CENTER FOR ECONOMIC BEHAVIOR & INEQUALITY

CEBI WORKING PAPER SERIES

Working Paper 03/24

FAMILY-FRIENDLY JOBS AND OCCUPATIONAL SORTING ACROSS GENDER: EVIDENCE FROM INTRODUCTION OF A MAXIMUM 40-HOUR WORKWEEK

Kathrine Aaby Lorentzen

ISSN 2596-447X

CEBI
Department of Economics
University of Copenhagen
www.cebi.ku.dk

Family-friendly Jobs and Occupational Sorting across Gender: Evidence from Introduction of a Maximum 40-hour Workweek*

Kathrine Aaby Lorentzen †

Version: February 9, 2024

Abstract

I study the impact of changes in the family-friendliness of jobs on occupational choice across gender in the labor market for physicians. I utilize a Danish working hour reform from 1981, which imposed a maximum 40-hour workweek for junior hospital physicians. To do so, I digitize archival data and combine it with administrative data. I find that the reform affected some specialties more than others, with the least flexible (acute) specialties being the most affected. I exploit these differing treatment intensities across specialties in order to estimate the impact of family-friendly hours constraints on specialty choice across gender in a dose-response difference-in-differences design. My results suggest that the fraction of females in a specialty increased more where introduction of a 40-hour workweek had more bite. In addition, I find that female fertility increased more in highly affected specialties. I do not find similar effects for males. My results suggest that preferences for working hours are important for occupational choice across gender and hence gender inequality in wages.

*I thank my supervisor Torben Heien Nielsen as well as Frederik Plesner Lyngse, Ida Lykke Kristiansen, Johanna Kokot, the PhD students at CEBI, seminar participants at University of Copenhagen, DGPE Workshop and EuHEA PhD & Supervisor Conference for helpful comments and discussions. I thank The Danish Association of Junior Hospital Doctors for lending me copies of historical collective agreements and for providing me with helpful institutional insights. Furthermore, I thank Carl Benjamin Simpson for assisting me with digitization of archival data.

The paper builds upon my master's thesis from August 2023.

I gratefully acknowledge funding from the Danish Independent Research Fund (grant no. 9061-00035B). The activities of CEBI are financed by the Danish National Research Foundation (grant no. DNRF134).

†University of Copenhagen, Center for Economic Behavior and Inequality. E-mail: kalo@econ.ku.dk

1 Introduction

The gender wage gap has decreased significantly over the last hundred years. Nevertheless, substantial gender inequality in wage rates remains a worldwide phenomenon (Kleven, Landais, and Søgaard, 2019). Convergence in wage rates up until now can largely be attributed to advances in the relative human capital levels of women and men. Today, however, the gender education gap has been reversed and differences in labor-market experience have become muted. Instead, occupation and industry make up some of the largest measured causes (Blau and Kahn, 2017). What explains this link between occupation and industry and gender inequality in wages? According to the Goldin (2014) model the answer is a combination of convex wage schedules and differences in preferences for working hours which lead to a gender wage gap in equilibrium through occupational sorting. However, there is still limited empirical evidence on the interplay of the model's mechanisms. One important exception is Wasserman (2023b) who shows that capping working hours at 80 hours causally affect occupational sorting among female medical residents in the US. Using average hourly earnings, she simulates the effect of the implied changes in specialties' gender composition on the profession's gender wage gap.

In this paper, I add to this empirical evidence on Goldin (2014)'s theory by studying a reform, which imposed hours constraints on a whole different margin: a cap at 40 weekly working hours. Furthermore, I exploit that this reform played out in the Danish labor market for physicians – a highly regulated labor market ideal for studying wage structures. As such, I study each of the Goldin (2014) model components in turn.

First, analyzing wage schedules set out in collective agreements, I document that physicians' wage schedules were in fact convex. I show that the scope for convex returns increased with the type and frequency of shifts and the potential for overtime hours. Intuitively, convexities arise when providing employees with flexibility is costly for the production of health. That is, when postponing treatment to the next day could have fatal consequences for the patient. One important aspect of the Danish labor market for physicians during the considered period is that wages were determined exclusively by the collective agreement. As such, wages increase deterministically with tenure and vary according to the particular hours worked only (e.g. the type/frequency of shifts). This allows me to unambiguously conclude that the observed sorting into specialties could have resulted from selection on convex returns.

Second, I study whether preferences for working hours drive physicians' specialty choice. To do so, I use a Danish working hour reform from 1981 which imposed a maximum 40-hour workweek for junior hospital physicians. To find out how the reform affected physicians' workweek, I digitize archival data on posted positions across medical specialties dating back to 1972. Absent data on pre-policy hours, I infer how binding the reform was across specialties from the reform-induced percentage change in the number of posted positions in a specialty. That is, I exploit that the 40-hour workweek acts as a labor demand shock. Following Wasserman (2023b), I employ an estimation strategy that exploits that the working hour reform lead to greater reductions in the workweek within some specialties than others and use these differing treatment intensities across specialties to compare the specialty choices of individuals who chose their specialty prior to the reform to the choices of individuals who made their choice after the reform.

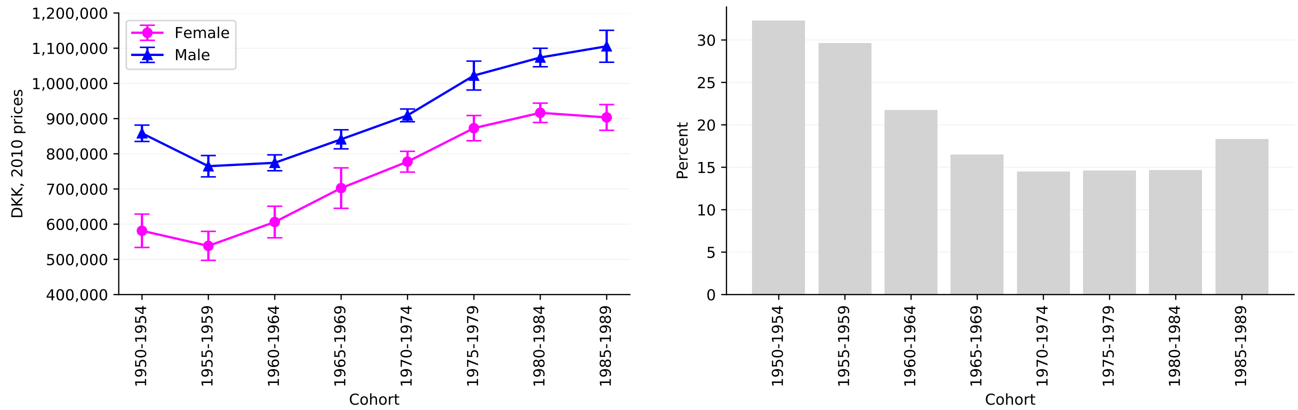
I show that highly affected specialties were more likely to be acute, as measured by their propensity to admit patients during the weekend. I find that the fraction of women in a specialty increased more where introduction of a 40-hour workweek had more bite. Specifically, I find that increasing the reform intensity by 10 percentage points leads to a 1.9 percentage point increase in the fraction of females in a specialty. A back-of-the-envelope calculation suggests that reducing the workweek from 60 to 40 hours leads to an increase in the specialty's female share of 9.5 percentage points.

Lastly, I study how the introduction of the maximum 40-hour workweek affected family formation. Previous research has shown that new mothers self-select into more family-friendly occupations after child arrival (Kleven, Landais, and Søgaard, 2019) and are more likely to leave the labor force if they had an inflexible job (Herr and Wolfram, 2012). Women's preferences for working hours might thus stem from their desire to have children. Consistent with this hypothesis, I find that female fertility increased the most within specialties highly affected by the maximum 40-hour workweek. At the same time, male fertility is unaffected.

My paper contributes to an extensive literature on gender disparities in wages and their determinants (see e.g. Blau and Kahn, 2017 for a review). Most closely related to the present study is the previously mentioned paper by Wasserman (2023b) in which she leverages a working hour reform which capped the maximum workweek at 80 hours to estimate the causal effect of hours constraints on medical residents' specialty choices. She finds that women, but not men, respond by entering affected specialties at a higher rate. I contribute to this work by studying a reform which capped the workweek at 40 hours and hence mimics a shock to family-friendliness of occupations. I show that Wasserman (2023b)'s findings carry over when hours are capped at 40 rather than 80 hours. Bolotnyy and Emanuel (2022) study the gender earnings gap among bus and train operators and find that even in this unionized workplace there is an earnings gap of 11 percent. They find that the gap is caused by a lower propensity to work overtime and a higher propensity to take unpaid time off among women. The setting I consider in this paper is similar to theirs in the sense that I too consider a unionized workplace in which wage-setting is non-discriminatory by construction. I contribute to this work by studying the impact of hours constraints on occupational choice in another unionized labor market with gender disparities in earnings (see fig. 1.1).

My paper furthermore contributes to the literature on the interaction between labor supply and fertility. Numerous studies have documented a negative correlation between female labor supply and fertility and researchers have studied the potential causal mechanisms accountable. Because labor supply and fertility are likely simultaneously determined, researchers have come up with valid instruments for completed fertility in order to estimate the causal impact of fertility on labor supply. These instruments include the sibling sex mix as in Angrist and Evans (1998) and the arguably unanticipated arrival of twins as in Rosenzweig and Wolpin (1980). Recent research has found evidence of causal effects working in the other direction: from labor supply to fertility. For example, Fadlon, Nielsen, and Lyngse (2022) find that early career sorting causally affect fertility while Wasserman (2023a) find that a cap on hours at 80 hours affects female fertility timing. Wasserman (2023a) finds evidence that fertility decisions both respond to occupations' time demands as well as play a role in the *choice* of occupation. I contribute to this work by studying how a maximum 40-hour workweek affects completed fertility for both women *and* men.

This paper is organized as follows. Section 2 describes the institutional setting and the working hour reform. Section 3 presents the conceptual framework and shows the convex wage structures that were present pre-reform. Section 4 presents the data. Section 5 explains the empirical strategy. Section 6 presents the results. Section 7 concludes.



(a) Average Earnings 30 Years after Graduation

(b) Earnings Gap 30 Years after Graduation

Figure 1.1: Earnings 30 Years after Graduation

Notes: Figure (a) shows average earnings 30 years after graduation (e.g. earnings are measured in 1980 for the 1950 cohort etc.). Earnings are in 2010 prices to ensure comparability over time. Figure (b) shows the corresponding gender earnings gap in percent of average male earnings.

Source: *Own calculations based on data from Statistics Denmark and the Danish Authorization Register*

2 Institutional Setting

In this section, I describe the aspects of the Danish labor market for physicians necessary to understand the 1981 working hour reform. This reform was the result of a collective bargaining agreement and I describe its background, content and show its direct implications for the demand for physicians at Danish hospitals. Finally, in section 2.4, I describe the career path of a physician at the time of the reform. This section sets the stage for how one should think about occupational choice among physicians and at what point in the career this choice is made.

2.1 The Danish Labor Market for Physicians: A Unionized Labor Market

In Denmark, collective bargaining between unions and employer organizations through which conditions such as working hours and wages are determined is common (Kreiner and Svarer, 2022). In fact, it is a key aspect of the so-called 'Danish labor market model'. The Danish labor market for physicians is no exception.

