Do Long Work Hours Constrain Women's Fertility Choices?

Melanie Wasserman[*]

February 2023

Abstract

Women who work in time-intensive occupations are more likely to postpone having children. This paper investigates whether this postponement is due to the constraints that long hours impose on women's fertility choices. I study the introduction of a policy that capped the weekly hours of early career physicians in the U.S., resulting in differential reductions in hours across medical specialties. By constructing a novel linkage between data on physicians and administrative birth records from two large states, I estimate the effect of a specialty's hours reduction on women's fertility timing. The effect of the reform is theoretically ambiguous. Among inframarginal women—those who would have chosen time-intensive specialties absent the reform—the reform reduces hours worked, which may increase the propensity to have children during residency. But the reform could also change the composition of women who enter time-intensive specialties, leading to a rise or fall in a specialty's fertility rate. I find empirical support for each of these channels, suggesting that an occupation's long work hours both constrain women's fertility choices and induce selection of women based on fertility preferences.

Keywords: long hours, fertility, gender
JEL codes: J13, J16, J24

---

[*] UCLA Anderson School of Management. E-mail: melanie.wasserman@anderson.ucla.edu. I have benefited from discussions with David Autor, Martha Bailey, Manasi Deshpande, Esther Duflo, Paola Giuliano, Danielle Li, Adam Sacarny, Romain Wacziarg, and Heidi Williams. Funding is gratefully acknowledged from the National Science Foundation (NSF) Graduate Research Fellowship, the National Institute on Aging, Grant Number T32-AG000186, and the MIT Shultz Fund.

1 Introduction

Women who work in time-intensive occupations tend to postpone having children. Figure 1 uses U.S. Census data for over 100 large occupations to provide a snapshot of this relationship. Panel I (II) plots the fraction of college-educated women (men) employed in an occupation who have at least one child against the fraction of men in the same occupation who work 50 or more hours per week. For women ages 30-34, there is a strong negative relationship between the propensity to have any children and rates of overwork (Panel I.A). This negative relationship weakens as women get older, but remains statistically significant among women ages 40-44, suggesting that postponement may translate to lower completed fertility (Panels I.B and I.C). In contrast, for men of all ages, there is a *positive* relationship between an occupation's time intensity and having children (Panel II).[1]

There are two explanations for the negative relationship between long work hours and women's fertility timing. First, preferences for having children or the presence of children may cause women to work in less time-intensive occupations. Indeed, a large literature documents the negative effects of motherhood on female labor supply, including whether to work at all as well as the number of work hours, conditional on employment (Angrist and Evans, 1998; Bertrand et al., 2010; Angelov et al., 2016; Kleven et al., 2019). The second explanation is the reverse causal relationship: occupational time demands may constrain women's fertility choices, including the number and timing of children. As of yet, there is little research on this explanation.

This paper investigates the ramifications of occupational time demands for women's fertility timing. I focus on a large professional occupation, physicians, which has a particularly time-intensive early career training period known as medical residency. Each medical specialty has distinct time demands during residency.[2] I confirm that the descriptive patterns among physicians mirror those of the broader college-educated population: female physicians in specialties with

---

[1] The corresponding figures for number of children are found in Appendix Figure A.1.

[2] Although medical residency is the most time-intensive portion of a physician's career, physicians work long hours post-residency as well. U.S. Census data indicates that among physicians ages 40-44, 68 percent of men and 43 percent of women work 50 or more hours per week. Among lawyers ages 40-44, 58 percent of men and 33 percent of women work 50 or more hours per week.

time-intensive residencies tend to postpone having children.

In order to estimate the effect of time demands on women's fertility choices, I study a policy that introduced an arguably exogenous change in the weekly hours of medical residents: the 2003 Accreditation Council for Graduate Medical Education (ACGME) duty hour reform. In particular, the policy capped the weekly hours of early career physicians in the U.S., resulting in differential reductions in hours across medical specialties. I leverage the timing of the reform and pre-policy variation in hours across medical specialties to estimate the effect of a specialty's hours requirements on women's propensity to have children during residency.[3]

A reduction in a specialty's hours can affect its female fertility rate through two channels: directly through the hours reduction and indirectly through changes in the composition of women who choose the specialty. The direct channel represents the effect of an hours reduction on the fertility choices of inframarginal women—those who would have chosen the specialty absent the reform. The reform may also have an indirect effect on a specialty's fertility rate by encouraging the entry of women who are more or less likely than inframarginal women to have children during residency. Combining the direct and indirect channels, the overall effect of the reform is theoretically ambiguous: it could cause a specialty's fertility rate to rise, fall, or stay the same.

To empirically disentangle the direct and indirect effects of an hours reduction on women's fertility timing, I exploit the fact that there were heterogeneous effects of the reform on women's specialty choices across U.S. states.[4] I focus on two states with divergent effects: California and Texas. In California, the reform caused *no change* in women's choice of medical specialty. As a result, using California medical residents, I can isolate the direct effect of an hours reduction on women's fertility choices, holding specialty choice constant. By contrast, in Texas, the reform encouraged women to enter time-intensive specialties, implying that both direct and and indirect effects are potentially operative.

Using a novel linkage of data on early career physicians and administrative birth records in Cal-

---

[3] Wasserman (forthcoming) uses a similar empirical strategy to estimate the nationwide effects of the reform on physicians' specialty choices, by gender.

[4] Using nationwide data on physicians, Wasserman (forthcoming) documents that the reform induced women to enter historically time-intensive specialties, while it had little effect on men's specialty choices.

ifornia and Texas, I estimate the effect of the reform on women's fertility during residency. In California, I find that reducing women's work hours—holding specialty choice constant—increases childbirth during residency. In particular, a four hour per week decrease resulted in 0.02 additional children during the first three years of residency, an increase of 15 percent over the average pre-reform level. In contrast, in Texas there is no evidence of a fertility increase due to the reform. If anything, the point estimates are negative, suggesting that women's sorting based on fertility preferences offsets the direct effects of the reform. While I am unable to observe completed fertility, these results imply that an occupation's long work hours both constrain women's fertility choices and induce selection of women based on fertility preferences.

This paper contributes to a large literature that studies the ramifications of various institutions, laws, and norms for women's fertility choices (Goldin and Katz, 2002; Goldin, 2006; Bailey, 2006, 2010; Bailey and Collins, 2011; Myers, 2017). It is widely documented that college-educated women are more likely to delay marriage and motherhood than their non-college-educated counterparts (Goldin, 2021). Since individuals simultaneously choose educational/career investments and family formation, isolating the effect of early career time demands on fertility timing has proved challenging. Park and Rim (2020) document that female lawyers have children later than male lawyers and this difference is especially pronounced among those vying for partnership, which requires billing and working extremely long hours. The present paper adds to this body of evidence by analyzing the consequences of a sharp change in the time demands of specific career paths for women's fertility timing.

This paper also contributes to the smaller literature on the detrimental effects of long hours requirements on women's earnings, promotions, and occupational choices (Gicheva, 2013; Cha and Weeden, 2014; Goldin, 2014; Cortés and Pan, 2016, 2019; Wasserman, forthcoming). Adding to this body of evidence, I document that the implications of long hours extend beyond labor market outcomes to the timing of family formation.

2 Institutional Details

2.1 Fertility Choices during Medical Residency

Medical residency is known for its extremely long hours. Despite its time intensity, having children during this time period is not a rare occurrence. Physicians start residency, on average, when they are 28 years old, positioning residency to coincide with when college-educated women typically have children. Data from the nationally representative Young Physicians' Survey (YPS) 1991 sample reveal that nearly one quarter of women have children during residency. More recent institution- and specialty-specific surveys indicate that between 25 and 35 percent of female residents have children during residency (Hamilton et al., 2012; Smith et al., 2013; Turner et al., 2012; Chen et al., 2013). Although women generally have children at younger ages than men, female physicians in time-intensive specialties such as General Surgery are less likely than male physicians to have children during residency (Smith et al., 2013; Chen et al., 2013). These patterns suggest that women in demanding specialties could have a preference for later childbearing or, alternatively, that residency time demands differentially constrain women's fertility timing.

Why might long work hours differentially constrain women's fertility choices? Having children imposes more demands on women's than on men's time. In the short-run, pregnancy and giving birth represent substantial health shocks from which women require time to recover. Longer term, well-documented gender gaps in time spent on childcare extend to the physician population: female physicians spend eight hours more per week on childcare than their male counterparts (Jolly et al., 2014).[5] For all residents, however, there is little scope to adjust one's hours along the intensive and extensive margins in response to having children. Rules set by medical specialty boards and residency programs stipulate that an individual must make up missed time or repeat a residency training year if absent from work for more than 4-8 weeks (Davis et al., 2001; Gabbe et al., 2003; Willett et al., 2010). In addition, there are few part-time residency positions. Other costs of taking time off include resentment from co-residents who may be asked to work extra hours to cover for

---

[5] These differences could arise due to gender norms governing time use, intra-household comparative advantage, or preferences.

their colleague who is on leave (Turner et al., 2012).

2.2 ACGME Duty Hour Reform

In the late 1990s and early 2000s, federal legislators, the Occupational Safety and Health Administration (OSHA), and state governments started to scrutinize the role of excessive resident hours in generating medical errors and potentially compromising patient safety. In response, the Accreditation Council for Graduate Medical Education (ACGME) enacted a duty hour reform in 2003, which had four main provisions:

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

This reform represented a watershed change to graduate medical education, the most substantial since the early 1900s (Ludmerer, 2015). It is important to note that while this reform changed the time intensity of medical residency, the length of training and compensation during the training period remained unchanged. The relatively quick implementation of the reform coupled with its focus on weekly hours makes it well suited to studying the ramifications of hours worked on fertility choices during residency.

To document the effect of the reform on hours worked, I use the Current Population Survey and surveys of medical residents. Figure 2 Panel I uses reports of hours worked in the previous week from the CPS monthly files to plot the average weekly hours of physicians from 1989 through 2014 for medical residents and non-resident physicians. As medical resident status is not observed in the CPS, I impute it based on an individual's age (<35), occupation (physician), and if the individual works in a hospital. In the years preceding the introduction of the duty hour reform in 2003, medical residents worked, on average, 64 hours in the previous week, well above the average of 50 hours worked by non-resident physicians. While there has been a long-term decline in the hours of non-resident physicians, the hours for resident physicians are stable prior to the reform and drop when the reform is enacted in 2003 (Panel I.A). Relative to fully trained physicians, residents experience a four hour per week decline, on average, after the reform.

Given that the reform restricted average hours per week to 80, we expect the upper end of the hours distribution to be primarily affected. Panel II.A shows the fraction of resident physicians who report working more than 80 hours in the previous week falls after the introduction of the policy, while there is little change among non-resident physicians. Since remainder of the empirical analysis focuses on physicians' fertility choices in California and Texas, I also present state-specific evidence on hours worked. While noisier, the patterns are similar for physicians in these two states (Panels I.B and II.B).

In order to estimate the effect of the reform on women's fertility timing, I leverage variation in the extent to which the hours cap was binding across specialties. The reform should cause steeper hours reductions in the most time-intensive specialties, such as the surgical specialites, where the typical resident pre-reform worked far in excess of 80 hours per week (Philibert et al., 2009). I test whether this is the case in Appendix Figure A.2, which uses surveys of medical residents to plot 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. The slope of the line 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-reform.[6]

2.3 Anticipated Effects of the Reform on Fertility Timing

To guide our understanding of how a reduction in hours affects fertility decisions, I offer a simple conceptual framework in which physicians jointly choose their medical specialty and whether

---

[6] State-specific data on hours worked by specialty before and after the reform is not available.

to have children during residency (formalized in Appendix B). Specialties are characterized by their hours worked during residency and wages post-residency. Individuals exhibit heterogeneity in preferences for having children during residency as well as the relative valuation of non-market time and wages. The framework generates the following predictions regarding specialty and fertility choices during residency. If women incur an additional disutility of hours worked when they have children, then they are less likely than men to enter time-intensive specialties. In addition, conditional on entering a time-intensive specialty, women are less likely than men to have children during residency.

