutoff unchanged.

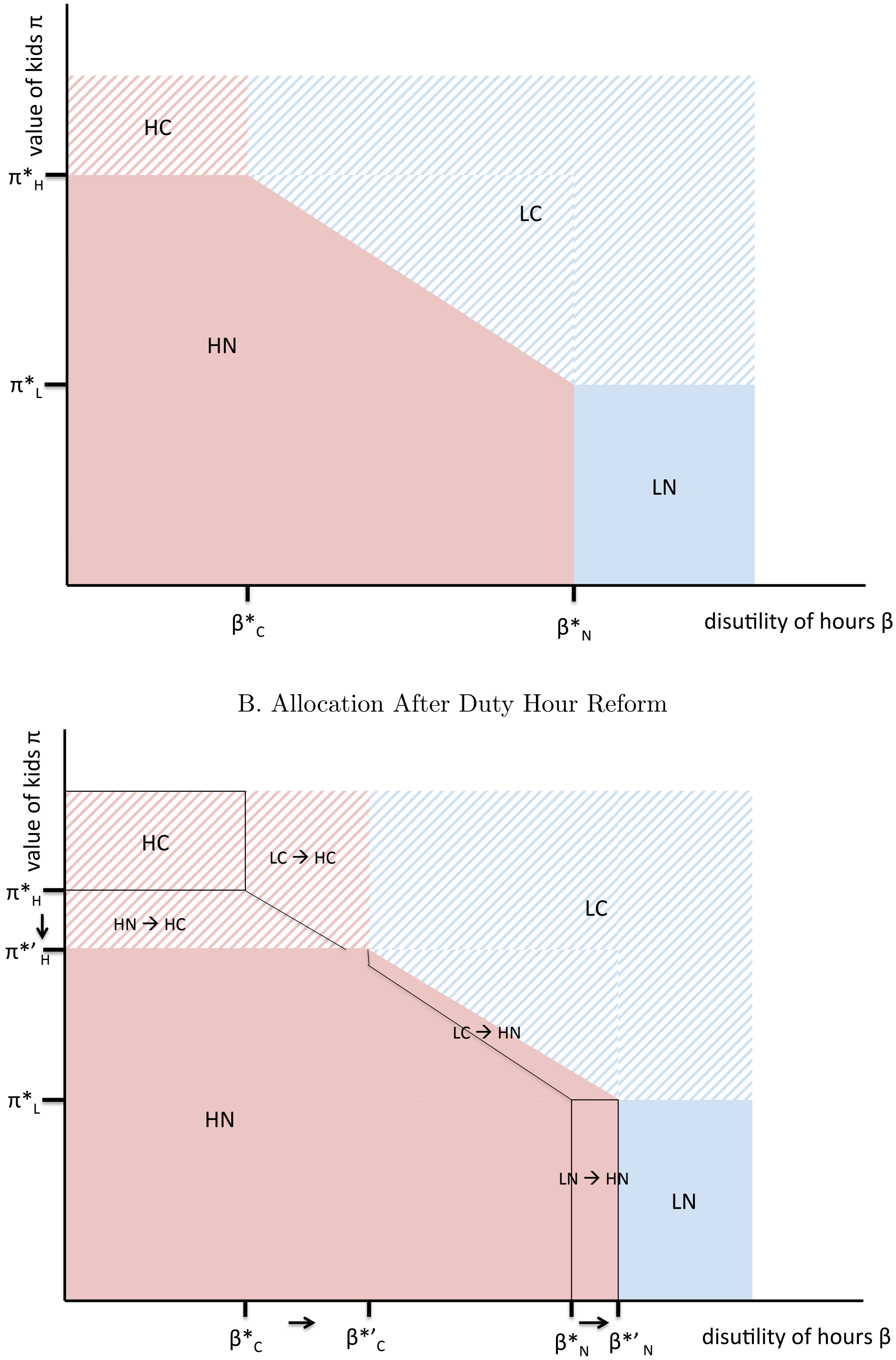
A few key insights emerge from this simple framework. First, the decrease in h_H induces net entry into H due to the shift upward of the β cutoffs, which serve primarily to expand the population of individuals willing to enter H specialties, from both LC and LN. I denote these individuals who change specialties, but do not change their fertility choice, "specialty compliers with children" ($LC \rightarrow HC$) and "specialty compliers without children" ($LN \rightarrow HN$). Second, there is an expansion of the HC region, due to the fertility compliers and specialty compliers with children. Third, some individuals with intermediate values of β and π, who would have chosen L and had children, now choose H and do not have children. I denote this group the "fertility-marginal specialty compliers" ($LC \rightarrow HN$). The intuition behind this switch is the wage advantage in H now outweighs the disutility of hours difference between H and L (which has fallen) and the utility of having a child. Since for women, the disutility of hours is convex when they have children, the β_C^* cutoff rises by more than the β_N^* cutoff in response to the reduction in hours, potentially inducing more women than men to enter H.

The final insight concerns the direction of the change in the female fertility *rate* in H and L. Given the above discussion, the direction of the change depends on the magnitude and composition of the new entrants to H, the magnitude and composition of those individuals who exit L, and the magnitude of the fertility compliers. In H, there are two potentially offsetting effects: the fertility compliers, who increase the fertility rate, and new entrants, who could increase or decrease the fertility rate. Under the assumption of independently and uniformly distributed β and π, the fertility rate rises in H. Under other distributional assumptions, the fertility rate in H can increase or decrease. For example, consider the case in which the valuation of children and disutility of hours are highly positively correlated. In this scenario, the HC and LN regions of Figure B.1 Panel A are sparsely populated. The introduction of the duty hour reform would induce individuals to enter H who are disproportionately drawn from the fertility-marginal specialty compliers. If this is the case, then the fertility rate in H could stay the same or even fall in response to the reform.