The interests of Danish physicians are protected by the Danish Medical Association (Danish: *Den almindelige danske lægeforening*[1]) henceforth DADL. DADL is an NGO which (during my sample period and till this day) consists of the following three labor unions (Den almindelige danske Lægeforening, 1977; Lægeforeningen, 2022):

- The Danish Association of Junior Hospital Doctors (Danish: *Foreningen af yngre læger*)
- The Danish Organization of General Practitioners (Danish: *Praktiserende Lægers Organisation*)
- The Danish Association of Medical Specialists (Danish: *Foreningen af Speciallæger*)

The working hour reform (which I am going to describe in section 2.2) was the result of a Collective Agreement which covered members of The Danish Association of Junior Hospital Doctors (Danish: Foreningen af yngre læger,

[1] In 2007 the name was changed to *Lægeforeningen* (Winther Jensen, 2007)

henceforth FAYL). At the time of the reform in 1981 FAYL's members were physicians in subordinate positions employed at hospitals, institutes and the like (Den almindelige danske Lægeforening, 1981).[2]

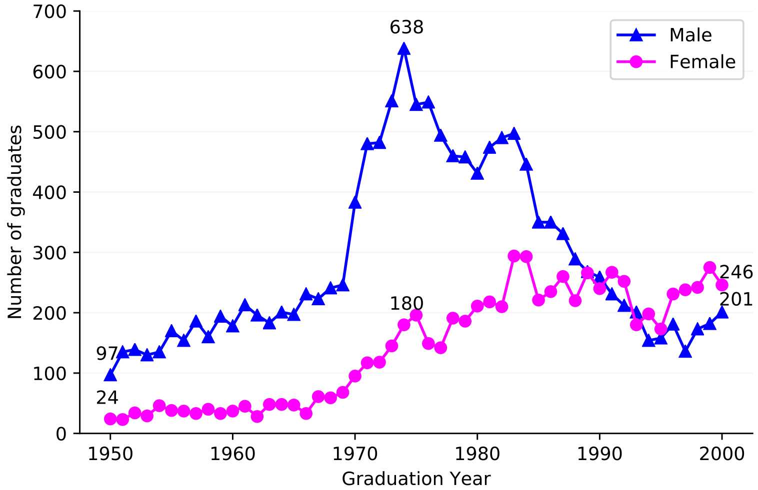
2.2 The 1981 Collective Agreement

The working hour reform was the result of a controversial collective bargaining agreement. The parties (the Danish Association of Junior Hospital Doctors and the public employer) negotiated for almost a full year. The point of contention was introduction of a 40-hour workweek - or rather the specifics of its implementation.

The idea about a 40-hour workweek had long been discussed but attempts at reaching an actual agreement had proven unsuccessful. At a board of representatives meeting in the fall of 1975, FAYL adopted the aim of the 40-hour workweek. The motivation was twofold: 1) a normalization of the workweek and 2) solidarity with large medical school graduation cohorts (Steensen, 1981).

Throughout the 1960's, Danish universities experienced explosive growth in their student intake and the situation was especially challenging at the faculties of medical sciences. Despite the large cohorts, universities were not allowed to reject qualified applicants throughout the 1960's and entry limitations were not in place until 1976-77 (Hansen, 2015). That the universities could not reject applicants was a political order rooted in the politics of the welfare state at the time: free entry should ensure equality as well as the quality of the workforce in a time with high demand for labor (Hansen, 2018). Within medicine this lead to an excess supply of physicians. Since medical school took 6.5 years to complete, the large student intake in the 60's materialized as big graduation cohorts in the 70's. Figure 2.1 shows registered physicians by gender from 1950 until 2000. It is evident from the graph that the size of graduation cohorts more than doubled within a few years in the beginning of the 70's.

Figure 2.1: Number of Graduates by Year



Source: *Own calculations based on data from the Danish Authorization Register.*

FAYL hence entered the year of 1981 with a long list of structural problems. One item on that list was unemploy-

[2]Hence, for some of FAYL's members the term 'junior doctor' is somewhat misleading. In principle a member of FAYL could be a specialist physician in a non-permanent position. The permanent staff, e.g. chief physicians, however, were not part of FAYL and thus not covered by the collective agreement.

ment issues. The number of unemployed Danish physicians was 314 in January of 1981 while 800-900 physicians worked in Sweden (Andersen Nexø et al., 2004). At a board of representatives meeting held in November 1980, the chairman described these issues as follows:

> "[...] An increasing number of physicians who leave the university with large student debt at high interest rates, only to enter directly into meaningless unemployment, emigration or - in the best-case scenario - random, short term employment" (Steensen, 1981, p. 171).

FAYL entered into negotiations with a collective bargaining proposal which had approaching a 40-hour workweek as the main objective. A key feature of the proposal was that the 40-hour workweek should be approached through increases in employment (i.e. through job sharing), while ensuring a satisfactory wage level and postgraduate education (Christy, 1981c).

The public employer's collective bargaining proposal demanded an entirely new collective agreement. 24-hour shifts should be divided into 8-hour shifts and there were given no guarantees for employment effects. The proposal entailed significant reductions in wages but no guarantees for reductions in the workweek (Christy, 1981a). This proposal came as a shock to FAYL. The overarching worry was that 8-hour shifts would harm both continuity of patient treatment and the physicians' postgraduate education (Andersen Nexø et al., 2004).

A decisive factor of the 1981 collective agreement was that the umbrella organization The Danish Confederation of Professional Associations (Danish: *Akademikernes Centralorganisation*) henceforth AC) acted as the central negotiating body. Hence, FAYL was reliant on the other AC members' support. On the 24th of March 1981 negotiations between the public employer and AC broke down - heavily influenced by FAYL's conflict (Andersen Nexø et al., 2004). On March 31st the conflict escalated as the public employers warned about lockout (Christy, 1981b). Negotiations continued in the Conciliation Board which resulted in a compromise proposal. AC balloted their members and on May 8th the Public Conciliator announced that the proposal had been adopted by majority vote - despite that 98 percent of FAYL's voting members voted against the proposal. This lead more than 4.000 young physicians to engage in temporary renunciations of their authorizations in order to go on strike in week 21 of 1981. The public employer agreed to discuss "guidelines" and interpretations of the collective agreement's provisions. The new negotiations continued all the way until new year (Andersen Nexø et al., 2004).
The final collective agreement contained the following main components (Foreningen af yngre læger, 1982):

- A 40-hour workweek
- A "20-hour rule": 20 out of the 40 hours should be daytime duty placed in the interval from 08.00 am to 04.00 pm during weekdays (Monday-Friday).
- Overtime should be avoided. If this is not possible then it should be compensated with time off. If the extent of overtime work reaches a level where it can be converted into more positions, then this should be done.

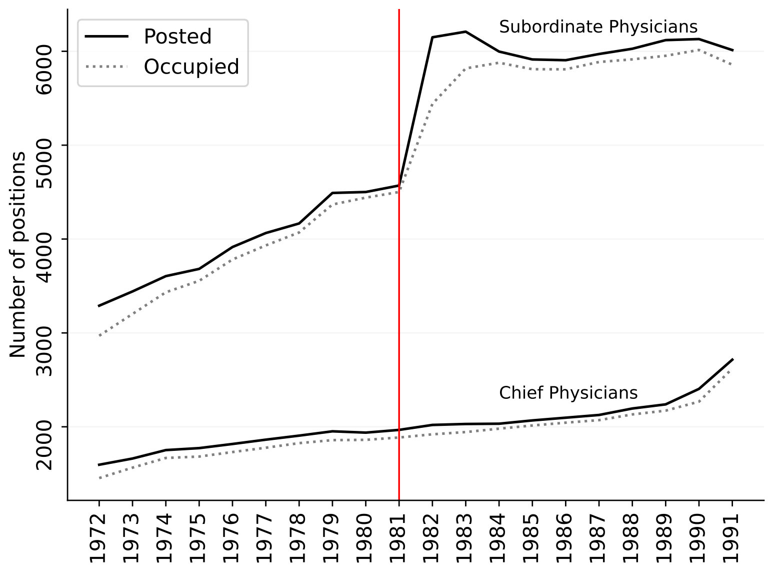
In the following section I show how these components of the collective agreement affected the demand for labor at Danish hospitals and I explain how I use this increase in labor demand to inform about the reform's differential impact across medical specialties.

2.3 Impact of the 1981 Collective Agreement: Change in Posted Positions

I do not have data on physicians' pre-policy hours. To measure the impact of the reform, I instead infer how binding the introduction of the maximum 40-hour workweek (and the 20 hour rule) was across specialties based on posted positions (Danish: *normerede stillinger*). That is, I exploit that the reform creates a positive labor demand shock. Furthermore, I exploit that this labor demand shock should differ across medical specialties depending on 1) the number of pre-policy hours in excess of 40 hours and 2) the extent to which scheduling of patient treatment is flexible. Hence, the extent to which the reform "binds" should depend on whether the specialty was characterized by long working hours pre-reform and the extent to which some of these hours could be shifted across daytime and non-daytime duty[3].

Figure 2.2 shows the number of posted and occupied positions at Danish hospitals from 1972 to 1991. The figure shows a sharp increase in the number of posted positions for subordinate physicians in the reform implementation year 1982 with approximately half of these positions becoming occupied within the same year. Hence, the graph shows how the reform transformed the tightness of the labor market (and thus solved the unemployment issues) in just one year. Furthermore, it is clear from the graph that the working hour reform affected FAYL's members and not the chief physicians who were covered by another collective agreement.

Figure 2.2: Posted and Occupied Positions at Danish Hospitals



Notes: This figure shows the number of posted (Danish: *normerede*) and occupied positions at Danish hospitals. The vertical red line indicates the pre-reform year 1981.

Source: *Own calculations based on data from Sundhedsstyrelsen (1972) - Sundhedsstyrelsen (1991)*

Dividing the 32 specialties that were present in 1981 into five specialty categories shows that increases in the number of posted positions were largest for the surgical specialties in absolute and relative terms. Surgical specialties experienced an increase in the number of posted positions of 44.30 percent between 1981 and 1982. At the same time, para clinical specialties were much less affected with an increase of just 8.40 percent. This is in accordance with the claim that the reform should bind the most in specialties characterized by long and inflexible

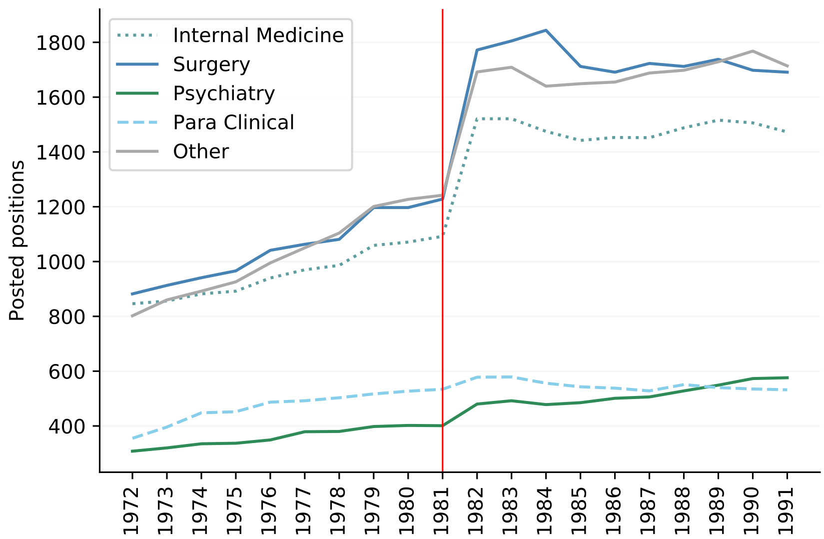
[3]The definition of daytime duty is duty in the interval from 08.00 am to 04.00 pm during weekdays cf. section 2.2

hours. Para clinical specialties are specialties that primarily assist other specialties in their treatment of patients e.g. radiologists that assist with x-ray examinations. These specialties are less acute and less duty-heavy than surgical specialties with most of the working hours taking place during daytime. For this reason, the 40-hour workweek and the 20-hour rule should be less binding in these specialties. Since the 20-hour rule demanded that 20 out of the 40 hours should be placed during daytime, only 20 hours were left per physician to cover non-daytime duty. Since postponing treatment till the beginning of the next daytime duty starting at 08.00 am is not possible for patients with very acute conditions, the working hour reform should bind the most in specialties with a large proportion of acute patients.