There is also an ambiguous prediction regarding the effect of a specialty's reduction in hours on a specialty's female fertility rate. The ambiguity arises due to two potentially offsetting phenomena. First, an hours reduction could increase the rates of having children during residency for inframarginal women, that is, those women who would have chosen a time-intensive specialty absent the reform (the direct effect). Second, an hours reduction may change the composition of women who enter time-intensive specialties: women induced to enter a specialty due to the reform may differ from inframarginal women in their preferences over fertility timing (the indirect effect). Notably, marginal women may be drawn from (1) women who always have children during residency, (2) women who never have children during residency, and (3) women who switch from having children in a low hours specialty to not having children in a high hours specialty.[7] Depending on the magnitude of the fertility increase among inframarginal women and the composition of marginal women, a specialty's fertility rate can rise, fall or stay the same in response to a reduction in hours.

To disentangle the direct and indirect effects, the empirical analysis proceeds in three steps. First, I provide evidence of heterogeneous effects of the reform on women's specialty choices in California and Texas. Then, I leverage the absence of indirect effects in California in order to isolate the direct effect. Finally, I estimate the effects of the reform on women's fertility timing in

---

[7] Note that the composition of women who switch specialties hinges on the distribution of these types throughout the physician population. Most women do not have children during residency, so it is possible that women induced to enter time-intensive specialties due to the reform are disproportionately drawn from group (2).

Texas, taking into consideration direct and indirect effects of the reform.

3 Data and Summary Statistics

3.1 Data Sources

American Medical Association Physician Masterfile: In order to estimate the effect of the reform on specialty choice, I use the American Medical Association (AMA) Physician Masterfile, which has information on demographic characteristics (gender, age, and birthplace), medical training history (medical school, residency training institution, start date of residency) and primary specialty for all physicians in the U.S. For the subset of physicians who did any part of their residency training in California or Texas, the file includes first and last names, which permits linkages to birth records.

Vital Statistics Birth Records: In order to analyze the effect of the duty hours reform on the family formation decisions of medical residents, I construct a new linkage between the AMA Masterfile and confidential Vital Statistics birth records from California and Texas, 1993-2013 (California Health and Human Services Agency, 2016; Texas Department of State Health Services, 2016). The birth records include parental identifiers (mother/father names, dates of birth), child birthdate and demographic characteristics. Due to limited information on fathers in the CA birth records, it is only possible to link female physicians. Using female physician first name, last name (maiden and/or current), year of birth and birthplace (U.S. state or country), I conduct a probabilistic merge of the physicians in the Masterfile with CA/TX Vital Statistics birth records. Texas birth certificates additionally record mother occupation, which I use as a post-merge verification of the quality of the match.

Other Data Sources: I classify specialties based on their pre-policy time intensity during residency training with use of a 1999 nationally representative survey of approximately 2,000 second year medical residents' hours worked, conducted by Baldwin Jr et al. (2003). The study reports the average hours worked during the second year of residency for twenty-one specialty categories.[8]

3.2 Sample Restrictions and Summary Statistics

The sample is limited to U.S. medical school graduates (USMGs) who started residency training between 1993 and 2010, and completed the first three years of residency in California or Texas.[9] The sample ends in 2010 in order to observe fertility outcomes for the first three years of residency without censoring and to avoid confounding the effects of a 2011 reform, which limited the maximum shift length of first year medical residents to 16 hours, with the 2003 duty hours reform studied in this paper. I exclude foreign graduates and individuals who graduated from osteopathic medical schools and participated in an M.D. residency program, as there is a high incidence of missing specialty information among this population, which increases over time. 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 is 12,616 female physicians in California and 7,106 in Texas.

Appendix Table A.1 presents summary statistics for the full sample of physicians, physicians who were U.S. medical school graduates, and the California and Texas USMG samples (for both men and women). In comparison to all female USMGs, CA physicians are less likely to be born

---

[8] The twenty-one 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, Preventive Medicine, 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.

[9] 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.

in the U.S. and more likely to have attended a ranked medical school. Texas physicians are comparable to the female USMG population, with the exception that they are slightly younger. The bottom portion of the table reports summary statistics for fertility outcomes during the first three years of residency.[10] Since residency lasts at least three years, I observe with near certainty the fertility of individuals who completed their first three years of residency training in California or Texas.[11] Among California medical residents, the mean 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 California, female residents in Texas are more likely to have children during residency, with 19 percent having at least one child and the mean number of children 0.20.[12]

3.3 Descriptive Evidence on Entry and Fertility Timing

Using the linked Masterfile and Vital Statistics birth records, Figure 3 plots the relationship between pre-policy fertility during residency, pre-policy fraction of residents in each specialty who are female, and pre-policy average hours per week.[13] As documented in Wasserman (forthcoming), women's representation declines substantially with specialty hours requirements. Consistent with the conceptual framework and Figure 1, female fertility is also negatively correlated with residency time demands. The higher fertility rates among Texas residents are particularly apparent in the low hours specialties, whereas the fertility rates in high hours specialties are comparable across the two states. While data limitations preclude linking male physicians to Vital Statistics records, in Appendix Figure A.4 I use the 1991 Young Physicians Survey to replicate these relationships for both male and female physicians. Women's propensity to have children during residency is negatively correlated with residency hours, while there is a positive relationship for men.

---

[10] There are slightly fewer observations (36 in CA and 13 in TX) due to missing name information, which is used to match with the Vital Statistics data.

[11] The Young Physicians' Survey (YPS) 1991 sample documents that a small percentage (8%) of female physicians have children before or during medical school, while a much larger percentage (25%) have children during residency training.

[12] I confirmed these differences in the average number of children in TX and CA using data on young female physicians from the 2000 Census.

[13] Due to confidentiality restrictions regarding small cells, the fertility rates for two specialties could not be reported. The patterns are similar for having any children during the first three years of residency (Appendix Figure A.3).

4 Effect of the Reform on Specialty Choice and Fertility

The empirical analysis proceeds in three steps. First, I document that the reform had heterogeneous effects on women's specialty choices in California and Texas. Then, I leverage the absence of indirect effects in California in order to isolate the direct effect of the reform on women's fertility timing. Finally, I estimate the effects of the reform on fertility timing in Texas, which encompasses both direct and indirect effects of the reform.

4.1 Specialty Choice

4.1.1 Estimation

In order to estimate the effect of the reform on women's specialty choices, I contrast the propensity of women to enter more and less time-intensive specialties, before and after the reform, using the following conditional logit specification:

$$\Pr(C_{it}=s)=\frac{\exp(\lambda_1(\text{Hours}_{s,1999}\times\text{During}_t)+\lambda_2(\text{Hours}_{s,1999}\times\text{Post}_t)+\text{X}_i'\delta_s+\alpha_s)}{\sum_{s'\in S}\exp(\lambda_1(\text{Hours}_{s',1999}\times\text{During}_t)+\lambda_2(\text{Hours}_{s',1999}\times\text{Post}_t)+\text{X}_i'\delta_{s'}+\alpha_{s'})} \quad (1)$$

where C_{it} represents the specialty outcome of individual i who graduated from medical school/entered a residency program in year t, $\text{Hours}_{s,1999}$ represents a specialty's pre-policy hours, During_t is an indicator for entering residency in 2001-2002, and Post_t is an indicator for entering residency 2003-2010. Specialty fixed effects are represented by α_s and X_i is a vector of individual controls, including age at medical school graduation and rank of medical school attended. I estimate separate effects for two groups: (1) 2001-2002 medical school graduates, who had already chosen their specialty but could still change their fertility during the first three years of residency in response to the reform, and (2) 2003-2010 medical school graduates, who had the opportunity to change their medical specialty and their fertility in response to the reform.[14] Standard errors are heteroskedastic

---

[14] Note that this specification is different from the one used in Wasserman (forthcoming), which allows the effect of the reform on specialty choice to evolve during the post-reform period. To maintain consistency with the fertility

robust. I test the robustness of the results to controlling for resident demographic characteristics, including age and whether their medical school was ranked by U.S. News and World Report. I also include selected specialty time trends, to account for the strong secular trend in women's entry into Ob/Gyn and the declining interest among all physicians in Primary Care specialties: Family Practice, Internal Medicine, and Pediatrics.

4.1.2 Results

Table 1 reports the average marginal effects of the reform on specialty entry for individuals who completed residency in California (Panel A) and Texas (Panel B).[15] Based on the results in Panel A columns 1-3, it is evident that the reform had no effect on women's specialty choice in California. In Texas, however, a reduction in hours significantly encouraged specialty entry among women, consistent with nationwide effects documented in Wasserman (forthcoming). An additional hour pre-policy increased the probability that women chose a specialty post-reform by 0.036 percentage points. For completeness, I also include the results for men. After the reform, men's entry into time-intensive specialties increased in CA and decreased in TX.

What explains the state-level differences in the effects of the reform on women's entry? In Appendix C I investigate the correlates of state-level heterogeneity in the effects of the reform on specialty sorting, using all available U.S. states. The reform induced women to enter time-intensive specialties in states where at baseline (1) there was lower female representation in time-intensive specialties, (2) it was easier to obtain a residency position, and (3) female residents were more likely to have more children. Across all states, California is in the top third for female representation and competitiveness and the bottom third for fertility, consistent with the observed low effect of the reform on female specialty entry. Texas, on the other hand, is among the top third for the effects of the reform on specialty entry, middle third for female fertility and competitiveness, and bottom third for female representation.

---

analysis below, I split cohorts based on their exposure to the reform during residency. The effect of the reform on specialty choice does evolve over time in TX, as documented in the event study plots in Appendix Figure A.5.

[15] The coefficients are reported in Appendix Table A.2.

In California, the absence of effects of the reform on women's specialty entry creates a unique opportunity to analyze the effects of an occupations' time requirements on women's fertility timing, holding occupational choice constant. In the absence of compositional changes, the estimated fertility effect represents the direct effect of the reform on the fertility timing of inframarginal women. As discussed in Section 2.3, the strong positive effect of the reform on women's specialty entry in Texas implies that the effects of the reform on fertility timing will represent a combination of direct and indirect effects: the reform can alter a specialty's fertility rate by reducing the hours of inframarginal women and/or by inducing women with differing fertility preferences to enter time-intensive specialties.

4.1.3 Tests of Internal Validity

I examine the validity of the identifying assumption, that the propensity to enter more versus less time-intensive specialties was stable pre-reform, using an event-study version of Equation 1, with 2000 as the omitted year. The results are graphically depicted in Appendix Figure A.5, which reveals no significant trends prior to the reform.