Figure B.1: Graphical Depiction of Conceptual Framework

A. Example of Initial Allocation

B. Allocation After Duty Hour Reform



Note: This figure presents a graphical depiction of an allocation of individuals into specialty and fertility choices, based on their parameter values. An individual's disutility of hours β is plotted on the x-axis, and an individual's valuation of children π is plotted on the y-axis. LC represents the choice of low hours specialty and to have children, LN represents the choice of low hours specialty and no children, HC represents the choice of high hours specialty and to have children, HN represents the choice of high hours specialty and no children.

C Empirical Strategy for Characterizing Fertility of Marginal Entrant

In this Appendix, I formalize the empirical strategy for separating the direct from the compositional effects of the reform on a specialty's female fertility rate. Consider the fertility rate during residency for women in specialty s who started residency in year t, denoted by $F_{st} = \frac{K_{st}}{R_{st}}$, where K_{st} represents the number of children born to the female residents R_{st}. Suppose that the number of women in a given specialty and cohort is a function of h_{st}, the hours of work required in the specialty for that residency cohort, $R_{st} = r(h_{st})$. Furthermore, suppose that the number of children born to women is a function of the number of women and the number of hours worked, $K_{st} = k(r(h_{st}), h_{st})$. As the expression is defined, hours requirements may affect the number of children indirectly through the number of women $r(h_{st})$ and directly through the second term. The effect of a change in hours on the fertility rate is the total derivative of the fertility rate with respect to hours (dropping subscripts for simplicity):

$$
\begin{aligned}
\frac{d\frac{k(r(h),h)}{r(h)}}{dh} &= \frac{\left[\frac{\partial k(r(h),h)}{\partial r(h)} \cdot \frac{dr(h)}{dh} + \frac{\partial k(r(h),h)}{\partial h}\right] \cdot r(h) - \frac{dr(h)}{dh} \cdot k(r(h),h)}{[r(h)]^2} \\
&= \underbrace{\frac{\partial \frac{k(r(h),h)}{r(h)}}{\partial h}}_{\text{fert. compliers}} + \underbrace{\frac{d \ln r(h)}{dh}}_{\text{specialty compliers}} \cdot \underbrace{\left[\frac{\partial k(r(h),h)}{\partial r(h)} - \frac{k(r(h),h)}{r(h)}\right]}_{\text{marginal - average fertility}}
\end{aligned} \tag{7}
$$

The final expression has three components, each of which is of independent interest. The first term is the fertility complier effect: the change in fertility with respect to a change in hours worked, holding the composition $r(h)$ of a specialty constant. This is the effect of labor supply on fertility, which, to my knowledge, has not previously been estimated with use of a natural experiment. The second term quantifies the specialty complier effect, that is, the effect of a reduction in hours on female entry into a specialty. The third term represents the difference between the marginal entrant's and average entrant's fertility. Note that the marginal entrant group is comprised of the fertility-marginal specialty compliers, the specialty compliers with children, and the specialty compliers without children. As discussed in the conceptual framework, under the assumption of uniformly distributed types, the fertility rate of the marginal entrant should be larger than the fertility rate of the average entrant. Under other distributional assumptions, the fertility rate of the marginal entrant could be higher or lower than the fertility rate of the average entrant.

By rearranging the final expression from equation (7), the fertility of the marginal entrant relative to the average entrant is: $MF = \frac{\beta_2 - \beta_1}{\beta_{entry}}$, where β_{entry} is the estimate of the reform's effect on female entry into more time-intensive specialties from Section 4.[57] To provide intuition for this expression, the numerator partials out the fertility complier effect from the combined fertility effect, leaving the effect of compositional changes on a specialty's fertility rate, and then scales it by the entry effect to convert it to per person terms.[58] I use the expression derived above, $MF = \frac{\beta_2 - \beta_1}{\beta_{entry}}$, to quantify the difference between the fertility rate of the marginal and average female entrant for CA. The inputs are the estimates of β_1 and β_2 from Table 5 column 3 and the CA-specific percentage increase in female entry due to the reform, β_{entry} , reported in Appendix Table A.9. The difference between the fertility rate during the first three years of residency of marginal and average female entrant is: $MF = \frac{0.00077 - 0.00097}{0.0038} = -0.05$. Given that the CA fertility rate is 0.13, this difference is substantial.[59]

[57]Since the numerator and denominator are divided by the same first-stage relationship, the first-stage coefficient cancels from the expression.

[58]The expression for the fertility of marginal entrant is equivalent to the approach outlined in Gruber et al. (1999), with one exception. Their analysis does not include the direct effect, β_1. In the Gruber et al. (1999) setting, legislation to restrict/grant access to abortion is used as an instrument for the birth rate, and they compute the outcomes of the marginal child not born due to abortion legislation through two-stage least squares estimation. The exclusion restriction – that abortion legislation did not influence the outcomes of inframarginal children, except through its effect on the birth rate – permits this interpretation. In my setting, absent β_1, the expression is equivalent to two-stage least squares estimation of the effect of new entry on fertility, where the reform serves as an instrument for new entry. The exclusion restriction – that the only effect of the reform on fertility is through the induced entry of women into time-intensive specialties – is clearly not reasonable given my above results on specialty entry, which prompts the modification of the approach to include the effect of the reform on fertility of inframarginal women.

[59]Although the TX estimates are overall noisy, a similar computation yields a difference between the fertility rates of the marginal and average female entrant of -0.037.