In order to check that this was in fact the case, I compute specialties' weekend admission rate. This measure is inspired by Doyle et al. (2015) who use it as a measure of the acuteness of diagnoses. As in Doyle et al. (2015), the closer the weekend admission rate is to 2/7, the more acute the specialties must be, since a weekend admission rate of 2/7 is equivalent to weekend admission being just as likely as weekday admission. Figure 2.4 shows that the weekend admission rate was indeed positively correlated with the percentage change in the number of posted positions in a specialty from 1981 to 1982.

In the empirical analysis I am going to use the reform-induced percentage change in a specialty's posted positions as a measure of reform intensity across specialties. The idea is that this measure is informative about a specialty's average pre-policy working hours. Because positions only take on integer values, this measure will be most accurate for higher numbers of pre-policy hours. Figure A.1 in the appendix plots this relationship for different hypothetical values of pre-policy average weekly working hours.

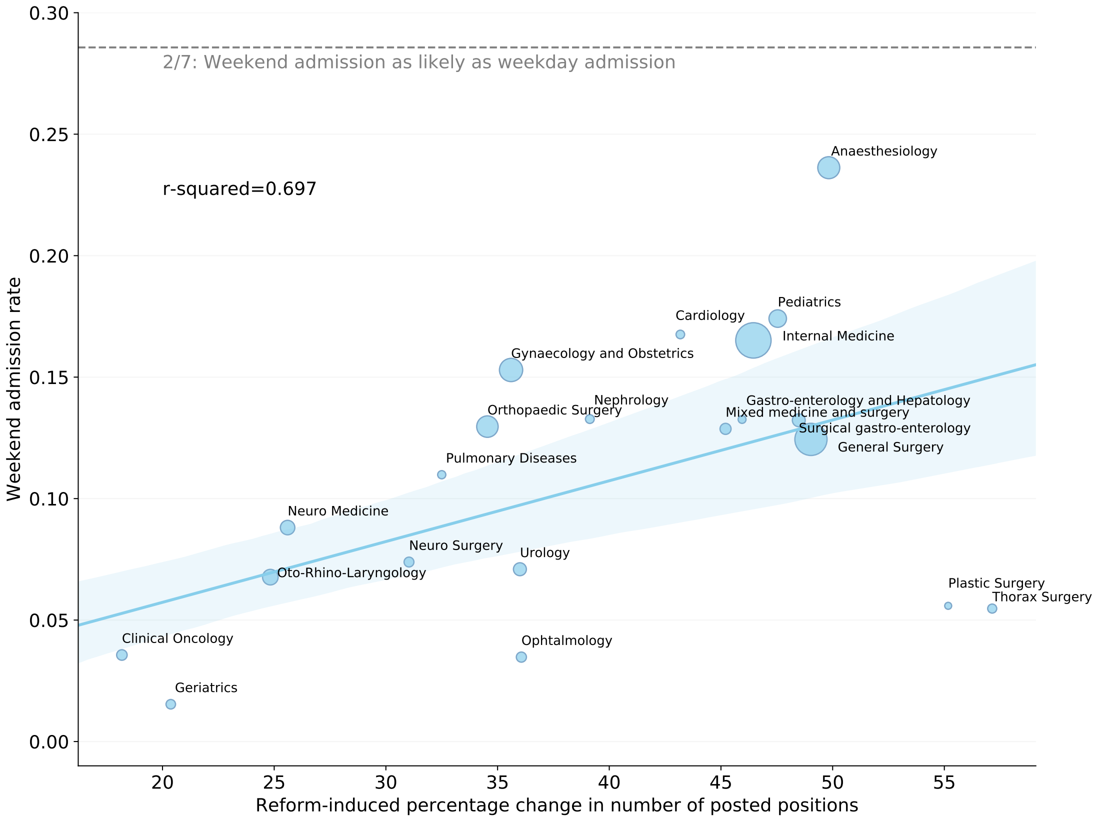
Figure 2.3: Posted Positions by Specialty Category (1972-1991)



Notes: This figure shows the number of posted positions by specialty category. The specialty grouping used can be found in table A.1 in the appendix. The vertical red line indicates the pre-reform year 1981.

Source: *Own calculations based on data from Sundhedsstyrelsen (1972) - Sundhedsstyrelsen (1991)*

Figure 2.4: Correlation with Weekend Admission Rate



Notes: This figure shows the correlation between the percentage increase in the number of posted positions in a specialty between 1981 and 1982 and the specialty's weekend admission rate. The weekend admission rate measures the fraction of a specialty's patients that are admitted on a Saturday or a Sunday. The shaded area shows a 95 percent confidence interval.

Source: *Own calculations based on data from Sundhedsstyrelsen (1972) - Sundhedsstyrelsen (1991) and Statistics Denmark*

2.4 Physicians' career path

In the empirical analysis I am going to estimate the effect of the introduction of the 40-hour workweek on specialty choice across gender. Hence, it is necessary to understand *when* the specialty choice is made and thereby *whose* specialty choice was potentially affected.

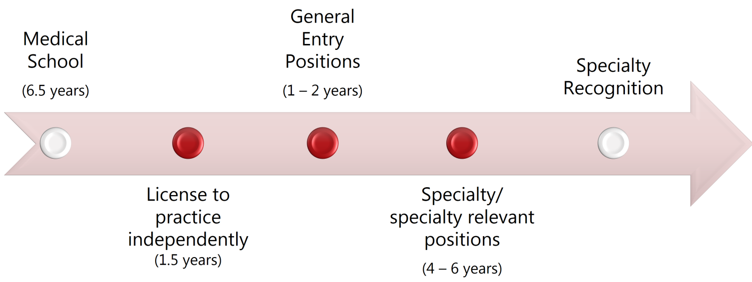
Figure 2.5 shows an outline of a physician's career path from medical school to specialization. The first step is medical school which, covering my sample period and up until 2001, took 6.5 years (Bendixen et al., 2015). Since 1976 (and up until 2008 (Kjær, Qvesel, and Kodal, 2010)) medical school was followed by 18 months of clinical education for those wanting to advance beyond resident physician or temporary deputy for a family physician (Foreningen af yngre læger, 1978). After these 18 months of clinical training, the physician obtained a "license to practice independently" and could now practice without being under supervision. At this point the physician could choose between going into family practice (which required another 18 months of training) or pursuing a specialization[4].

The first part of specialist medical training consisted of 1-2 years of general education (Danish: *almenuddannelse*). At the time of the reform, close to two thirds of the specialties had 2 years of general education with common content. After the general education came the main education (Danish: *hoveduddannelse*) and in some specialties

[4]Family practice was not a specialty in Denmark until 1994.

the side education (Danish: *sideuddannelse*) which amounted to 4-6 years of employment within the specialty or at specialty relevant hospital wards. The minimum total duration of specialist medical training was 6, 6.5 or 7 years, depending on the specialty (Den almindelige danske Lægeforening, 1977).

Figure 2.5: Timeline: Postgraduate training



3 Conceptual Framework

In this section, I describe Goldin (2014)'s theory of occupational pay differences and provide an interpretation of the model in the context of the physician labor market.

3.1 Goldin (2014)'s Theory of Occupational Pay Differences

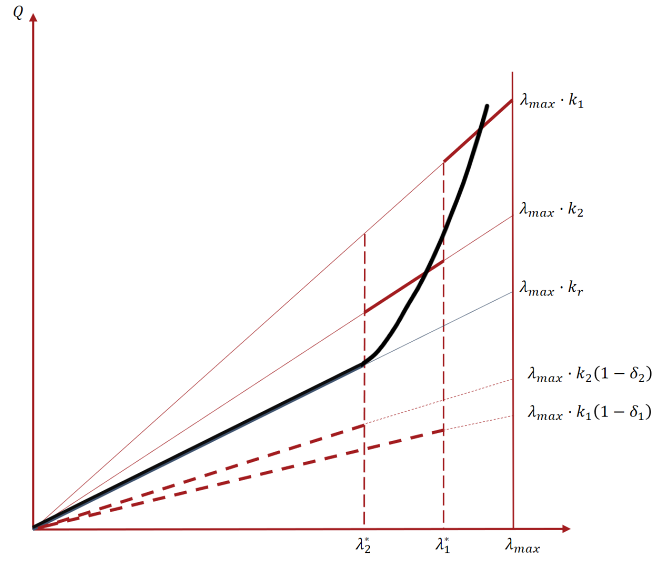
According to Goldin (2014) nonlinearities (convexities) in pay schedules arise when there is a discrete drop in output if the number of hours worked by individual i λ_i (or some measure of the flexibility of those hours) falls below a certain threshold in occupation j. Mathematically[5]:

$$Q = \begin{cases} \lambda_i k_j & \text{if } \lambda_i > \lambda_j^* \\ \lambda_i k_j \cdot (1 - \delta_j) & \text{if } \lambda_i \leq \lambda_j^* \end{cases} \tag{3.1}$$

where k_j is output per hour λ_i in occupation j when λ_i is above the threshold λ_j^* and where δ_j is the drop in output per hour if not. For three occupations where the penalty for hours reductions differ such that $\delta_1 > \delta_2 > \delta_r = 0$ and where the hourly output for hours worked beyond the threshold value also differs such that $k_1 > k_2 > k_r$ there is then an optimal occupational choice given individual i's working hour preferences. For example, an individual with preferences above λ_1^* will work in the most remunerative occupation (occupation 1) while an individual with preferences for working hours below λ_2^* will work in what Goldin calls "the reservation occupation" (occupation r) in which there is no penalty for hours reductions (δ_r=0). Goldin's illustration of the model is replicated in fig. 3.1. Thus, Goldin's model yields a prediction about how a gender wage gap arises in equilibrium: if females' preferences for working hours are on average below those of males (e.g. because of differences in preferences for consumption versus leisure due to child-rearing), then females will self-select into less remunerative occupations at a higher rate than males. Similarly, males will self-select into occupations with convex returns to long hours at a higher rate than females. As a consequence a gender wage gap emerges in equilibrium due to compensating differentials.

[5]The notation used here is identical to that used by Goldin (2014)

Figure 3.1: Replication of figure 4 from Goldin (2014)



3.2 Interpretation of Goldin (2014)'s Model in the Physician Labor Market

In the physician labor market, I am going to interpret Goldin's model as follows: specialties (occupations) differ in their relationship between hours worked and those hours' implications for patient health (output Q). In specialties were the conditions of patients are not acute, patient treatment can safely be postponed till the next day or till after the weekend. In these occupations there is a linear relationship between hours worked and patient health: there is no discrete drop in patient health if the patient is not treated immediately. In acute specialties, however, the relationship between hours worked and patient health is different. If an acute patient is not cared for within a short amount of time it could have fatal consequences. In fact, in some cases it could even have deadly consequences (in which case one might imagine that Q drops to zero). In Goldin's model λ is both the number of hours and a measure of which *particular* hours. Both dimensions are important for patient health for the following reasons:

- **The number of hours** are important for patient health due to continuity of patient treatment. Particularly, if a surgeon is in the middle of performing surgery on a patient when his daytime-duty or shift ends, then it is important for patient health that the physician continues working (i.e. works overtime).
- **The particular hours** are important for patient health because acute patients may come in any time of day/night as well as on weekends.