4.2 Fertility Timing

4.2.1 Estimation

In order to estimate the effect of the reform on fertility choices, I contrast the fertility outcomes of women in more and less time-intensive specialties, before and after the duty hour reform:

$$Y_{ist} = \beta_0 + \beta_1(\text{Hours}_s \times \text{During}_t) + \beta_2(\text{Hours}_s \times \text{Post}_t) + \alpha_s + \gamma_t + X'_{ist}\delta + \varepsilon_{ist} \tag{2}$$

where Y_{ist} is the fertility outcome of individual i from residency cohort t who entered specialty s, $Hours_s$ are the average pre-policy hours of specialty s, $During_t$ is an indicator for starting residency training in 2001 or 2002, $Post_t$ is an indicator for starting residency training in 2003 onward, 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 separate grouping for individuals who started residency in 2001 or 2002 captures their partial exposure to the duty hours reform, since they were in residency at the time the reform was enacted and likely experienced a reduction in hours. Standard errors are clustered at the specialty level.[16]

Note that an alternative strategy for separately identifying the direct and indirect effects is to focus on the $During_t$ cohorts, that is, those residents who were already in residency (and had therefore chosen their specialty) at the time the reform was enacted. I did not use this strategy for a few reasons. First, due to only two cohorts falling into this category, there is a limited number of residents, leading the coefficients to be imprecisely estimated (the standard errors are more than twice as large for the coefficients on $\text{Hours}_s \times \text{During}_t$). Second, these cohorts had less exposure to the reform during the first three years of residency. A physician who started residency in 2001 experienced one year of reduced hours, during her third year of residency. Third, these cohorts were surprised by the reform, which may have limited their capacity to plan/have children within the first three years of residency.

4.2.2 Results

The results from the OLS estimation of equation 2 are presented in Table 2. The coefficient β_2 on the interaction $(Hours_s \times Post_t)$ is reported in the first row of each panel. Columns 1-3 present results with the progressive inclusion of individual demographic and educational control variables. Starting with column 1, there is a positive and statistically significant effect of the duty hours reform on fertility in California and a negative effect in Texas. These results change very little with the inclusion of controls for resident age or medical school or specialty time trends (columns 2-4). In order to gauge the sensitivity of the results to the potential incidence of false positive matches between the CA/TX Masterfile and the Vital Statistics data, in column 5 I restrict the sample to individuals without common last names, defined by their frequency. This restriction yields little change to the magnitude of the coefficients, though the loss in sample size reduces the

---

[16] The OLS specification permits the inclusion of more granular controls for medical school and age.

precision of the estimates. Overall, it appears that there was a positive effect of the reduction in hours on a specialty's fertility rate in California and a negative effect in Texas.

As discussed in Section 2.3, the combined effect of the reform on a specialty's fertility rate depends on (1) direct effect on the fertility among inframarginal women, and (2) the indirect effect on fertility due to new specialty entrants. In California the reform does not affect women's propensity to enter a time-intensive specialty, implying the estimated positive effects on fertility represent the direct effect. Taking the coefficient from column 1 for California, the duty hours reform increased the number of children female residents had during their first three years of residency by 0.00093 children.[17] Scaling this coefficient by the first stage relationship, -0.17, and the average decline of four hours across all specialties, a four hour per week reduction due to the duty hours reform corresponds to an increase of 0.02 children. Given that the pre-policy average number of children during the first three years of residency is 0.13, this effect amounts to more than a 15 percent increase in fertility during the first three years of residency.

In Texas, by contrast, the reform increases women's entry into time-intensive specialties, implying that the estimated negative effects on fertility capture both the direct and the indirect effects. While it is not possible to disentangle these two effects in Texas, it appears the indirect effect is negative: women who prefer to not have children during residency or who are willing to trade off higher earnings for later childbearing differentially select into time-intensive specialties due to the reform.

4.2.3 Tests of Internal Validity

Parallel trends: The identifying assumption for the analysis to yield unbiased estimates of the total effect of the duty hours reform on fertility, is that absent the reform, the fertility outcomes of individuals in more and less time-intensive specialties would have evolved similarly over time. 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 duty hours

---

[17] Since there are fewer than 50 clusters, I have also computed wild cluster bootstrapped t-statistics (Cameron et al., 2008), which imply that coefficients on Post in columns 3 and 4 remain statistically different from zero.

reform. Appendix Figure A.6 presents the year-by-year coefficients from the following event study specification:

$$Y_{ist} = \beta_0 + \sum_{m=1993}^{2010} \beta_m(\text{Hours}_s \times 1[t = m]) + \alpha_s + \gamma_t + X'_{ist}\delta + \varepsilon_{ist} \tag{3}$$

where β_m represents year-by-year contrasts of fertility outcomes of women in more and less time-intensive specialties, and all other variables are defined above. While the yearly coefficients are less precisely estimated, there is little evidence of pre-trends. In CA, there appears (if anything) to be a negative trend in the years preceding the duty hours reform. Immediately after the duty hours reform is implemented, the coefficients increase substantially and are positive, but the positive effects are not sustained after 2007.

Contemporaneous policy changes: In July 2004, California's Paid Family Leave (PFL) policy went into effect, which entitled parents partial wage replacement for six weeks after childbirth. While it is possible that this policy contributed to the positive effects of the duty hour reform on California residents' fertility, it is unlikely to be the case. Prior to the introduction of PFL, women were already entitled to partial wage replacement for four weeks prior to childbirth and six weeks after childbirth through CA's temporary disability insurance (TDI). PFL was designed to extend parental leave, i.e. to start immediately after TDI leave ends (Rossin-Slater et al., 2013). Prior evidence shows that modest extensions of parental leave do not affect the decision to have children and PFL in particular decreased fertility (Dahl et al., 2016; Bailey et al., 2019).[18] As discussed in Section 2.1, specialty boards and residency programs limit the number of weeks that a resident can take leave during a residency year before being required to make up the time or repeat the year, making additional leave taking through PFL particularly costly.

Another contemporaneous policy change is the 2003 Texas Medical Liability Act, which made medical malpractice lawsuits more difficult to file, increased evidentiary requirements, and capped

---

[18] In contrast, Lalive and Zweimüller (2009) and Raute (2019) document that more substantial leave extensions and expansions of benefits lead to increased fertility, particularly for higher-order births.

recovery damages. Since specialties that are time-intensive during residency—such as surgical specialties—have a higher risk of medical malpractice claims, it is possible that women were drawn to more time-intensive specialties due to their reduced risk of medical malpractice claims rather than their reduced hours (Jena et al., 2011). Medical claims and payouts dropped sharply after the Texas reform, but there is not evidence on whether the drop differed by specialty (Hyman et al., 2013). Evidence from medical malpractice reforms across various states shows that reductions in damages were *smaller* in high-risk specialties relative to low-risk specialties (Seabury et al., 2014). Furthermore, while risk of malpractice claims varies substantially across specialties and states, physicians' fear of malpractice claims is relative constant across high- and low-risk specialties, and across high- and low-risk states (Carrier et al., 2010). Perhaps for this reason, research on the effects of malpractice liability on physician labor supply has produced mixed findings: some studies find that reductions in medical liability increase physicians per capita, particularly in rural areas and high-risk specialties, and others find no or even negative effects.[19] The literature has not examined whether malpractice reforms differentially affect physician labor supply by physician gender.

Composition changes: The absence of specialty entry effects in California permits isolation of the direct effects of the hours reduction. Even in the absence of a change in the *quantity* of women, there could be a change in the *composition* of women who enter time-intensive specialties. In Appendix Table A.3 I investigate whether the reform affects specialty composition using two demographic characteristics of physicians: age and whether they attended a ranked medical school. The reform encourages older physicians (both male and female) to enter time-intensive specialties, but the effects are of modest magnitude. There is no change in the medical school ranking of physicians who enter time-intensive specialties. Interestingly, the composition of Texas residents does not change much in response to the reform, suggesting that these characteristics may not

---

[19] See, for example, Klick and Stratmann (2007), Helland and Seabury (2015), and Paik et al. (2016). Looking specifically at the Texas 2003 reform, Hyman et al. (2013) finds no effect on physician labor supply.

capture underlying fertility preferences.[20] As a further test of the role of demographic changes in contributing to the reform's effects on fertility, Table 2 columns 2 and 3 include controls for age and medical school, which do not change the direct effects of the reform.

5 Conclusion

Long work hours have become a fixture of the U.S. labor market, especially among professional occupations (Kuhn and Lozano, 2008). Since certain occupations disproportionately reward working long hours—and women work fewer hours than men—these temporal demands are central to understanding gender disparities in pay (Goldin, 2014). This paper demonstrates that the ramifications of long hours extend beyond women's labor market outcomes, namely, to decisions of when to have children. By studying a reform that restricted the work hours of physicians during their early career years, I find that long hours constrain women's fertility choices and induce occupational sorting based on fertility preferences. While this paper is unable to analyze completed fertility, it represents a first step in examining the consequences of occupational time demands for fertility choices.

---

[20] I do not have access to data on whether a physician has children during medical school in order to test whether the reform causes women with children to enter time-intensive specialties.

References

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

Angelov, Nikolay, Per Johansson, and Erica Lindahl, "Parenthood and the Gender Gap in Pay," *Journal of Labor Economics*, 2016, *34* (3), 545–579.

Angrist, Joshua D. and William N. Evans, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *American Economic Review*, 1998, *88* (3), 450–477.

Bailey, Martha J, "More Power to the Pill: The Impact of Contraceptive Freedom on Womens Lifecyle Labor Supply," *The Quarterly Journal of Economics*, 2006, *121* (February), 289–320.

_ , ""Momma's Got the Pill": How Anthony Comstock and Griswold v . Connecticut Shaped US Childbearing," *American Economic Review*, 2010, *100* (1), 98–129.

_ **and William J Collins**, "Did Improvements in Household Technology Cause the Baby Boom? Evidence from Electrification, Appliance Diffusion, and the Amish," *American Economic Journal: Macroeconomics*, 2011, *3* (2), 189–217.

Bailey, Martha, Tanya Baker, Elena Patel, and Shanti Ramnath, "The Long-Term Effects of California's 2004 Paid Family Leave Act on Women's Careers: Evidence from U.S. Tax Data," *NBER Working Paper #26416*, 2019.

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.

Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz, "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors," *American Economic Journal: Applied Economics*, 2010, *2* (3), 228–255.

California Health and Human Services Agency, "Vital Statistics Birth Data," 2016.

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.

Carrier, Emily R., James D. Reschovsky, Michelle M. Mello, Ralph C. Mayrell, and David Katz, "Physicians' fears of malpractice lawsuits are not assuaged by tort reforms," *Health Affairs*, 9 2010, *29*, 1585–1592.

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," *Journal of Labor Economics*, 2019, *37* (2).

Dahl, Gordon B., Katrine V. Løken, Magne Mogstad, and Kari Vea Salvanes, "What Is the Case for Paid Maternity Leave?," *The Review of Economics and Statistics*, 10 2016, *98* (4), 655–670.

Davis, Jane L, Susan Baillie, Carol S Hodgson, Louis Vontver, and Lawrence D Platt, "Maternity Leave : Existing Policies in Obstetrics and Gynecology Residency Programs," 2001, *98* (6), 1093–1098.

Gabbe, Steven G., Maria a. Morgan, Michael L. Power, Jay Schulkin, and Sterling B. Williams, "Duty hours and pregnancy outcome among residents in obstetrics and gynecology," *Obstetrics and Gynecology*, 2003, *102* (5), 948–951.

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, C., *Career and Family: Women's Century-Long Journey Toward Equity*, Princeton University Press, 2021.

Goldin, Claudia, "The Quiet Revolution That Transformed Women's Employment, Education, and Family," *AEA Papers and Proceedings*, 2006.

_ , "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.

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.

Helland, Eric and Seth A. Seabury, "Tort reform and physician labor supply: A review of the evidence," *International Review of Law and Economics*, 2015, *42*, 192–202.