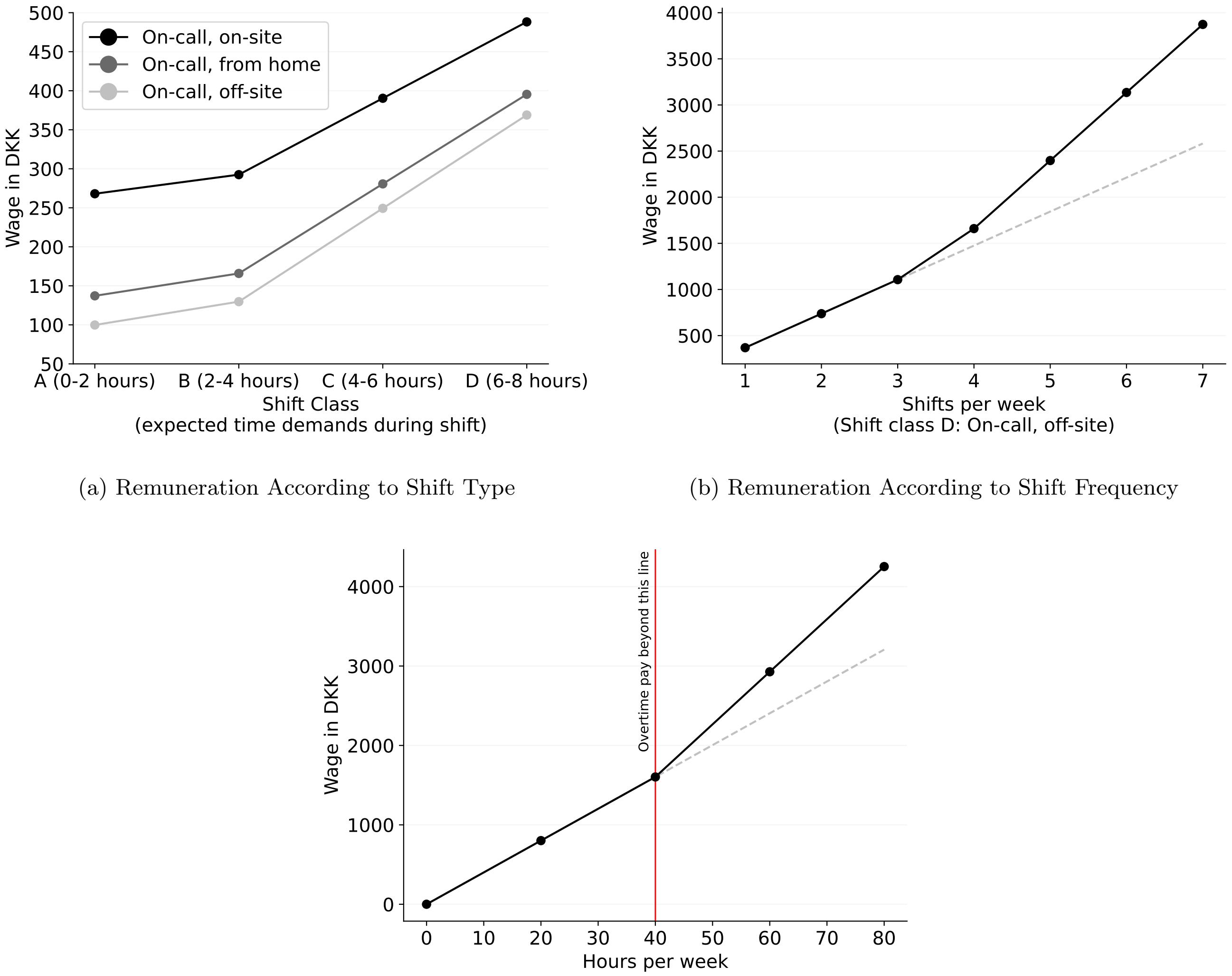
Given my interpretation above, the prediction of Goldin's model is that wage schedules should be convex in acute specialties, with the most acute specialties having the highest convex returns to long/particular hours (due to compensating differentials).

3.3 Necessary Condition: Existence of Convex Wage Schedules

In this section, I examine whether convexities were present in wage schedules pre-reform – and hence whether occupational choices could potentially have been influenced by selection on convex returns as Goldin's theory predicts. Because wages were determined through the collective agreement, there is no need to estimate a wage-setting curve. The pre-reform (1977) collective agreement has provisions with wages for full-time employment, wages for different types of shifts and their frequency and wage rates for overtime. This is convenient because it allows me to determine whether wage schedules were in fact convex, based on simple calculations using the collective agreement's provisions. Figure 3.2 shows examples of relationships between hours worked and wages from the 1977 collective agreement. Figure 3.2a shows the relationship between different shift classes (”A”,”B”,”C” and ”D”) and the wage for such a shift. A shift class is characterized by its expected time demands during a shift. Hence, in shift class ”A” the physician works 0-2 hours in expectation, while a physician in shift class ”B” works 2-4 hours in expectation. Convexities arise in the wage schedule as wages increase more when moving from shift class ”B” to ”C” than they did when moving from shift class ”A” to ”B” (even though both constitute an increase in expected time demands of two hours). Figure 3.2b shows the relationship between the number of shifts per week and total wages. Again, convexities arise because the fourth shift pays 50 percent more than the previous three, while the fifth to seventh shift pays 100 percent more than the first three did. Finally, fig. 3.2c shows the relationship between hours worked per week and total wages. In this case, convexities arise because hours in excess of full-time (40-hours) are remunerated with overtime compensation.

How does this connect to Goldin's theory of occupational pay differences? The wages set out in the collective agreement are the same across all specialties for all subordinate physicians with the same seniority level. However, specialties differ in the extent to which they benefit from these convex returns. For example, a hospital ward's duty team is given a shift class based on its historical effective working hours during shifts.[6] Hence, specialties that are very acute and duty-heavy will tend to 1) be associated with higher expected time demands during a shift (and hence higher shift classes), 2) have more available shifts that physicians can take and 3) have more overtime hours. This is in correspondence with Goldin's theory (given my interpretation of the model provided in section 3.2).

[6]cf. paragraph 3(a) of section 9 in the 1977 collective agreement.



(a) Remuneration According to Shift Type

(b) Remuneration According to Shift Frequency

(c) Remuneration for Overtime Hours

Figure 3.2: Convex Wage Schedules in the 1977 (pre-reform) Collective Agreement

Notes: All wages are for the lowest seniority level and are in 1977 prices (as stated in the Collective Agreement). Dashed lines show what the wage schedule would have looked like, had it been linear. In Panel (c) the wage for a 40-hour workweek is computed as annual full-time earnings divided by 48 (the number of yearly workweeks cf. paragraph 2 of section 15 in the 1977 collective agreement) and for hours below 40, wages are computed from the implied hourly wage.

Source: *Own calculations based on numbers from the 1977 Collective Agreement* (Foreningen af yngre læger, 1977)

3.4 Predicted Effects of Introduction of a Maximum 40-hour Workweek

Now that I have shown that wage schedules were in fact convex, I next turn to what Goldin's model predicts should happen when a cap on working hours is introduced. The model's predictions, however, depend on the underlying distributions of preferences for working hours and specialty characteristics. These preferences determine the fraction of individuals whose optimal specialty choice is responsive to changes in working hours. For some, the introduction of a cap on hours will lead to selection *out* of formerly time-intensive specialties once the higher earnings motive is removed, for others the cap will lead to selection *into* such specialties as the time-requirements are now in better correspondence with their working hour preferences.

Individuals who would have chosen a time intensive specialty absent the cap: A cap on working hours at 40 hours will affect the specialty choice of those who would have preferred to work in a more time intensive specialty absent the cap, and where the prospect of working long/particular hours, and thereby earning higher wages, is pivotal to that decision. That is, where the utility derived from a specialty's non-wage characteristics is not so high, that the individual would still choose the specialty once the higher earnings motive is removed. Such individuals will sort into pre-reform time intensive specialties at a *lower* rate when hours are capped at 40.

Individuals who would not have chosen a time intensive specialty absent the cap: A cap on working hours at 40 hours will affect the specialty choice of those who would not choose a time intensive specialty absent the cap, and where the disutility from working long/particular hours is pivotal to that decision. That is, where the utility derived from a specialty's other characteristics is sufficiently high, such that removing the disutility from long/particular hours renders this specialty choice optimal. Such individuals will sort into pre-reform time-intensive specialties at a *higher* rate when hours are capped at 40.

To the extent that females' preferences for working hours are on average below those of males, females will tend to fall into the latter category while men will tend to fall into the former. Exact predictions will, however, depend on the distributions of preferences for other specialty characteristics across gender as well as on the distributions of preferences for working hours across gender. If the fraction of females who respond to the cap by selecting into formerly time-intensive specialties is greater than the fraction who responds by selecting out of such specialties and if the opposite is true for males, then Goldin's model yields the predictions presented in table 3.1

Table 3.1: Predicted Effect of Introduction of a Maximum 40-hour Workweek

	Occupational Choice (Propensity to enter pre-reform time-intensive specialty)
Male	↓
Female	↑
Net effect on pre-reform time-intensive specialty	Female Ratio ↑

4 Data and Descriptives

4.1 Data Sources

In the empirical analysis, I am going to estimate the impact of the introduction of a 40-hour workweek on specialty choice and fertility. In order to do so, I combine data from the Danish Authorization Register with digitized archival data on the effect of the reform across specialties (shown in section 2.3) and additional administrative data from Statistics Denmark.

From the Danish Authorization Register I obtain information on physicians' year of birth, gender, registration date, country where the education was obtained and specialty recognitions. The registration date, which is coinciding with the graduation date for individuals educated in Denmark, is the variable that I use to determine whether an individual's specialty choice is affected by the reform. I cannot observe specialty choices before an individual obtains a specialty recognition in the Authorization Register. Hence, I cannot observe short run specialty choices. Therefore I use an individual's first specialty recognition as the specialty outcome (to avoid the inclusion of remade choices). I combine these data with administrative data on individual earnings, workplace identifiers, industry, sector and labor market attachment (available from 1980 and onwards) as well as parent identifiers. The administrative registers do not have information on hours worked, which is why I infer how binding the introduction of the 40-hour workweek was across specialties based on posted positions. I retrieve this data from 20 annual statistics books authored by the Danish Health Authority (Medicinalstatistiske Meddelelser: Lægestillinger og Sengepladser på Institutioner (1972-1980) and Sygehusstatistik: Lægestillinger og Sengepladser på Institutioner (1981-1991)). I digitize this data using Adobe Pro's built-in optical character recognition (OCR) tool. To correct for OCR mistakes, I cross-check the numbers by recomputing column sums and manually checking entries in case of discrepancies. From this data, I compute the percentage increase in a specialty's posted positions and use it is a measure of how binding the 40-hour constraint was across specialties (the reform intensity in a specialty).

4.2 Summary Statistics

Table 4.1 shows summary statistics measured in the pre-reform implementation year 1981 for individuals who graduate between 1950 (the beginning of my sample period) and 1981. These summary statistics show that 22.4 percent of pre-reform working age graduates were female. These females were more likely to be employed at a hospital than males and less likely to be a specialist, a general practitioner and employed in the private sector. Differences, that might in part be attributable to the fact that females were on average almost one and a half years younger. The table shows that a male earned on average 78,701 DKK more than a female amounting to an (unadjusted) gender wage gap of 29.72 percent.

Table 4.2 shows summary statistics by specialty category for individuals who obtained their specialty recognition before 1983. That is, individuals who chose specialty before introduction of the 40-hour workweek. The table shows clear indications of occupational segregation by gender with 33 percent of psychiatrists being female against only 4 percent of surgeons. Consistent with Goldin's theory, psychiatrists had the lowest average earnings while surgeons had the highest.

Table 4.1: Summary Statistics for Analysis Data: Pre-reform Graduates

	(1) Total mean	(2) Male mean	(3) Female mean	(4) Male-Female
Graduation year	1970.30	1969.89	1971.71	-1.81***
Age	39.05	39.37	37.93	1.44***
Years since graduation	10.70	11.11	9.29	1.81***
Employed in DK (d)	0.84	0.85	0.81	0.03***
Hospital (d)	0.44	0.43	0.47	-0.04***
Specialist (d)	0.24	0.25	0.20	0.05***
Private sector (d)	0.27	0.30	0.19	0.10***
General Practice (d)	0.12	0.14	0.06	0.08***
Labor Income (DKK)	247,005	264,843	186,141	78,701***
Observations	12,301	9,547	2,754	12,301

Notes: The table shows the mean for the full sample of working age (age $<$ 67) physicians with a Danish medical degree from 1950-1981.

All variables are measured in 1981 (before reform implementation).

Source: *Own calculations based on data from Statistics Denmark and the Danish Authorization Register.*

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

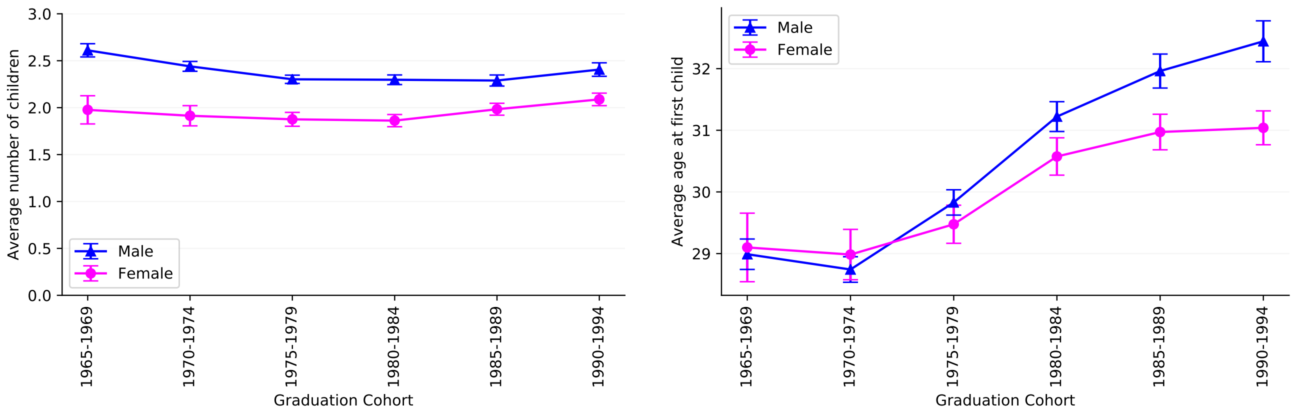
Table 4.2: Summary Statistics by Specialty Category (1981)

	(1) Internal Medicine mean	(2) Surgery mean	(3) Psychiatry mean	(4) Para Clinical mean	(5) Other mean
Graduation year	1963.27	1962.77	1963.66	1962.71	1964.34
Age	45.31	45.90	46.00	46.11	44.61
Years since graduation	17.73	18.23	17.34	18.29	16.66
Female (d)	0.15	0.04	0.33	0.23	0.21
Private sector (d)	0.20	0.19	0.23	0.15	0.37
Hospital (d)	0.63	0.71	0.54	0.70	0.50
Labor Income	327,367	348,041	315,016	338,457	335,576
Observations	578	551	406	471	1,235

Notes: The table shows the mean for a subsample of working age physicians with a medical degree from 1950 or later who obtain their specialty recognition before 1983. All variables are measured in 1981 (before reform implementation).

Source: *Own calculations based on data from Statistics Denmark and the Danish Authorization Register.*

Figure 4.1a shows completed fertility (number of children) by 5-year graduation cohorts. During the entire analysis period, male physicians have more children than female physicians on average, although these differences become smaller towards the end of the period. When looking at age at first child, however, there are no significant differences between males and females in the beginning of my analysis period, but differences emerge from 1980 and onwards as males seem to become fathers at a higher and higher age, while the rise in mothers' age at first child levels off.



(a) Completed Fertility

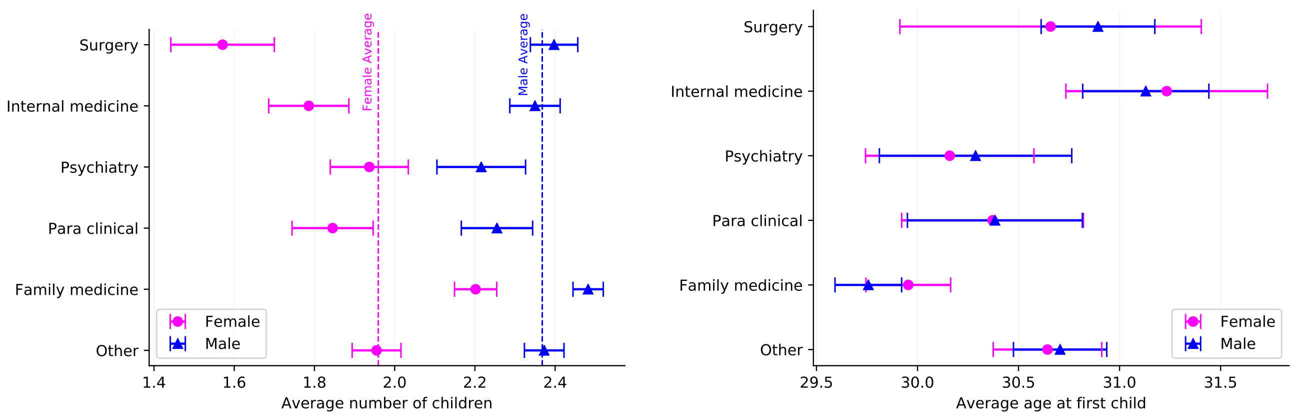
(b) Fertility Timing

Figure 4.1: Completed Fertility and Fertility Timing 1965-1994

Notes: Figure (a) shows the average number of children by graduation cohort and by gender for individuals who are observed in the data until they are at least 45. Errorbars indicate 95 percent confidence intervals.

Source: *Own calculations based on data from Statistics Denmark and the Danish Authorization Register*

Turning to the number of children by specialty category, fig. 4.3a shows significant differences between males and females, with these differences being most pronounced within the surgical specialties. A male surgeon has close to one more child than a female surgeon on average. Besides variation across gender, the figure shows considerable variation in the number of children across specialties for females and much less variation for males. Hence, the pattern observed here is consistent with an explanation where females sort into specialties based on desired fertility as well as with an explanation where the specialty itself constrains female fertility.



(a) Completed Fertility

(b) Fertility Timing

Figure 4.3: Completed Fertility and Fertility Timing 1965-1994

Notes: Figure (a) and (b) show the average number of children and average age at first child by specialty category and by gender for 1965-1994 cohorts. Errorbars indicate 95 percent confidence intervals.

Source: *Own calculations based on data from Statistics Denmark and the Danish Authorization Register*

5 Empirical Framework

In section 2.3 I showed that the introduction of the maximum 40-hour workweek affected some specialties more than others. In this section, I explain how I use this source of exogenous variation in a specialty's working hours to estimate the effect on specialty choice and fertility in a dose-response difference-in-differences design. The research design follows Wasserman (2023b) and exploits that physicians who choose their specialty-path before the reform face different specialty-specific time requirements than physicians who choose their specialty after the reform is implemented.

5.1 Impact on Occupational Sorting: Specialty Choice

In order to estimate the effect of the reduction in hours requirements on specialty choice across gender, I consider both the net effect on a specialty's gender ratio and the gender-specific response (following Wasserman, 2023b). When estimating the effect on a specialty's gender ratio, I estimate the following two-way fixed effects regression model:

$$FemaleFrac_{st} = \delta(ReformIntensity_s \cdot Post_t) + \gamma_s + \alpha_s t + \lambda_t + \epsilon_{st} \tag{5.1}$$

Where $FemaleFrac_{st}$ is the fraction of females in specialty s from graduation cohort t, $ReformIntensity_s$ is the reform-induced percentage change in the number of posted positions in specialty s and $Post_t$ is a dummy for affected cohorts (I return to the definition of this variable in section 5.3 below). The terms γ_s and λ_t are specialty and time fixed effects while $\alpha_s t$ is a specialty-specific linear trend. Hence, the parameter γ_s captures time-invariant but specialty-specific characteristics while λ_t captures specialty-invariant but time-varying effects. The term $\alpha_s t$ captures specialty-specific preexisting trends (under the assumption of linearity). The parameter of interest is δ which measures the effect of the maximum 40-hour workweek on a specialty's (cohort-specific) female ratio.

Since changes in a specialty's female ratio can result from changes in both male and female sorting behavior, I furthermore estimate the effect on the gender-specific response. In order to do this, I estimate the effect on the fraction of females(males) from the female(male) graduation cohort t who enters a specialty s:

$$FemaleCohortFrac_{st} = \delta(ReformIntensity_s \cdot Post_t) + \gamma_s + \alpha_s t + \lambda_t + \epsilon_{st} \tag{5.2}$$

Where the right-hand side variables are defined as in eq. (5.1). I estimate eq. (5.2) for females and males separately (using $MaleCohortFrac_{st}$ as the dependent variable in the latter case).

5.2 Impact on Fertility

When estimating the impact on fertility outcomes, I consider the following (now individual-level) equation[7]:

$$Fertility_{ist} = \delta(ReformIntensity_s \cdot Post_t) + \gamma_s + \alpha_s t + \lambda_t + \beta_1 YOB_{ist} + \beta_2 specialist81_{ist} + \epsilon_{ist} \tag{5.3}$$

[7] This is reminiscent of Wasserman (2023a) although I consider different measures of fertility and a different set of individual-level controls.

Where $Fertility_{ist}$ is the fertility outcome of individual i (who chooses/has chosen specialty s and who is from a graduation cohort t, which might belong to the post-period) while YOB_{ist} is the year of birth of individual i and $specialist81_{ist}$ is a dummy for whether the individual was a specialist physician in 1981. Controlling for year of birth captures generation effects (correlated with whether an individual belongs to a cohort from the post-period) which might affect fertility. The other variables are defined as in eq. (5.2). I estimate eq. (5.3) separately for females and males using multiple measures of fertility including the external margin (a dummy for whether the individual has any children) and the internal margin (the number of children).

5.3 Defining the treatment period

In the absence of data on physicians' short-run specialty choices, I assign individuals to the treatment period based on their graduation cohort. I estimate eq. (5.1) - eq. (5.3) using different definitions of the $Post_t$-dummy. That is, I vary the graduation year for the onset of treatment. The resulting estimates reflect varying degrees of control/treatment period misallocation risk. Consider the following definition of the $Post_t$-dummy when considering the effect on specialty choice:

$$Post_t = \begin{cases} 1 & \text{if graduation year} \geq 1981 \\ 0 & \text{else} \end{cases} \tag{5.4}$$

Using this definition there is no risk of including untreated observations in the treatment period as none of them have obtained their license to practice independently by the time of reform implementation and hence none of them have had to consider which specialty-track to pursue. There is, however, a significant risk that the pre-period is contaminated with treated individuals. For example a graduate from summer 1979 could have obtained the right to practice independently towards the end of 1980 and started in a general entry position in 1981. Since around 2/3 of the specialties had the same requirements for these general entry positions, this individual might have been able to switch specialties after the introduction of the 40-hour workweek without having to start over. Hence, redefining eq. (5.4) such that treatment switches on for earlier graduation cohorts decreases the risk of treated individuals being assigned to the pre-period while increasing the risk of untreated individuals being assigned to the treatment period.

The relevant cut-off for the post-dummy is different when measuring the effect on fertility instead of specialty choice. While the older cohorts may have chosen their specialty-path before introduction of the 40-hour workweek, they may still have the opportunity to adjust their fertility decisions. Other cohorts have not chosen their specialty-path yet when the 40-hour workweek is introduced and hence sort into specialties based on realized or desired fertility. Hence, the relevant cut-off depends on whether one is interested in the effect of the 40-hour workweek on the fertility of individuals in the specialty holding constant specialty choice or not. That is, the effect on fertility for individuals who already sorted into the specialty. If increases in the fertility of the individuals in a specialty happens only through sorting, then one should expect to see effects only for cohorts who graduate close to 1981. If, however, individuals in a particular specialty react by adjusting their fertility decisions, then one should expect to see effects for earlier graduation cohorts.