Hyman, David A., Charles Silver, Bernard Black, and Myungho Paik, "Does tort reform affect physician supply? Evidence from Texas," *International Review of Law and Economics*, 4 2013, *42*, 203–218.

Jena, Anupam B, Seth Seabury, Darius Lakdawalla, and Amitabh Chandra, "Malpractice Risk According to Physician Specialty," *The New England Journal of Medicine*, 2011, *365*, 629–636.

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.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard, "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 2019, *11* (4), 181–209.

Klick, Jonathan and Thomas Stratmann, "Medical Malpractice Reform and Physicians in High-Risk Specialties," *Journal of Legal Studies*, 2007, *36*, S121–S142.

Kuhn, Peter and Fernando Lozano, "The expanding workweek? Understanding trends in long work hours among U.S. men, 1979-2006," *Journal of Labor Economics*, 2008, *26* (2), 311–343.

Lalive, Rafael and Josef Zweimüller, "How Does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments*," *The Quarterly Journal of Economics*, 08 2009, *124* (3), 1363–1402.

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.

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

Myers, Caitlin Knowles, "The Power of Abortion Policy: Reexamining the Effects of Young Women's Access to Reproductive Control," *Journal of Political Economy*, 2017, *125* (6), 2178–2224.

Paik, Myungho, Bernard Black, and David A. Hyman, "Damage Caps and the Labor Supply of Physicians: Evidence from the Third Reform Wave," *American Law and Economics Review*, 09 2016, *18* (2), 463–505.

Park, Kyung H and Nayoung Rim, "The Disparate Impact of Up-or-Out Promotion Policy on Fertility Timing," *American Law and Economics Review*, 03 2020, *22* (1), 127–172.

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.

Raute, Anna, "Can financial incentives reduce the baby gap? Evidence from a reform in maternity leave benefits," *Journal of Public Economics*, 2019, *169*, 203–222.

Rossin-Slater, Maya, Christopher Ruhm, and Jane Waldfogel, "The Effects of California's Paid Family Leave Program on Mothers' Leave-Taking and Subsequent Labor Market Outcomes," *Journal of Policy Analysis and Management*, 2013, *32* (2), 224–245.

Seabury, Seth A., Eric Helland, and Anupam B. Jena, "Medical malpractice reform: Noneconomic damages caps reduced payments 15 percent, with varied effects by specialty," *Health Affairs*, 2014, *33*, 2048–2056.

Smith, Caitlin, Joseph M Galante, Jonathan L Pierce, and Lynette A Scherer, "The Surgical Residency Baby Boom : Changing Patterns of Childbearing During Residency Over a 30-Year Span," *Journal of Graduate Medical Education*, 2013, (December), 625–629.

Texas Department of State Health Services, "Vital Statistics Birth Data," 2016.

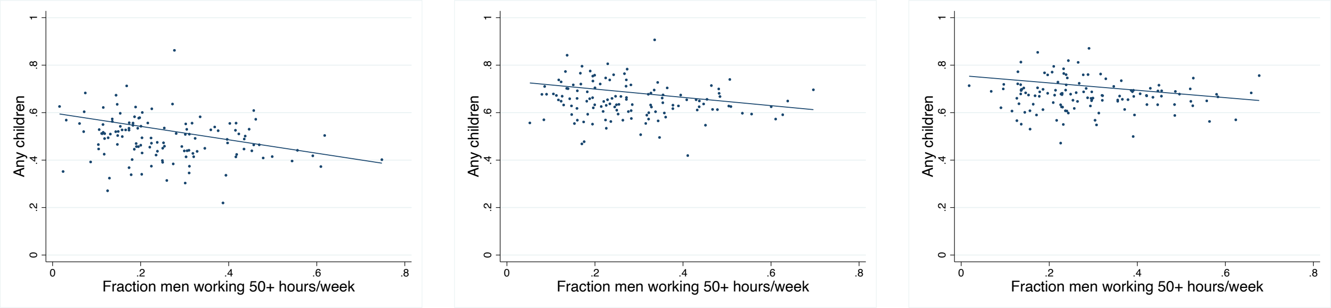
Turner, Patricia L, Kimberly Lumpkins, Joel Gabre, Maggie J Lin, Xinggang Liu, and Michael Terrin, "Pregnancy Among Women Surgeons: Trends over Time," *Archives of Surgery*, 2012, *147* (5), 474–9.

Wasserman, Melanie, "Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents," *Review of Economic Studies*, forthcoming.

Willett, Lisa L, Melissa F Wellons, Jason R Hartig, Lindsey Roenigk, Mukta Panda, Angela T Dearinger, Jeroan Allison, and Thomas K Houston, "Do Women Residents Delay Childbearing Due to Perceived Career Threats?," *Gender Issues*, 2010, *85* (4), 640–646.

Figure 1: Occupational Time Demands and Fertility

I. Women

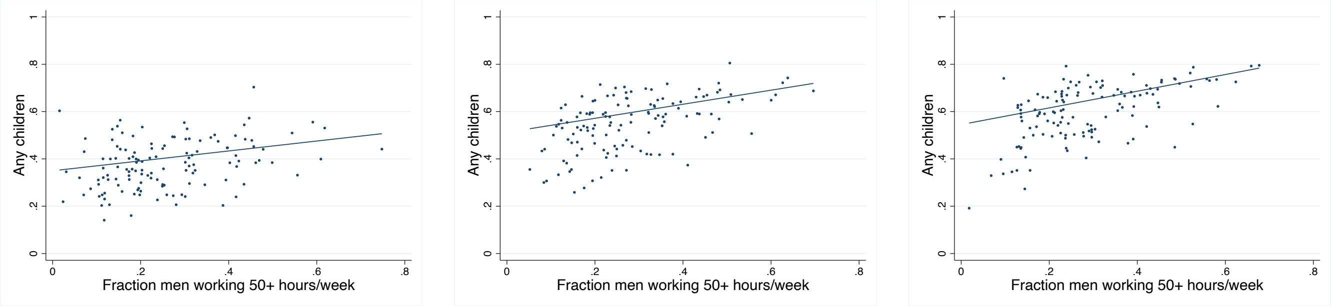


A. Age 30-34

B. Age 35-39

C. Age 40-44

II. Men



A. Age 30-34

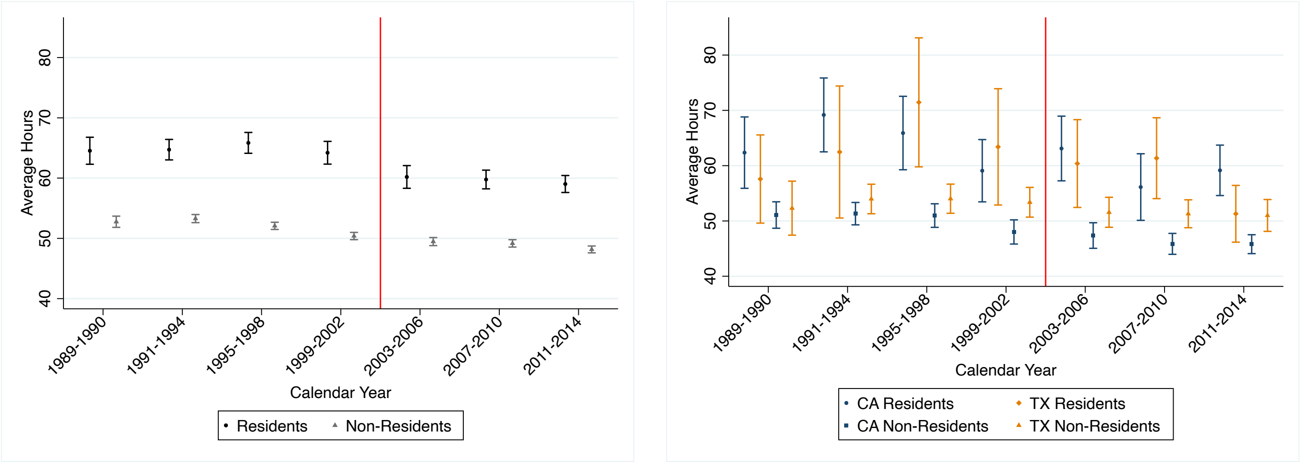
B. Age 35-39

C. Age 40-44

Source: U.S. Census 2000; American Community Survey 2005-2012 Note: The sample is limited to individuals with a Bachelor's degree or more who are in the labor force and have valid information for their occupation. An occupation is sufficiently large to include if it has at least 300 female observations in each age grouping. The fraction of men who usually work 50 or more hours per week is computed for each occupation and age grouping. Similarly, the fraction of women and men who have at least one child (including step-children and adopted children) is computed for each occupation and age grouping. The line of best fit is weighted by occupation cell size.

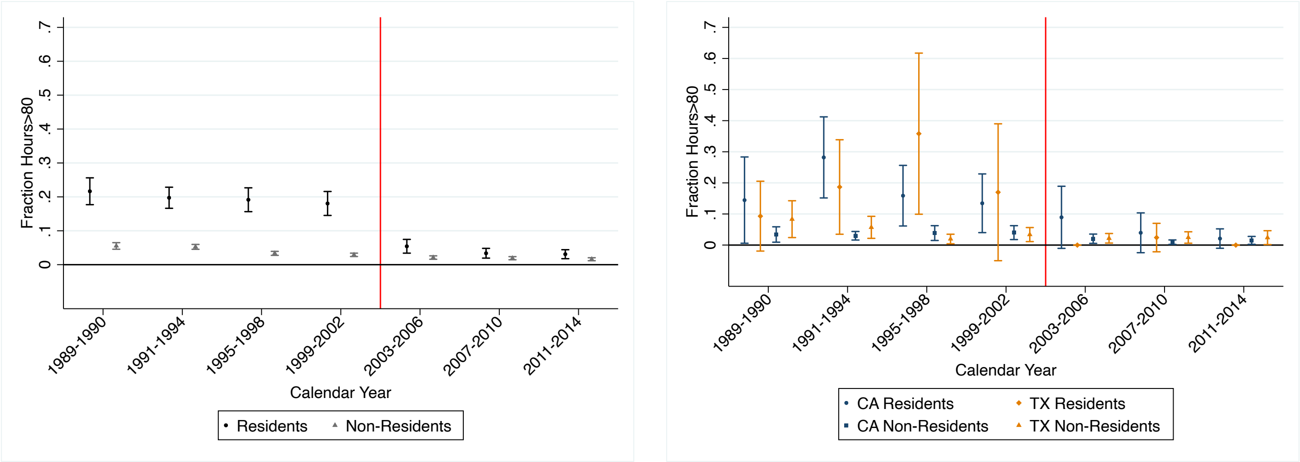
Figure 2: Hours Worked in the Prior Week among Resident and Non-Resident Physicians, 1989-2014

I. Average Hours


A. All Physicians

B. CA/TX Physicians

II. >80 Hours


A. All Physicians

B. CA/TX Physicians

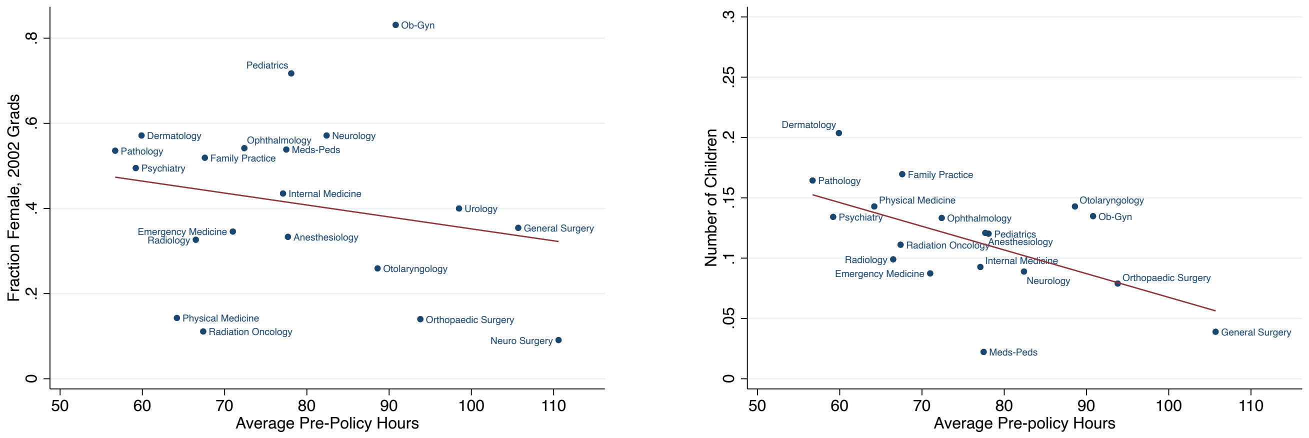
Source: Current Population Survey, monthly files January 1989-December 2014. Note: Panel I plots the average number of hours worked last week for physicians, separately for residents and non-residents. Panel II plots the fraction of physicians who worked more than 80 hours last week, separately for residents and non-residents. Resident status is imputed based on age (<35) and whether the individual works in a hospital. Panel A includes all physicians in the U.S. Panel B provides separate estimates for physicians working in California and Texas. CPS sampling weights are used. The whiskers represent 95% confidence intervals.

Figure 3: Pre-Policy Female Representation, Fertility, and Pre-Policy Hours Requirements

I. California

A. Female Representation

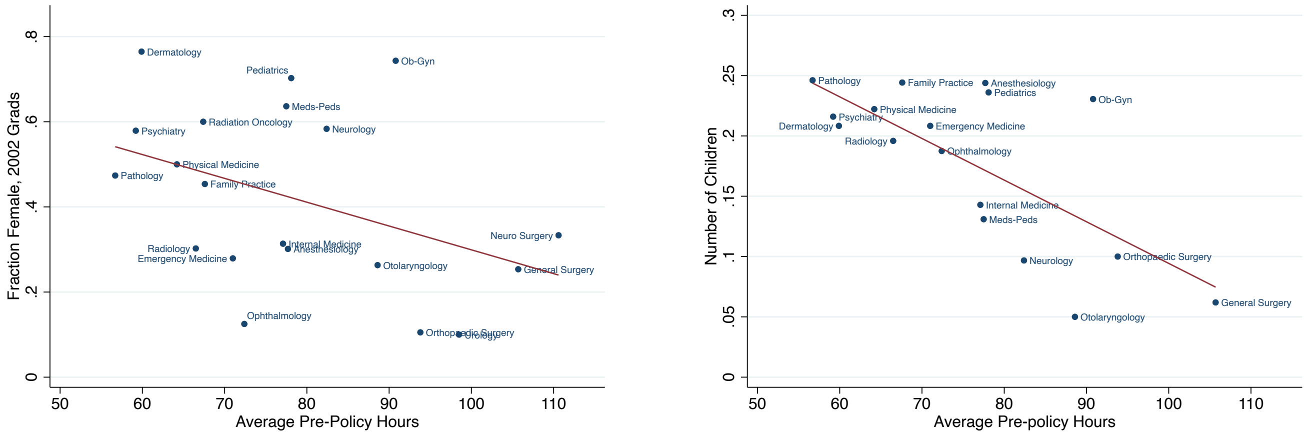
B. Female Fertility



II. Texas

A. Female Representation

B. Female Fertility



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 fertility sample includes female U.S. medical school graduates from years 1993 through 2010 who completed their first three years of residency training in CA. 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: Effect of the Reform on Specialty Choice

Dependent Variable: Specialty Outcome	Women			Men		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: California						
Avg Weekly Hours × Start 2003-2010	-0.002	-0.003	0.004	0.020***	0.025***	0.013**
	(0.007)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)
Avg Weekly Hours × Start 2001-2002	-0.013	-0.013	-0.01	0.005	0.006	0.002
	(0.012)	(0.012)	(0.012)	(0.010)	(0.010)	(0.010)
# individuals	12,616	12,616	12,616	14,017	14,017	14,017
Panel B: Texas						
Avg Weekly Hours × Start 2003-2010	0.036***	0.030***	0.041***	-0.018**	-0.016**	-0.032***
	(0.010)	(0.010)	(0.011)	(0.008)	(0.008)	(0.008)
Avg Weekly Hours × Start 2001-2002	0.004	0.001	0.008	0.002	0.002	-0.006
	(0.015)	(0.015)	(0.016)	(0.011)	(0.011)	(0.011)
# individuals	7,106	7,106	7,106	9,295	9,295	9,295
Specialty FE	X	X	X	X	X	X
Age, medical school ranking		X			X	
Primary Care / OBGYN time trends			X			X

Source: AMA Physician Masterfile Baldwin Jr et al. (2003). Note: This table reports the results of maximum likelihood estimation of a conditional logit model in which the baseline specification has specialty outcome as the dependent variable and the explanatory variables include specialty fixed effects and specialty hours interacted with an indicator for graduating medical school 2003 onwards and an indicator for graduating medical school in 2001/2. The average marginal effects associated with the coefficients on the interaction terms are reported. Panels A and B present results for the sample of USMGs who completed their first three years of residency in CA and TX, respectively. Heteroskedastic robust standard errors are in parentheses.

Table 2: Effect of the Reform on Women's Fertility Timing

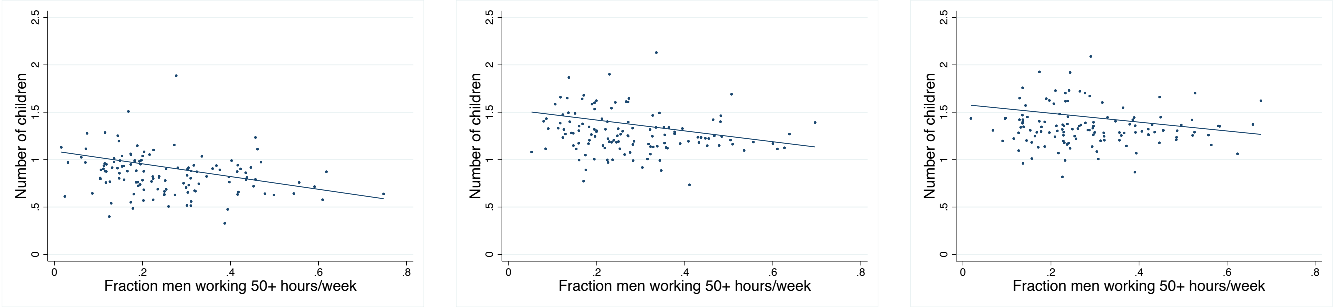
Dependent Variable: # children born during first three years of residency × 1000	(1)	(2)	(3)	(4)	(5)
Panel A: California					
Avg Weekly Hours × Start 2003-2010	0.93*	1.05**	1.07**	1.73***	1.17**
	(0.51)	(0.49)	(0.45)	(0.54)	(0.54)
Avg Weekly Hours × Start 2001-2002	-0.38	-0.38	-0.31	0.00	0.08
	(1.18)	(1.19)	(1.15)	(1.03)	(1.25)
R-squared	0.008	0.025	0.039	0.040	0.043
N	12,580	12,580	12,580	12,580	10,788
Panel B: Texas					
Avg Weekly Hours × Start 2003-2010	-1.58	-1.48	-1.74*	-2.23*	-1.51
	(0.98)	(0.95)	(0.99)	(1.10)	(0.98)
Avg Weekly Hours × Start 2001-2002	0.40	0.55	0.38	0.11	1.72
	(2.24)	(2.25)	(2.28)	(2.49)	(2.26)
R-squared	0.029	0.038	0.059	0.059	0.062
N	7093	7093	7093	7093	6468
Residency start year FE	X	X	X	X	X
Specialty FE	X	X	X	X	X
Age at medical school graduation FE		X	X	X	X
Medical school FE			X	X	X
Primary Care, OBGYN time trends				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-differences 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 after the reform (Start 2003-2010) or was doing residency at the time of the reform (Start 2001-2002). The omitted cohorts are individuals who started residency 1993-2000. Standard errors clustered at the specialty level are in parentheses. The sample includes individuals who did their first three years of residency in CA or TX and who started residency training between 1993 and 2010. Reference 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 1-4 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 Fertility Sample and CA Vital Statistics data, Column 5 excludes from sample individuals with common last names, defined as 50 or more occurrences.

A Appendix Figures and Tables

Figure A.1: Occupational Time Demands and Number of Children

I. Women



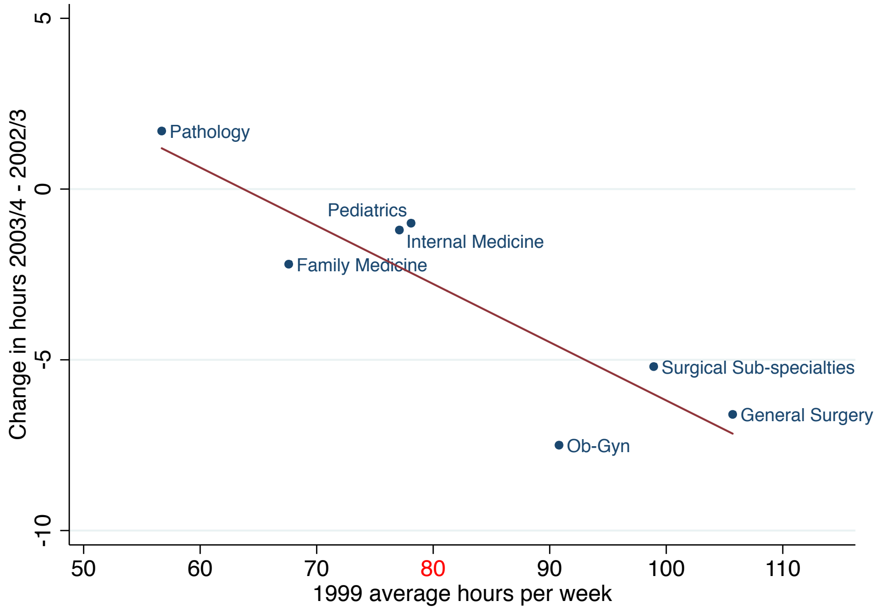
A. Age 30-34 B. Age 35-39 C. Age 40-44

II. Men

A. Age 30-34 B. Age 35-39 C. Age 40-44

Source: U.S. Census 2000; American Community Survey 2005-2012 Note: The sample is limited to individuals with a Bachelor's degree or more who are in the labor force and have valid information for their occupation. An occupation is sufficiently large to include if it has at least 300 female observations in each age grouping. The fraction of men who usually work 50 or more hours per week is computed for each occupation and age grouping. Similarly, the average number of children (including step-children and adopted children) is computed for each occupation, age grouping, and gender. The line of best fit is weighted by occupation cell size.