How should the pre-period then be defined when the outcome is fertility? Because of the female fecundity horizon, the probability that a female physician can respond in terms of fertility decreases with age. For example, Larsen and Yan (2000) find that female fecundability decreases monotonically from its global maximum in the start 20's to zero towards the end of the 40's. Furthermore, multiple studies find that factors such as social interaction and intergenerational transmission of values and behavior shape fertility timing (see e.g. Balbo, Billari, and Mills (2013) for a review). Hence, when estimating the effect of the reform on female fertility one might expect the following:

- females above the age of 49 are not able to increase their fertility (for biological reasons)
- the probability that a female physician is non-responsive increases with her age when her age is in the vicinity of the end of her fertile age (for biological and social/cultural reasons)

These factors will be most pronounced for the older cohorts. For example, graduates from 1970 are on average above the age of 40 at the time of reform implementation. These cohorts will also tend to have a specialty recognition at the time of the reform. Hence, it is reasonable to assume that both channels (the specialty sorting channel and the post-sorting fertility adjustment channel) will not be operating for the vast majority of individuals who belong to cohorts from 1970 and earlier.[8] Since males have a much longer fecundity horizon, it is less clear what one should expect here. However, in general their fertility choices will be restricted by their partner's fecundity horizon.

6 Results

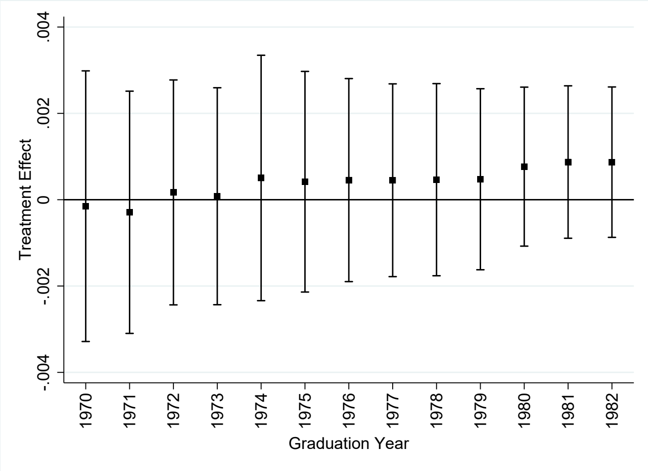
In this section, I present estimated effects of the working hour reform on specialty choice and fertility.

6.1 Effect on specialty choice

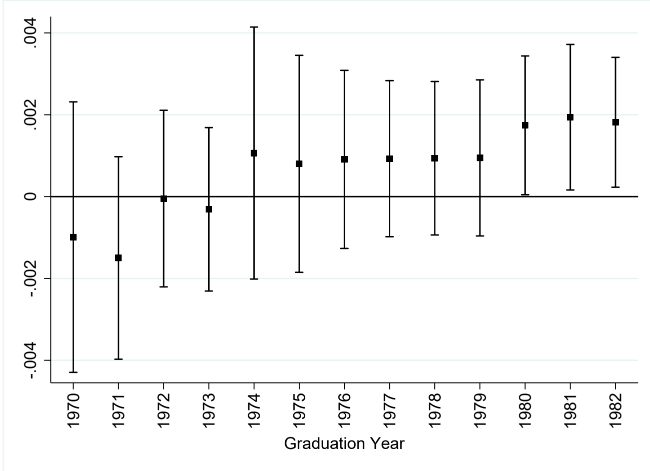
Figure 6.1 shows the estimated effect of the working hour reform on the fraction of females in a specialty using different "placebo-like" cut-offs as explained in section 5.3. That is, each year on the horizontal axis indicates the graduation year used to define onset of the treatment period. The figure shows that estimates from a specification without specialty-specific linear trends are insignificantly different from zero. The inclusion of specialty-specific trends, however, leads to a more than doubling of the parameter estimates for definitions of onset of treatment where the cohort is treated with a high degree of certainty (1980-182). For these definitions of the Post-dummy estimates are significantly different from zero. When onset of treatment is defined as 1981 (where everyone in the treatment period are treated for sure, but the pre-period might be contaminated with some treated individuals), the estimated effect of the working hour reform is 0.0019. A "naive"[9] interpretation of this estimate is then that a 10 percentage point increase in a specialty's reform intensity increases a specialty's fraction of females by 1.9 percentage points (recall that all surgical specialties experience reform intensities > 30 percent).

[8]Figure A.2 in the appendix illustrates a plausible response-pattern across cohorts for females.

[9]I refer to this as a naive interpretation because recent advancements in the Difference-in-Differences literature (see Callaway, Goodman-Bacon, and Sant'Anna (2021)) questions this standard interpretation of TWFE parameter estimates in a dose-response Difference-in-Differences design.



(a) Without Specialty-specific Linear Trends



(b) With Specialty-specific Linear Trends

Figure 6.1: Estimated Effect of Working Hour Reform on Specialties' Fraction of Females

Notes: Figure (a) and (b) show the estimated effect of the working hour reform (estimates of the parameter δ in eq. (5.1)) with and without the inclusion of specialty-specific linear trends using the sample that consists of 1950-1994 graduation cohorts. Errorbars indicate 95 percent confidence intervals.

Source: *Own calculations based on data from the Danish Authorization Register*

Table 6.1: Specialty Choice Results

	(1) Fraction Female	(2) Female Cohort Share	(3) Male Cohort Share
Panel A: Without specialty-specific linear trends			
Reform Intensity x Post	0.0009	0.0008***	0.0002
	(0.0009)	(0.0002)	(0.0003)
R-squared	0.619	0.625	0.705
Panel B: With specialty-specific linear trends			
Reform Intensity x Post	0.0019**	0.0006*	0.0003
	(0.0009)	(0.0003)	(0.0002)
R-squared	0.656	0.665	0.818
Year FE	Yes	Yes	Yes
Specialty FE	Yes	Yes	Yes
Observations	1,234	1,234	1,234

Notes: This table shows estimates of the parameter δ from eq. (5.1) with $Post_t = 1$ if graduation year $\geq$ 1981. The estimation sample consists of 1950-1994 cohorts. Standard errors are clustered at the specialty level (33 clusters) and are in parentheses.

Source: *Own calculations based on data from Statistics Denmark and the Danish Authorization Register.*

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

One potential reason why the inclusion of specialty-specific trends is important[10] could be that the fraction of females were already on the rise in more time-intensive specialties prior to the reform. This is plausible given the observed feminization of the medical profession previously presented in fig. 2.1. Consistent with this explanation, estimation using the fraction of females/males from the female/male cohort as the outcome variable yields estimates that are less sensitive to the inclusion of specialty-specific linear trends. Table 6.1 column (2) and (3) show parameter estimates from such specifications. For the male specification presented in column (3), the estimated effects are insignificantly different from zero with and without the inclusion of specialty-specific linear trends (p-value = 0.475 and 0.156, respectively). The estimated effects for females, however, are significant using both specifications (albeit only at the 10 percent level after the inclusion of specialty-specific trends, p-value = 0.086). After the inclusion of specialty-specific linear trends, the estimated effect of the working hour reform on the share of females that enter a specialty is 0.0006. Meaning that an increase of 10 percentage points in the reform intensity of a specialty increases the share of females that enter the specialty by 0.6 percentage points. Hence for surgical specialties which experienced reform intensities between 31.03 and 57.14 percent, predicted effects are between 1.9 and 3.4 percentage points (on a base of less than 3 percent for surgical specialties overall for the sub-sample presented in table 4.2).

The results presented in table 6.1 are robust to "donut"-specifications where I drop the 3-5 graduation cohorts leading up to 1981. Estimated effects are increasing in the size of the donut, consistent with the presumption that contamination of the pre-period should attenuate the estimated effect of the working hour reform.

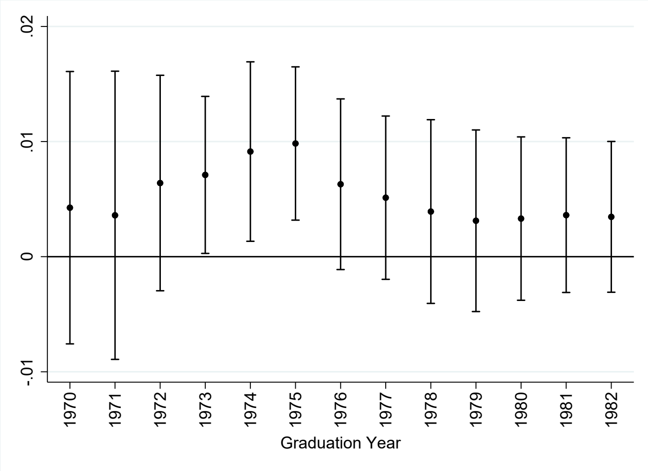
6.2 Effect on fertility

Figure 6.3 shows the estimated effect of the working hour reform on completed fertility (the number of children) for individuals who choose a particular specialty. The figure shows that for females, estimates of the treatment effect are positive and significant for definitions of the post-dummy from 1973-1975, consistent with the prediction that fertility should be most responsive for individuals who are close to their prime childbearing age. As was the case with specialty choice, the inclusion of specialty-specific linear trends affects the estimates.
For the 1975 cut-off, however, the estimate increases slightly but is less precisely estimated. The estimate for this cut-off is 0.012 (p-value = 0.052). For males, the inclusion of specialty-specific linear trends eliminates any significant positive effects. If anything, there is a slight indication of a negative effect of -0.005 at the 1973 cut-off (p-value = 0.077).

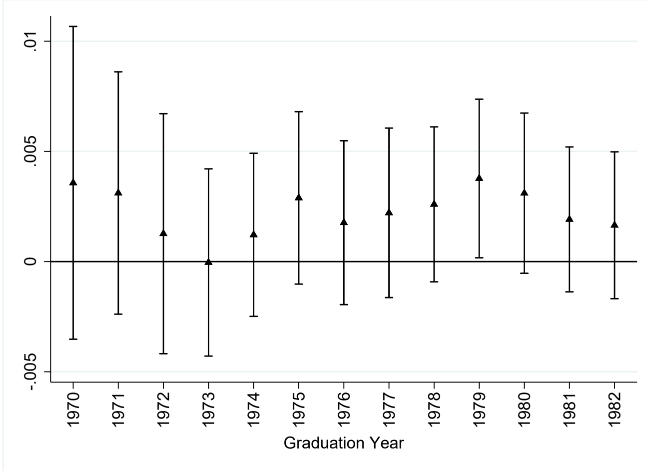
Table 6.2 shows estimated treatment effects for the 1975 cut-off for different outcome variables measuring completed fertility. The table shows that after controlling for specialty-specific trends, estimated effects on the propensity to have any children (the extensive margin) and to have more than one child are unaffected but less precisely estimated. The estimated effect on the propensity to have more than two children, however, increases from 0.003 to 0.007 and becomes significant at the 1 percent level. Hence, an increase in a specialty's treatment intensity of 10 percentage points leads to an increase in the probability of specialists within the specialty having more than two children of 7 percentage points. For males all estimates are insignificant (even though more precisely estimated than for females). Panel B of Table 6.2 shows that there is little indication that effects are driven by changes in fertility timing. Although the signs of the estimated effects on age at first/last child are consistent with an overall increase in the number of childbearing years.

[10]F-tests for joint significance of the specialty-specific linear trends yield p-values = 0.00 in all specifications.