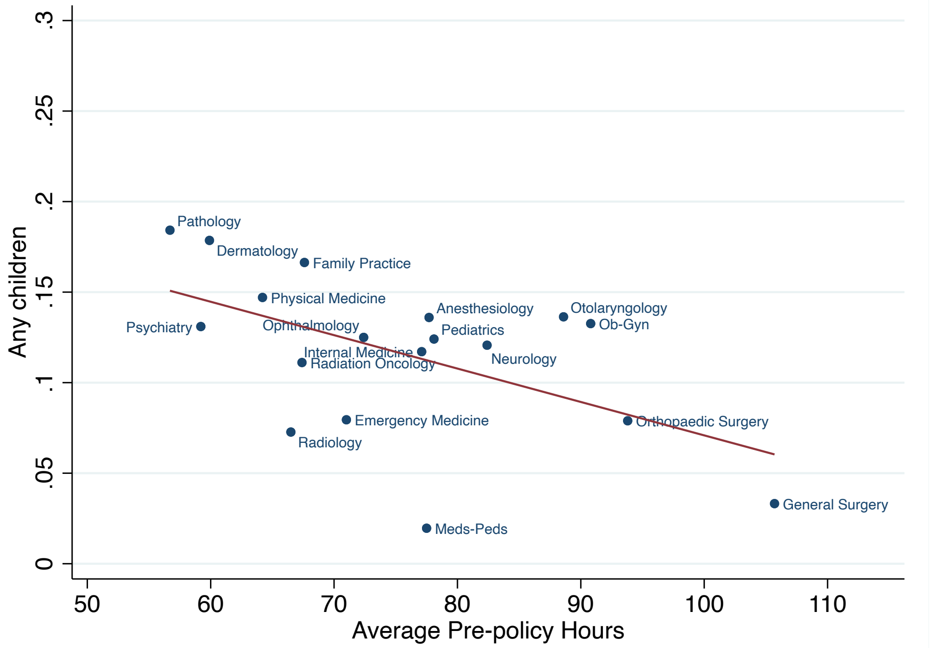
Figure A.2: Relationship between 1999 Hours and the Change in Hours 2002/3-2003/4, by Specialty



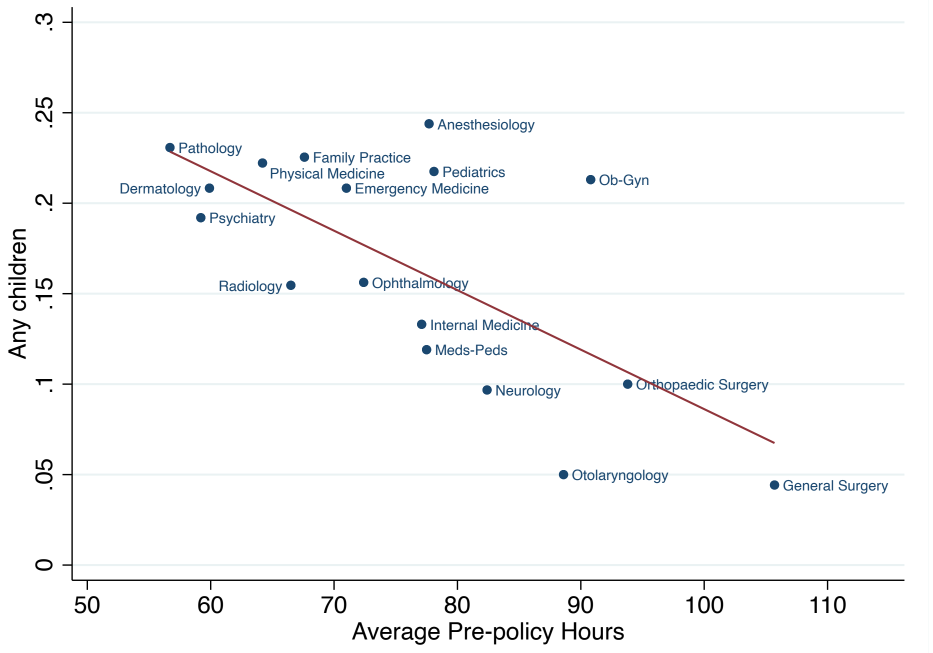
Source: Baldwin Jr et al. (2003), Landrigan et al. (2006), private correspondence with author. Note: This figure plots the average number of hours worked per week for second year medical resident physicians in 1999 on the x-axis. The change in average hours per week between residency years 2003/4 and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

Figure A.3: Pre-Policy Fertility and Pre-Policy Hours Requirements: Any Children

A. CA Vital Statistics Data



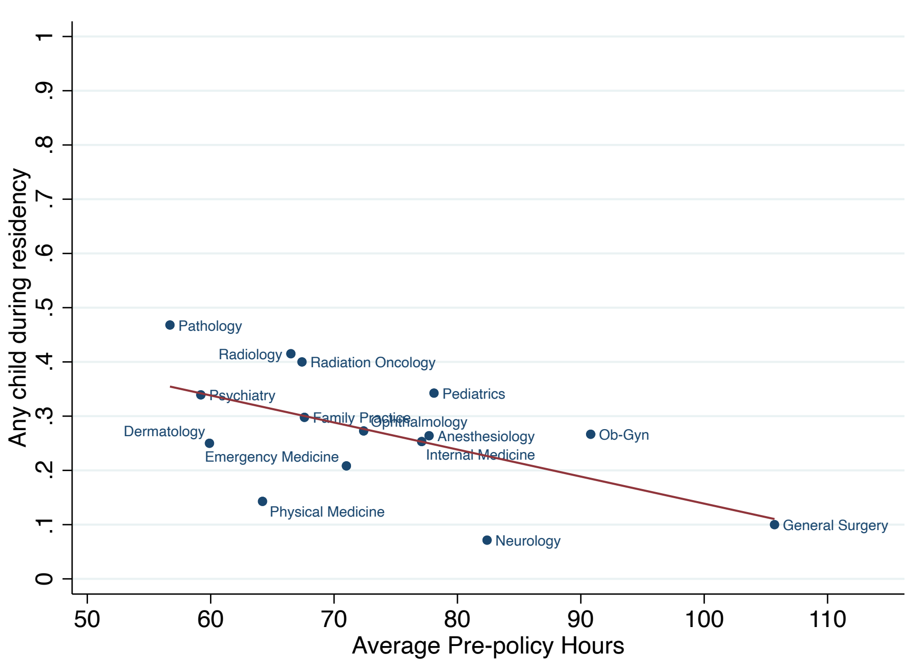
B. TX Vital Statistics Data



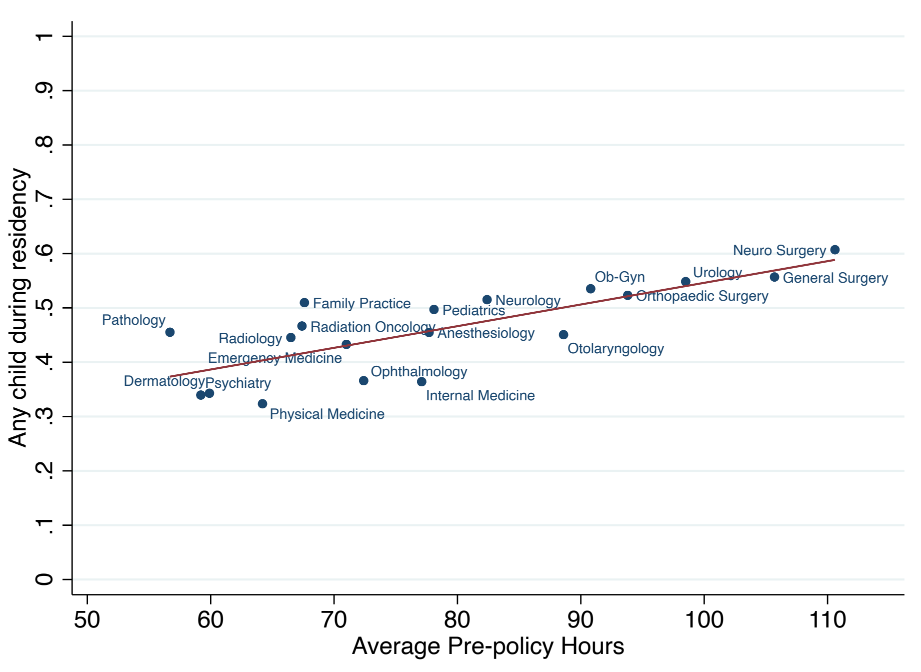
Source: AMA Physician Masterfile, California and Texas Vital Statistics birth records, Baldwin Jr et al. (2003) Note: This figure plots the the fraction of residents who have any children during the first three years of residency against the average pre-policy hours, by specialty. The CA fertility sample includes female U.S. medical school graduates from years 1993 through 2010 who completed their first three years of residency training in CA . 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.

Figure A.4: Pre-Policy Fertility and Pre-Policy Hours Requirements: Young Physicians Survey

A. Women



B. Men



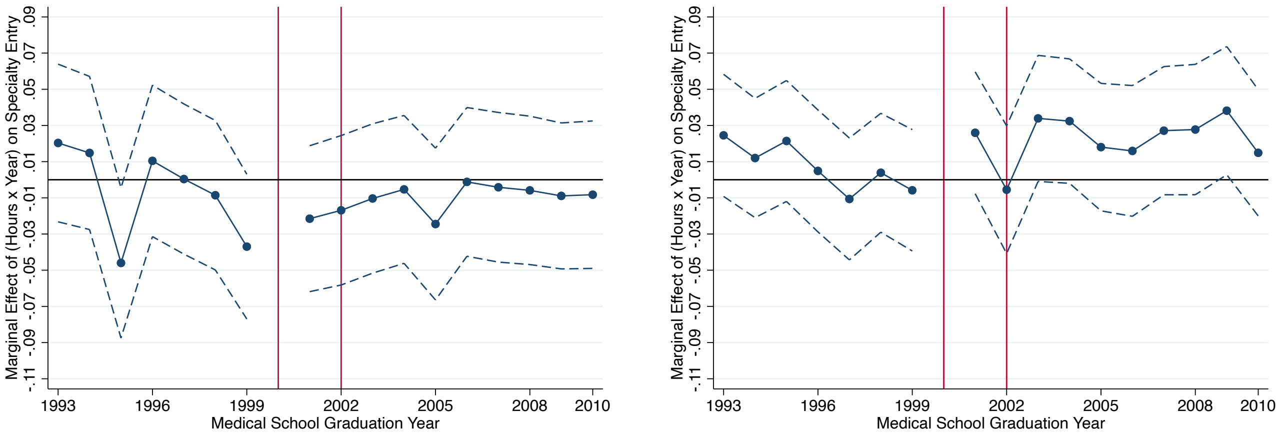
Source: Young Physicians Survey, 1991; Baldwin Jr et al. (2003). Note: This figure plots the fraction of physicians in each specialty who reported having children during residency training. The detailed specialties in the survey were crosswalked to the broad specialties used in Baldwin Jr et al. (2003). In Panel A (B), the sample is restricted to specialties with at least 10 female (male) respondents in the YPS.

Figure A.5: The Effect of the Reform on Specialty Choice: Event Study

I. California

A. Women

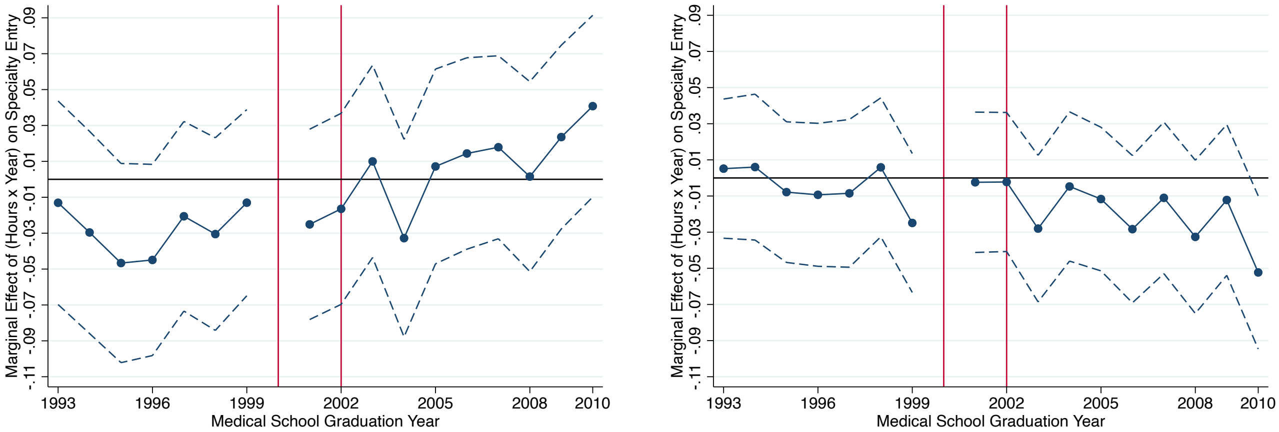
B. Men



I. Texas

B. Women

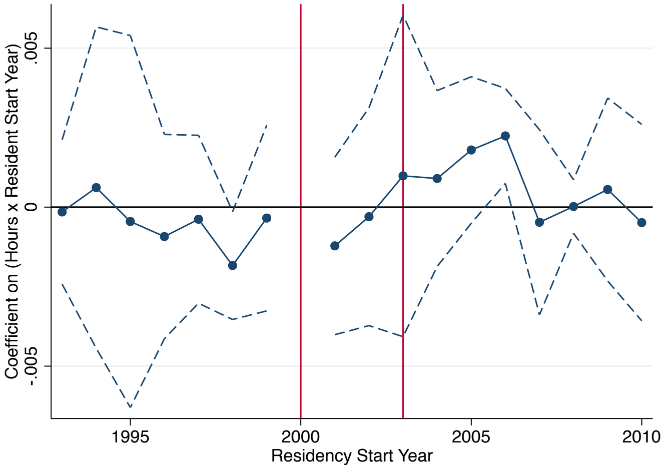
B. Men



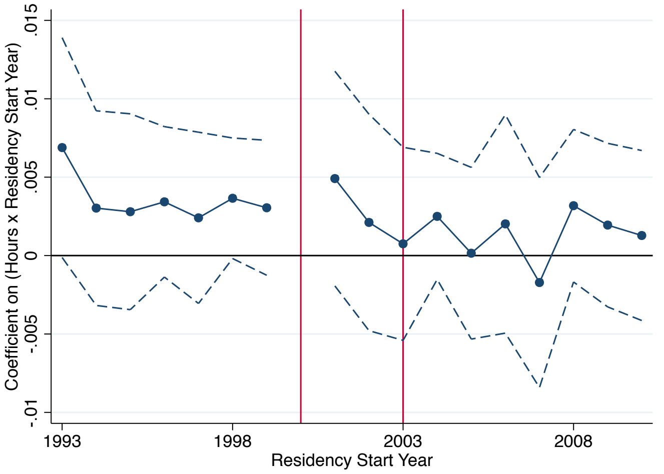
Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This figure plots the average marginal effects associated with the coefficients from the conditional logit event study. The sample is CA and TX. medical school graduates, 1993-2010. The dependent variable is specialty outcome and the explanatory variables are specialty fixed effects, and interactions of specialty pre-policy average hours with medical school cohort fixed effects. Cohort 2000 is omitted as the reference year. The solid line plots the average marginal effects of the interaction term (Hourss,1999 × Year). The dashed lines plot the 95% confidence intervals based on heteroskedastic robust standard errors.

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

A. California



B. Texas



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 hours 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 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.

Table A.1: Summary Statistics

	(1) Full Sample	(2) USMG Sample	(3) CA USMG Sample	(4) TX USMG Sample
Female	0.44	0.44	0.47	0.43
Age at medical school graduation	27.90	28.27	28.33	28.11
	(3.67)	(3.35)	(3.05)	(3.31)
Female	27.73	28.15	28.26	27.86
	(3.69)	(3.44)	(3.11)	(3.23)
Male	28.04	28.36	28.39	28.30
	(3.64)	(3.28)	(2.99)	(3.35)
U.S. Born	0.63	0.83	0.73	0.81
Female	0.63	0.83	0.74	0.81
Male	0.63	0.83	0.72	0.81
Ranked Medical School	0.33	0.48	0.68	0.47
Female	0.33	0.48	0.69	0.47
Male	0.32	0.48	0.67	0.47
Foreign Medical School	0.24	-	-	-
Osteopathic Medical School	0.08	-	-	-
N	414,075	281,477	26,633	16,401
Female	181,861	124,817	12,616	7,106
Male	232,214	156,660	14,017	9,295
Fertility During Residency (Women Only)				
Number of Children	-	-	0.13	0.20
			(0.36)	(0.44)
Any Children	-	-	0.12	0.19
			(0.32)	(0.39)
N			12,580	7,093

Source AMA Physician Masterfile, California and Texas Vital Statistics birth records, Baldwin Jr et al. (2003). Note: The Full Sample includes all medical school graduate from years 1993 to 2010, including foreign medical school graduates and osteopaths. The USMG sample includes U.S. medical school graduates from years 1993 through 2010. The CA (TX) sample includes female USMGs from years 1993 through 2010 who completed their first three years of residency training in California (Texas). 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 A.2: Effect of the Reform on Specialty Choice: Coefficients

Dependent Variable: Specialty Outcome	Women			Men		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: California						
Avg Weekly Hours × Start 2003-2010	-0.045	-0.075	0.094	0.447***	0.566***	0.293**
	(0.174)	(0.175)	(0.196)	(0.143)	(0.144)	(0.146)
Avg Weekly Hours × Start 2001-2002	-0.297	-0.314	-0.233	0.106	0.140	0.049
	(0.270)	(0.273)	(0.284)	(0.223)	(0.226)	(0.226)
# individuals	12,616	12,616	12,616	14,017	14,017	14,017
Panel B: Texas						
Avg Weekly Hours × Start 2003-2010	0.814***	0.685***	0.936***	-0.408**	-0.367**	-0.719***
	(0.219)	(0.220)	(0.249)	(0.170)	(0.171)	(0.178)
Avg Weekly Hours × Start 2001-2002	0.090	0.029	0.173	0.041	0.037	-0.128
	(0.346)	(0.348)	(0.360)	(0.250)	(0.254)	(0.255)
# individuals	7,106	7,106	7,106	9,295	9,295	9,295
Specialty FE	X	X	X	X	X	X
Age, medical school ranking		X			X	
Primary Care / OBGYN time trends			X			X

Source: AMA Physician Masterfile Baldwin Jr et al. (2003). Note: This table reports the results of maximum likelihood estimation of a conditional logit model in which the baseline specification has specialty outcome as the dependent variable and the explanatory variables include specialty fixed effects and specialty hours interacted with an indicator for graduating medical school 2003 onwards and an indicator for graduating medical school in 2001/2. The coefficient on the interaction terms are reported. Panels A and B present results for the sample of USMGs who completed their first three years of residency in CA and TX, respectively. Heteroskedastic robust standard errors are in parentheses.

Table A.3: The Effect of the Reform on Specialty Composition

Dependent Variable: Specialty Composition	Women		Men	
	(1) Age	(2) Ranked	(3) Age	(4) Ranked
Panel A: California				
Avg Weekly Hours × Start 2003-2010	0.011*	-0.178	0.012*	-0.248
	(0.006)	(0.120)	(0.006)	(0.177)
Avg Weekly Hours × Start 2001-2002	-0.011	-0.124	-0.005	-0.210
	(0.006)	(0.124)	(0.009)	(0.156)
# individuals	12,616	12,616	14,017	14,017
# observations	352	352	360	360
Panel B: Texas				
Avg Weekly Hours × Start 2003-2010	0.025*	-0.210	0.004	-0.030
	(0.013)	(0.222)	(0.011)	(0.141)
Avg Weekly Hours × Start 2001-2002	0.017	-0.319	0.004	0.069
	(0.014)	(0.319)	(0.011)	(0.112)
# individuals	7,106	7,106	9,295	9,295
# observations	346	346	358	358
Specialty FE	X	X	X	X

Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This table reports the coefficients from a difference-in-differences regression in which the dependent variable is demographic composition of individuals in a specialty from a particular medical school cohort. The independent variables are specialty fixed effects, medical school cohort fixed effects, and the interactions between pre-policy specialty hours and two indicators: for cohorts 2003 onwards and for cohorts 2001/2.

B Conceptual Framework for Specialty and Fertility Choice

To guide our understanding of how a reduction in hours can potentially affect occupational and fertility decisions, in the following simple conceptual framework, I model a medical student's choice of medical specialty and whether to have children during residency.[21] Then, I show how these choices can change as a result of the duty hour reform. This framework generates the following insights. First, if women incur an additional disutility of hours worked when they have children, then they will be less likely than men to enter high hours specialties. Second, conditional on entering high hours specialties, women will be less likely than men to have children during residency. Third, if the disutility of hours worked when women have children is convex, then it is possible that women are more likely than men to enter the high hours specialties after a reduction in hours. Fourth, depending on distributional assumptions regarding model parameters, the fertility rate of the high hours specialties can rise, fall or stay the same after a reduction in hours.

I formally represent the decision as a static, unconstrained utility maximization problem. 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 children. 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, here the focus of which are hours worked during residency and wages post-residency: (h_s, w_s).[22] This utility specification embeds a few key components. First, there is individual heterogeneity in the relative valuation of non-market time and wages. Specifically, individuals relative valuation is captured by β_i, with $\beta_i \sim [0, b]$. Second, there is also individual heterogeneity in valuation of children, where individuals obtain additional utility π_i, where $\pi_i \sim [-p, p]$, if they have a child. Since π_i can take on negative values, some individuals

---

[21] I abstract from the decision to enter the medical profession.

[22] 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.

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 $h_L < h_H$.[23]

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^*.[24] 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^*.[25] Higher values of π reflect a higher valuation of children and, all else equal, make individuals weakly more likely to have children.

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. We consider the 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 choose the high hours specialty. Depending on the value of π_i

---

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

[24] 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 $\beta_C^* < \beta_N^*$.

[25] 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 $\pi_L^* < \pi_H^*$.

an individual chooses whether to have children. For individuals with $\beta_C^* < \beta_i < \beta_N^*$, the disutility of hours is such that they may or may not choose a high hours specialty. 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 hours reform goes into effect? This manifests as a decrease in hours in the high hours specialty, h_H. For men, the duty hours reform causes a decline in β^*, which induces entry into H. The decline in hours does not change male fertility choices, as these are determined independently of hours worked. The implications of the duty hours 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 decrease in h_H also increases the β cutoffs, decreases the π_H^* cutoff, and leaves the π_L^* cutoff unchanged.