Top: Without specialty-specific linear trends

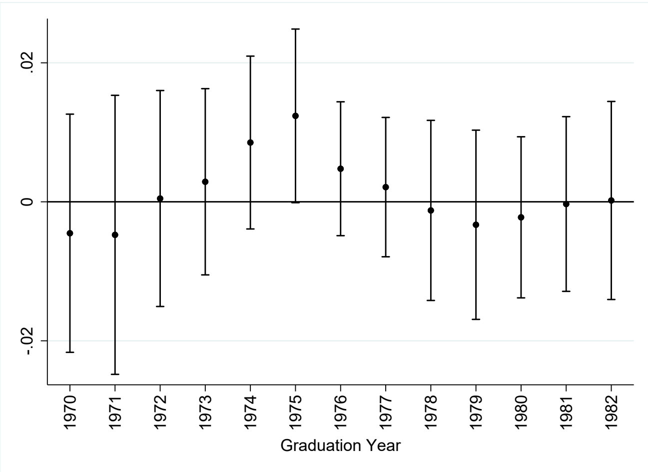


(a) Females

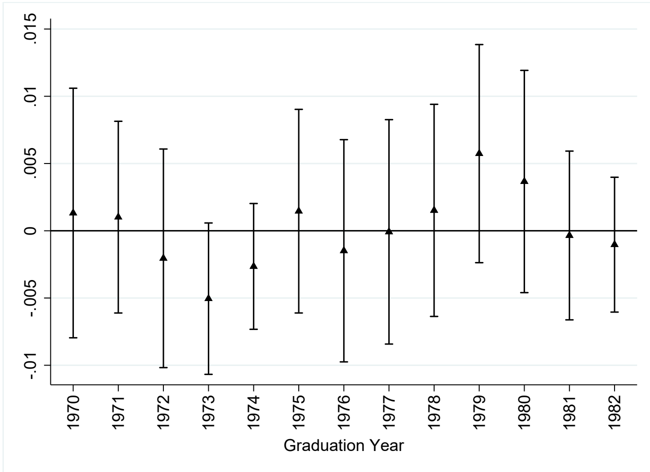


(b) Males

Bottom: With specialty-specific linear trends



(c) Females



(d) Males

Figure 6.3: Estimated Effect of Working Hour Reform on Number of Children

Notes: Figure (a) and (b) show the estimated effect of the working hour reform (estimates of the parameter δ in eq. (5.3)) for females (a) and males (b) using the sample that consists of 1965-1994 graduation cohorts. Errorbars indicate 95 percent confidence intervals. Source: *Own calculations based on data from Statistics Denmark and the Danish Authorization Register*

Table 6.2: Fertility Results

	Independent variable: Reform Intensity x Post			
	Females		Males	
Dependent variable	(1)	(2)	(3)	(4)
	Panel A: Completed Fertility			
Number of children	0.010***	0.012*	0.003	0.001
	(0.003)	(0.006)	(0.002)	(0.004)
Any children	0.003***	0.003	-0.001	-0.001
	(0.001)	(0.002)	(0.001)	(0.001)
More than one child	0.003***	0.003	0.000	0.001
	(0.001)	(0.002)	(0.001)	(0.002)
More than two children	0.003**	0.007***	0.001	-0.000
	(0.001)	(0.002)	(0.001)	(0.001)
Observations	2,756	2,756	5,809	5,809
	Panel B: Fertility Timing			
Age at First Child	-0.005	-0.023	-0.018	-0.017
	(0.014)	(0.026)	(0.010)	(0.016)
Age at Last Child	0.010	0.010	-0.002	0.013
	(0.013)	(0.026)	(0.014)	(0.029)
Observations	2,301	2,301	5,364	5,364
Specialty-specific trends	No	Yes	No	Yes
Year FE	Yes	Yes	Yes	Yes
Specialty FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: This table shows estimates of the parameter δ from eq. (5.3) with $Post_t = 1$ if graduation year $\geq$ 1975. The estimation sample consists of 1965-1994 cohorts who are observed in the data until they are at least 45 years old. Controls include year of birth and a dummy for whether the individual obtains his/her specialty recognition prior to 1982. Standard errors are clustered at the specialty level (33 clusters) and are in parentheses.

Source: *Own calculations based on data from Statistics Denmark and the Danish Authorization Register.*

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7 Conclusion

In this paper I examine the effect of the introduction of a maximum 40-hour workweek on specialty choice and fertility across gender. Following Wasserman (2023b), I exploit that the cap on hours binds differentially across specialties and estimate the effect of the policy in a dose-response difference-in-differences design. Absent available data on specialties' pre-policy hours, I exploit that the reform acts as a positive labor demand shock. I find that the introduction of a 40-hour workweek lead to increases in the share of females in formerly time-intensive specialties. My findings suggest that these increases are driven by increases in female selection into specialties that experienced reductions in hours coupled with non-response among males. I show that pre-policy sorting was subject to an environment with convex returns to long and particular hours, implying that sorting pre-policy could have resulted from selection on convex returns as Goldin (2014)'s theory of occupational pay differences predicts. The estimated effects on specialty sorting across gender are consistent with Goldin's theory of occupational pay differences: females sort into formerly time-intensive specialties at a higher rate indicative of a relatively high prevalence of preferences for lower working hours. Non-response among males, however, lends itself to multiple explanations. One potential explanation is that increased selection of formerly time-intensive specialties among the fraction of the male population who prefers lower working hours, is exactly offset by deselection of such specialties among those who would have chosen the specialty because of its convex returns. Another explanation consistent with the observed non-response is that males' preferences for specialty characteristics unrelated to wages are so strong relative to the earnings motive and the potential disutility from long working hours, that introduction of a 40-hour workweek does not affect which specialty they prefer. I find that female, but not male, fertility increased more where introduction of the 40-hour workweek had more bite. Due to the classic simultaneity problem of fertility and labor supply, this is both consistent with an explanation where females sort into specialties based on desired fertility as well as with an explanation where a specialty's time requirements causally affect realized fertility. However, effects on fertility arise for older graduation cohorts than for specialty choice, suggesting that part of the effect could be working through the latter channel.

My findings corroborate recent work by Wasserman (2023b) who finds that a cap at 80 hours affects the specialty choice of females but not males in a US context. I show that these results carry over in a setting where hours are capped at 40 hours. The results presented in this paper could however benefit from further analysis in order to strengthen the causal interpretation of the estimates. The possibility that pre-periods are contaminated with potentially treated individuals should, however, only lead to attenuation of parameter estimates. For estimation of fertility outcomes where the definition of the pre-period was less clear, supplementary analyses which take a closer look at heterogeneous effects by age could shed further light on whether estimated effects can be attributed to the working hour reform.

References

Andersen Nexø, Sniff et al. (2004). *Mellem kald og kamp - Yngre Læger i 100 år*. 1st ed. Lægeforeningens Forlag.

Angrist, Joshua D. and William N. Evans (1998). "Children and Their Parents' Labor Supply : Evidence from Exogenous Variation in Family Size". In: *The American Economic Review* 88.3, pp. 450–477.

Balbo, Nicoletta, Francesco C Billari, and Melinda Mills (2013). "Fertility in Advanced Societies : A Review of Research". In: *European Journal of Population* 29.1, pp. 1–38.

Bendixen, Gunnar et al. "Den store danske". URL: https://denstoredanske.lex.dk/l%C3%A6geuddannelse. *Date Accessed: 2023-10-06*.

Blau, Francine D. and Lawrence M. Kahn (2017). "The Gender Wage Gap: Extent, Trends, and Explanations". In: *Journal of Economic Literature* 55.3, pp. 789–865.

Bolotnyy, Valentin and Natalia Emanuel (2022). "Why Do Women Earn Less than Men? Evidence from Bus and Train Operators". In: *Journal of Labor Economics* 40.2, pp. 283–323.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna (2021). "Difference-in-Differences with a Continuous Treatment". In: URL: https://arxiv.org/abs/2107.02637.

Christy, Morten (1981a). "Den går ikke!!!" In: *Ugeskrift for Læger* 143/13, pp. 859–860.

— (1981b). "LOCK-OUT". In: *Ugeskrift for Læger* 143/18.

— (1981c). "Referat fra F.a.y.l.'s represæntantskabsmøde 13.-14. november 1980". In: *Ugeskrift for Læger* 143/3, pp. 181–184.

Den almindelige danske Lægeforening (1977). *Lægeforeningens Vejviser 1977*. København: Lægeforeningens Forlag.

— (1981). *Lægeforeningens Vejviser 1981*. København: Lægeforeningens Forlag.

Doyle, Joseph J. et al. (2015). "Measuring returns to hospital care: Evidence from ambulance referral patterns". In: *Journal of Political Economy* 123.1, pp. 170–214.

Fadlon, Itzik, Torben Heien Nielsen, and Frederik Plesner Lyngse (2022). "Causal Effects of Early Career Sorting on Labor and Marriage Market Choices: A Foundation for Gender Disparities and Norms". In: *CEBI Working Paper Series* 06/22.

Foreningen af yngre læger. "Takster og overenskomster 1977: Overenskomst om Løn- og ansættelsesvilkår for underordnede sygehuslæger mellem Amtsrådsforeningen i Danmark Københavns Hospitalsvæsen Finansministeriet Frederiksberg kommune og Foreningen af yngre læger: Gældende fra 1. apr".

Foreningen af yngre læger (1978). *Basishåndbog*. København: Foreningen af yngre læger.

Foreningen af yngre læger. "Takster og overenskomster 1982: Overenskomst om løn- og øvrige ansættelsesvilkår for underordnede sygehuslæger mellem Amtsrådsforeningen i Danmark Københavns kommune Finansministeriet Frederiksberg kommune og Foreningen af yngre læger".

Goldin, Claudia (2014). "A Grand Gender Convergence : Its Last Chapter". In: *American Economic Review* 104.4, pp. 1091–1119.

Hansen, Else (2015). "Universiteter og idealer i velfærdsstaten". In: *Uddannelseshistorie (year)* 2015, pp. 76–91.

— (2018). "Planlægningens (oversete) årti Argumenter for adgangsbegrænsning i 1970'erne". In: *Historisk Tidsskrift* 118.1, pp. 77–112.

Herr, Jane Leber and Catherine D. Wolfram (2012). "Work Environment and Opt-out Rates at Motherhood across High-Education Career Paths". In: *ILRReview* 65.4, pp. 928–950.

Kjær, Niels Kristian, Dorte Qvesel, and Troels Kodal (2010). "An evaluation of the 18- and 12-month basic postgraduate training programmes in Denmark". In: *Danish Medical Bulletin* 57/8.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard (2019). "Children and gender inequality: Evidence from Denmark". In: *American Economic Journal: Applied Economics* 11.4, pp. 181–209.

Kreiner, Claus Thustrup and Michael Svarer (2022). "Danish Flexicurity: Rights and Duties". In: *Journal of Economic Perspectives* 36.4, pp. 81–102.

Lægeforeningen. "Vedtægter for Lægeforeningen". URL: `https://www.laeger.dk/media/ie2kiwru/laegeforeningens-vedtaegter-2022.pdf`. *Date Accessed: 2023-07-24*.

Larsen, Ulla and Sharon Yan (2000). "The age pattern of fecundability: An analysis of French Canadian and Hutterite birth histories". In: *Social Biology* 47.1-2, pp. 34–50.

Rosenzweig, Mark R and Kenneth Wolpin (1980). "Life-Cycle Labor Supply and Fertility : Causal Inferences from Household Models". In: *Journal of Political Economy* 88.2, pp. 328–348.

Steensen, Jens Peter (1981). "Formandens mundtlige beretning". In: *Ugeskrift for Læger* 143/3, pp. 171–177.

Sundhedsdatastyrelsen. "Dokumentation af Registre". URL: `https://www.esundhed.dk/Dokumentation/DocumentationExtended?id=11`. *Date Accessed: 2023-03-30*.

Sundhedsstyrelsen (1972). *Medicinalstatistiske Meddelelser 1972:2: Normerede og besatte lægestillinger samt sengepladser på institutioner*. København: Sundhedsstyrelsen.

— (1973). *Medicinalstatistiske Meddelelser 1973:4: Normerede og besatte lægestillinger samt sengepladser på institutioner, maj 1973*. København: Sundhedsstyrelsen.

— (1974). *Medicinalstatistiske Meddelelser 1974:4: Lægestillinger og sengepladser på institutioner, maj 1974*. København: Sundhedsstyrelsen.

— (1975). *Medicinalstatistiske Meddelelser 1975:3: Lægestillinger og sengepladser på institutioner, maj 1975*. København: Sundhedsstyrelsen.

— (1976). *Medicinalstatistiske Meddelelser 1976:3: Lægestillinger og sengepladser på institutioner, maj 1976*. København: Sundhedsstyrelsen.

— (1977). *Medicinalstatistiske Meddelelser 1977:3: Lægestillinger og sengepladser på institutioner, maj 1977*. København: Sundhedsstyrelsen.

— (1978). *Medicinalstatistiske Meddelelser 1978:5: Lægestillinger og sengepladser på institutioner, maj 1978*. København: Sundhedsstyrelsen.

— (1979). *Medicinalstatistiske Meddelelser 1979:6: Lægestillinger og sengepladser på institutioner, maj 1979*. København: Sundhedsstyrelsen, p. 10.

— (1980). *Medicinalstatistiske Meddelelser 1980:2: Lægestillinger og Sengepladser på Institutioner Maj 1980*. København: Stougaard Jensen, p. 10.

— (1981). *Sygehusstatistik II:4:1981: Lægestillinger og sengepladser på institutioner Maj 1981*. København: Stougaard Jensen, p. 10. ISBN: 87-503-3945-1.

— (1982a). *Sygehusstatistik II:11:1982: Lægestillinger og sengepladser på institutioner November 1982*. København: Stougaard Jensen, p. 10. ISBN: 87-503-4437-4.

— (1982b). *Sygehusstatistik II:7:1982: Lægestillinger og sengepladser på institutioner Maj 1982*. Stougaard Jensen/København, p. 10. ISBN: 87-503-4171-5.

Sundhedsstyrelsen (1983). *Sygehusstatistik II:14:1983: Lægestillinger og sengepladser på institutioner april 1983.* København: Sundhedsstyrelsen, pp. 12–13.

— (1984). *Sygehusstatistik II:20:1984: Lægestillinger og sengepladser på institutioner, april 1984.* København: Sundhedsstyrelsen, p. 11.

— (1985). *Sygehusstatistik II:25:1985: Lægestillinger og sengepladser på institutioner, april 1985.* København: Sundhedsstyrelsen, pp. 10–11.

— (1986). *Sygehusstatistik II:30:1986: Lægestillinger og sengepladser på institutioner, april 1986.* København: Sundhedsstyrelsen.

— (1987). *Sygehusstatistik II:36:1987: Lægestillinger og sengepladser på institutioner, april 1987.* København: Sundhedsstyrelsen.

— (1988). *Sygehusstatistik II:40:1988: Lægestillinger og sengepladser på institutioner, april 1988.* København: Sundhedsstyrelsen.

— (1989). *Sygehusstatistik II:44:1989: Lægestillinger og sengepladser på institutioner, april 1989.* København: Sundhedsstyrelsen, pp. 14–15.

— (1990). *Sygehusstatistik II:47:1990: Lægestillinger og sengepladser på institutioner, april 1990.* København: Sundhedsstyrelsen, pp. 12–13.

— (1991). *Sygehusstatistik II:52:1991: Lægestillinger og sengepladser på institutioner, april 1991.* København: Sundhedsstyrelsen, pp. 18–19.

Sundhedsstyrelsen. "Data i autorisationsregisteret".

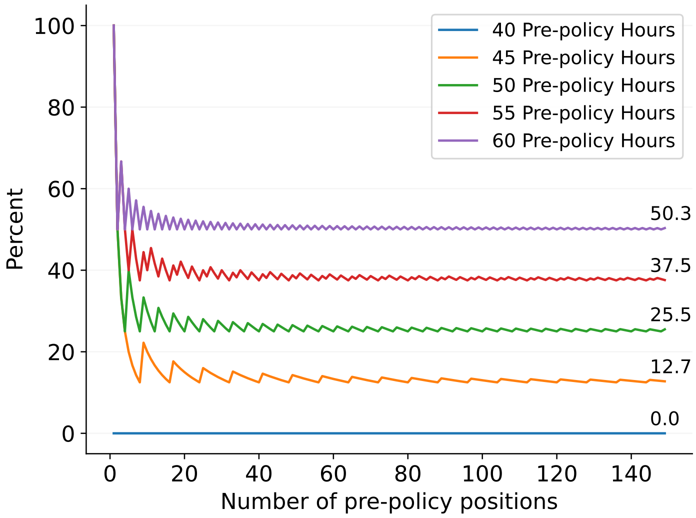
Wasserman, Melanie (2023a). "Do Long Work Hours Constrain Women's Fertility Choices ?" In: *Working paper* February 2023.

— (2023b). "Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents". In: *Review of Economic Studies* 90, pp. 1535–1568.

Winther Jensen, Jens. "Lægeforeningen". URL: https://ugeskriftet.dk/nyhed/laegeforeningen. *Date Accessed: 2023-07-24.*

A Supplementary Figures and Table

Figure A.1: Accuracy of Reform Intensity Variable



Notes: This figure shows the accuracy of the reform intensity variable given different hypothetical values of weekly pre-policy hours. It shows that the accuracy increases in the number of pre-policy positions.
Source: *Own calculations*

Figure A.2: Example of plausible response pattern across cohorts

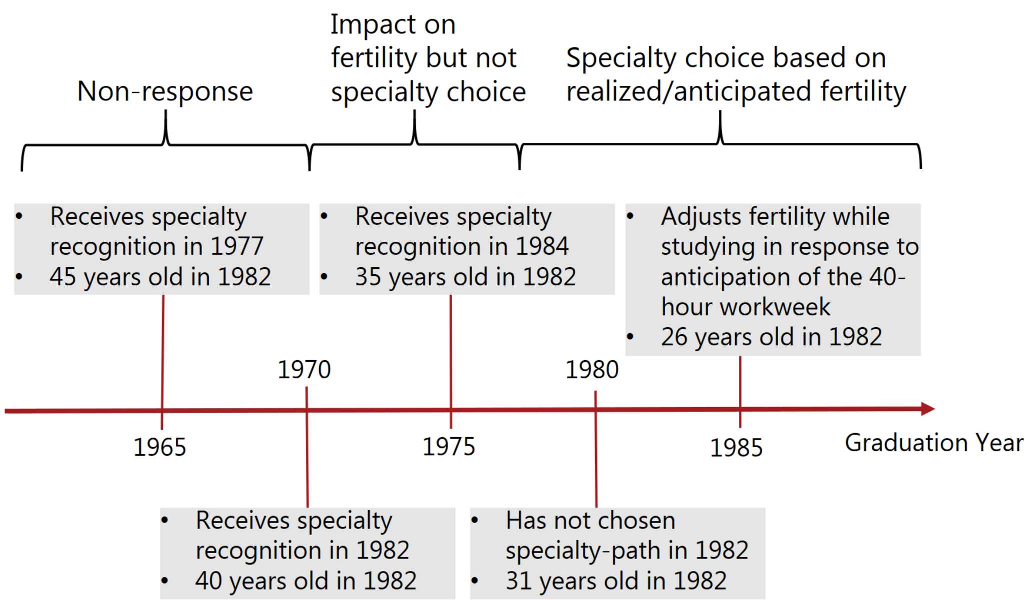


Table A.1: Impact of working hour reform across specialties

Original Specialty Name	Current Specialty Name	Percentage increase in number of posted positions
Surgery		
General Surgery	General Surgery	
Neuro Surgery	Neuro Surgery	
Orthopaedic Surgery	Orthopaedic Surgery	
Plastic Surgery	Plastic Surgery	
Thorax Surgery	Thorax Surgery	
Urology	Urology	
Surgical Gastroenterology	General Surgery	
Surgery average/weighted average		**44.49 / 44.30**
Internal Medicine		
Cardiology	Internal Medicine: Cardiology	
Physical Medicine and Rehabilitation	Internal Medicine: Rheumatology	
Internal Medicine	Internal Medicine	
Pulmonary Medicine	Internal Medicine: Pulmonary Diseases	
Nephrology	Internal Medicine: Nephrology	
Tropical Medicine	Internal Medicine: Infectious Diseases	
Medical Gastroenterology	Internal Medicine: Gastro-enterology and Hepatology	
Long-term Medicine	Internal Medicine: Geriatrics	
Internal Medicine average/weighted average		**28.85 / 39.29**
Psychiatry		
Child Psychiatry	Child and Youth Psychiatry	
Psychiatry	Psychiatry	
Psychiatry average/weighted average		**16.10 / 19.70**
Para Clinical		
Clinical Blood Type Serology	Clinical Immunology	
Clinical Physiology	Clinical Physiology and Nuclear Medicine	
Clinical Chemistry	Clinical Biochemistry	
Clinical Microbiology	Clinical Microbiology	
Diagnostic Radiology	Radiology	
Pathological Anatomy and Histology	Pathological Anatomy and Cytology	
Para Clinical average/weighted average		**9.14 / 8.40**
Other		
Anaesthesiology	Anaesthesiology	
Clinical Neurophysiology	Neuro Medicine	
Dermato-Venerology	Dermato-Venerology	
Gynaecology and Obstetrics	Gynaecology and Obstetrics	
Neuro Medicine	Neuro Medicine	
Ophtalmology	Ophtalmology	
Oto-rhino-laryngology	Oto-rhino-laryngology	
Pediatrics	Pediatrics	
Therapeutic Radiology	Clinical Oncology	
Other average/weighted average		**26.85 / 35.94**

Notes: The table shows the 32 specialties that existed in 1981/1982 along with their original and current specialty names. Weighted averages refer to the specialty-group specific average when the specialties within the group are weighted by the number of pre-policy (1981) posted positions.

Source: *Own calculations based on data from Sundhedsstyrelsen (1972) - Sundhedsstyrelsen (1991)*

B Description of Variables

Specialty Outcome

The measure of specialty choice that I use, is a physician's specialty recognition registered in the Authorization Register. I cannot observe their short-run specialty choices. Hence, to insure that the graduation year is informative about the set of information that the physician has when making his specialty choice, I restrict to the first registered specialty. I make sure to always use the branch specialty in cases where the physician receives a specialty recognition in a ground specialty (e.g. internal medicine) and a branch specialty (e.g. cardiology) on the same date. As the specialty classifications change over time, I use a specialty's current specialty name. The current specialty name is a translation variable present in the Authorization Register which is used for translation of original specialty names into current ones (Sundhedsdatastyrelsen, 2015). The choice of whether to use the original specialty name or the current one turns out to be important due to the inclusion of specialty fixed effects in the empirical analysis. Finally, I exclude family medicine as it was not a specialty until 1994.

Treatment Intensity

As mentioned in section 2.3, I compute the percentage change in posted positions within a specialty from 1981 to 1982 (the implementation year) in order to get a measure of the treatment intensity across specialties. I link this variable to the register data by first linking specialties' original specialty name with their current specialty name and thereafter merging the treatment intensity variable on the remaining data using the current specialty name.

C Sample Restrictions

In this section I list the sample restrictions that I use in order to arrive at my analysis dataset:

i) No individuals without a valid social security number at the time of authorization: Individuals without a valid social security number at the time of authorization appear in the authorization register with a unique ID. If the physician obtains a social security number after authorization, it is added only in the case where the physician has informed the Danish Health Authority about it (Sundhedsdatastyrelsen, 2015).

ii) No individuals with graduation date prior to 1950: The electronic authorization register was started in 1982. The first registration included all working age authorized physicians (Sundhedsstyrelsen, 2014). Therefore, I restrict the sample to individuals registered from 1950 and onwards. This restriction imposes the assumption that all physicians who graduate in 1950 or later are no older than working age in 1982. Absent any definition of "working age" I assume that this implies no older than the statutory retirement age which was 67 in 1982.

iii) No individuals with graduation date from 1995 and onwards. Since the 1981 collective agreement was replaced by an entirely new collective agreement in 1995, I exclude these later graduation cohorts to avoid that the treatment group contains individuals who were never affected by the 1981 collective agreement.

iv) No individuals who do not have Denmark as their country of authorization. In some instances, the year of registration in the Authorization Register does not correspond to the medical school graduation year. This happens primarily in the case of immigrants who obtain their Danish authorization upon arrival in Denmark in which case the year of registration, license to practice independently and specialization may be the same. Since I assign individuals to either the treatment or control group based on their graduation cohort, I need the registration year to be informative about the year of graduation.

v) When using fertility as the outcome, I restrict to graduation cohorts from 1965 and onwards. This is done due to incomplete parent-child linkages for children born prior to 1960. This imposes the implicit assumption that people who graduate in 1965 have not had any children 5 years prior to graduating from medical school.