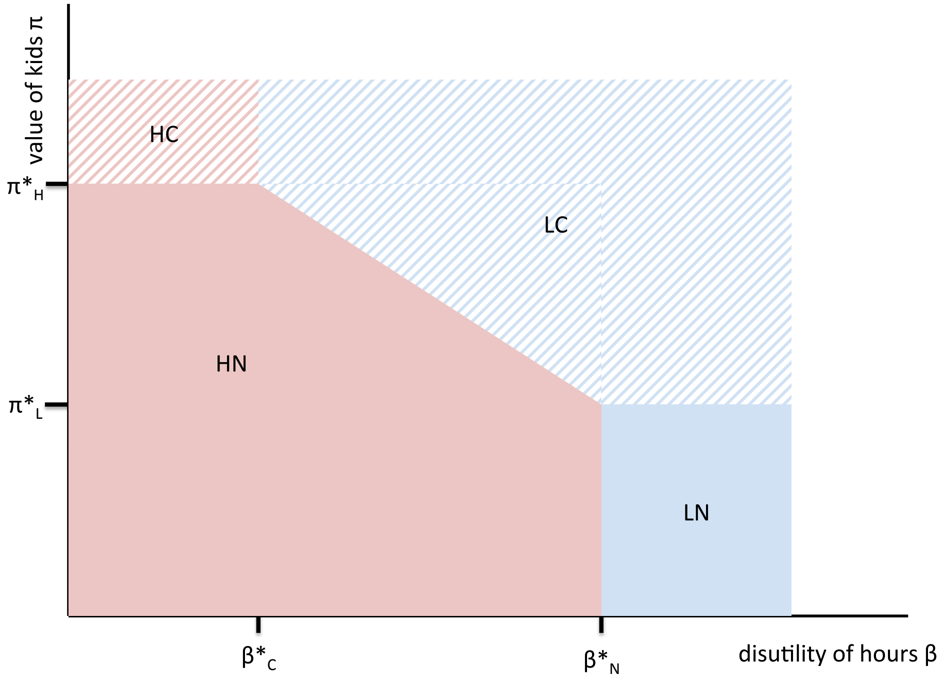
A few key insights emerge from this simple framework. First, the decrease in h_H induces net entry into H, due to the shift upward of the β cutoffs, which serve primarily to expand the population of individuals willing to enter H specialties, from both LC and LN. I denote these individuals who change 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.

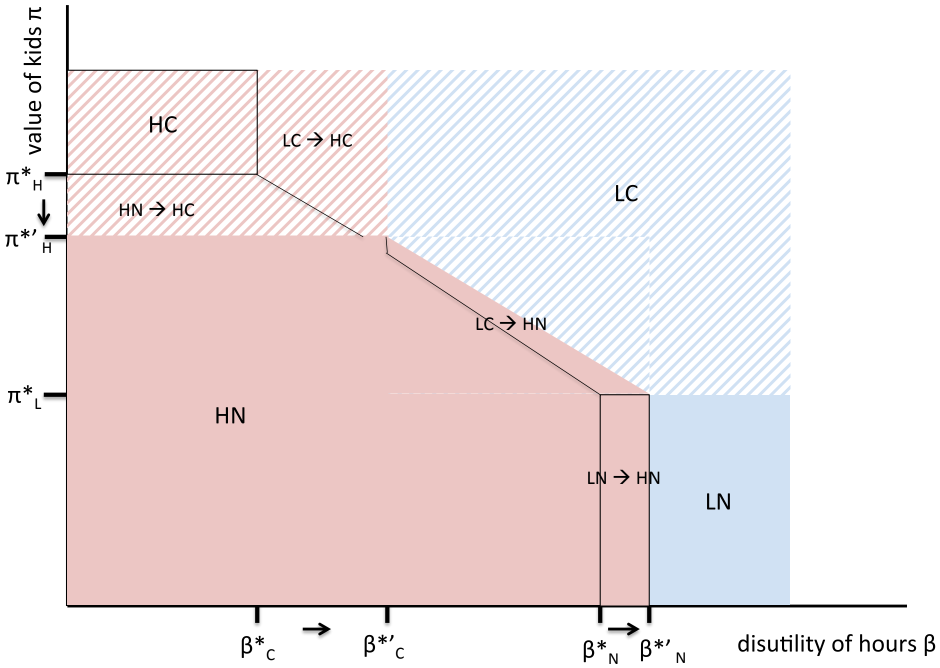
The last insight concerns the direction of the change in the fertility *rate* in H and L. Given the above discussion, the direction of the change hinges critically on the composition of the new entrants to H, the composition of those individuals who exit L, and the magnitude of the fertility compliers. In H, there are two potentially offsetting effects: the fertility compliers, who increase the fertility rate, and new entrants who choose HC or HN, which could serve to 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. 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 Appendix Figure B.1 Panel A are relatively sparsely populated. With the introduction of the duty hours reform, individuals who are induced to enter H may be disproportionately drawn from the fertility-marginal specialty compliers, who have intermediate valuations of children and disutility of hours worked, and switch to not having children when they enter H. If this is the case, then the fertility rate in H could stay the same or even fall in response to the duty hours reform.

Figure B.1: Graphical Depiction of Conceptual Framework

A. Example of Initial Allocation



B. Allocation After Duty Hours 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 individuals 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 not to have children.

C State Heterogeneity in the Effects of the Reform on Specialty Choice

This section investigates state-level heterogeneity in the effect of the duty hours reform. I estimate the effect of the reform on specialty entry in every other U.S. state with sufficient medical residents.[26] In order to do so, I assign individuals in the USMG sample to their last state of medical residency training (including fellowship).[27] I explore the relationship between the state-specific entry effects and four pre-policy state attributes: the fraction of residents in the state who are female; the competitiveness of residency programs in the state, as measured by the fraction of residents from ranked medical schools; the usual hours per week of medical residents in the state, as measured in the 2000 Census; and the fraction of female medical residents with children in the state, as measured in the 2000 Census.[28] The correlations are presented in Appendix Figure C.1, with labels for Texas and California. Taken together, it appears that the reform encouraged women to enter time-intensive specialties in states in which there was more room for the fraction female to grow and it was easier to obtain a residency position. Additionally, during the pre-reform period, female residents in these states tended to have more children, which could be attributed to women doing residency in specialties for which it is easier to have children, women exhibiting differing fertility preferences, or both. California is among the top third in female representation and competitiveness, and the bottom third in fertility, consistent with the observed low effect of the reform on female specialty entry. Texas, on the other hand, is among the top third in the effect of the

---

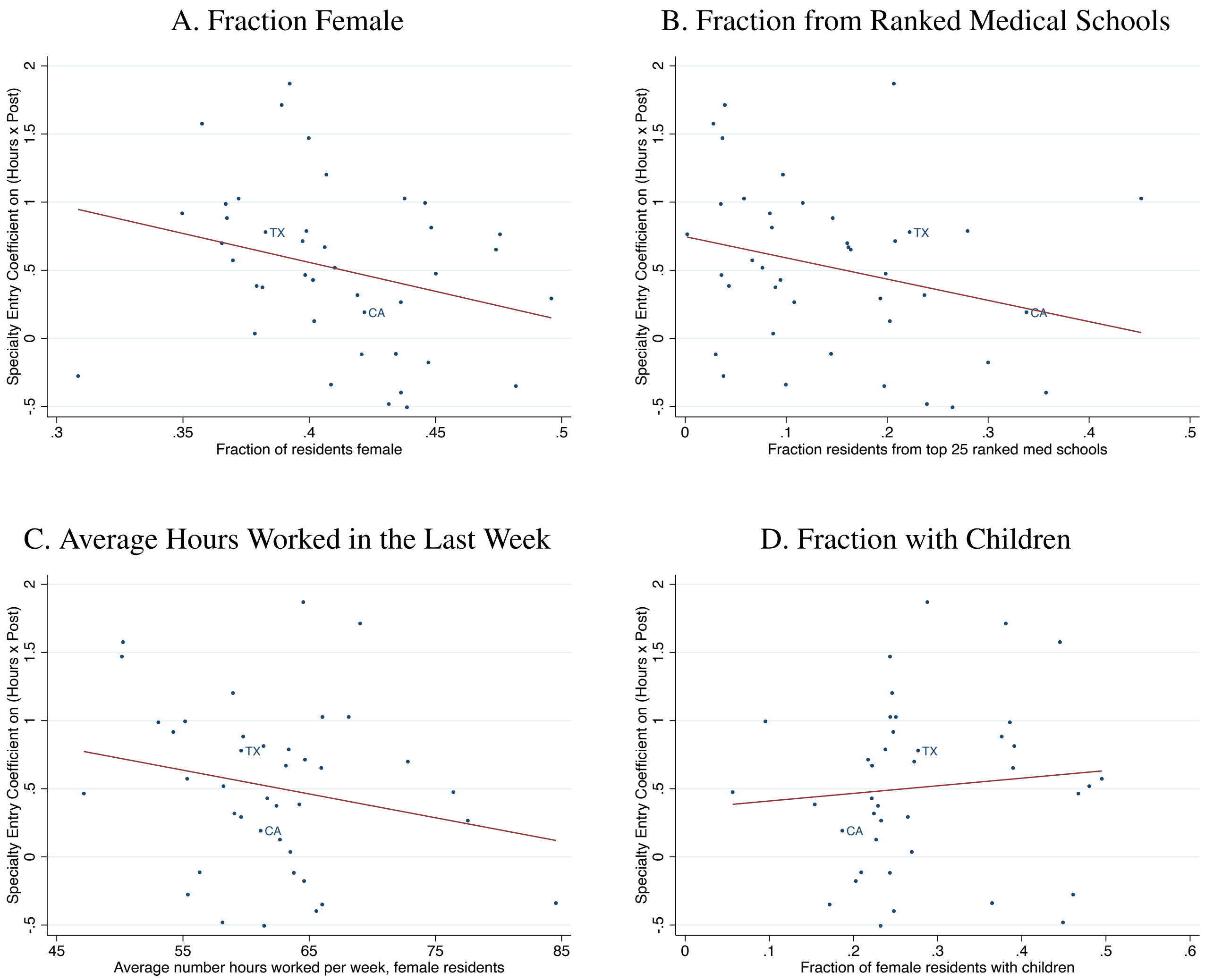
[26] For this analysis, I drop the 0.50 percent of individuals who do not have information regarding their last residency training state. I additionally exclude Alaska, Delaware, Idaho, Maine, Montana, New Hampshire, Nevada, North Dakota, South Dakota, Vermont, and Wyoming due to insufficient sample size. The sample includes Puerto Rico and Washington, D.C.

[27] The last state of residency training can potentially change in response to the duty hours reform, if an individual chooses to pursue additional training due to the reduction in hours and those programs are disproportionately found in select states. I confirmed that the estimates using last state of residency training and first state of residency training are similar in California and Texas using the richer data in the CA/TX Masterfile, which contains both variables. There are, however, some slight differences between the CA/TX coefficients in Appendix Table A.2 and those plotted below.

[28] It is also possible that states varied in the extent to which they complied with the regulations. If prospective medical residents were aware of state differences in compliance, then the states in which the hours were reduced considerably could experience greater entry of women. Due to limited data on hours by specialty and state pre- and post-reform, I cannot explore this channel.

reform on specialty entry, middle third in female fertility and competitiveness, and bottom third in female representation.

Figure C.1: State-Specific Female Entry Effects and Pre-Policy State Attributes


A. Fraction Female

B. Fraction from Ranked Medical Schools

C. Average Hours Worked in the Last Week

D. Fraction with Children

Source: AMA Physician Masterfile, Census 2000, Baldwin Jr et al. (2003). Note: This figure plots state-specific effects of the reform on women's specialty entry against various state-level characteristics: the fraction of residents in the state who are female; the competitiveness of residency programs in the state, as measured by the fraction of residents from top ranked medical schools; the usual hours per week of medical residents in the state, as measured in the 2000 Census; and the average number of children of medical residents in the state, as measured in the 2000 Census.