Online Appendix for "Hours Constraints, Occupational Choice, and Gender: Evidence from Medical Residents"

Melanie Wasserman

B For Online Publication: Data Appendix

Below is a description of the data sources utilized in this paper.

Data Source	Description	Years	Purpose
American Medical Association Physician Masterfile	Administrative and survey data on the universe of physicians in the U.S.	Medical school graduates 1993-2010	Medium-term specialty outcome analysis
American Medical Association Graduate Medical Education Census Track Data	Survey of residency program directors regarding resdiency program attributes	Residency program attributes 1996-1999, 2001-2010	Initial specialty entry analysis; Residency program analysis
American Association of Medical Colleges Matriculating Student Questionnaire	Survey of incoming U.S. medical school students that asks about specialty plans	Medical school entrants 1996-2010	Specialty preference analysis
Baldwin et al. (2001)	Survey of first and second year medical residents regarding their hours worked	1998/9	Classification of specialties by pre-policy residency hours

American Medical Association (AMA) Physician Masterfile: The AMA Physician Masterfile is a compilation of administrative and survey data sources on all physicians in the U.S. A physician's entry into the Masterfile is triggered by (a) entry into a U.S. medical school or (b) entry into a U.S. accredited medical residency or fellowship program. In addition, physicians who graduated from international medical schools are added to the Masterfile if they are licensed to practice in the U.S., conduct research or teach in the U.S., or completed a program to prepare for residency/fellowship training in the U.S. or a program for fully trained physicians to practice in the U.S. The Masterfile contains information on medical training, including medical school attended, year of medical school graduation, and medical specialty. Medical specialty is initially sourced from residency and fellowship programs attended and can be updated through self-reports in the Annual Census of Physicians.

American Medical Association (AMA) Graduate Medical Education (GME) Census Track Data: The AMA GME Census Track Data is an annual survey administered to all AAMC accredited residency and fellowship program directors in the U.S. The survey asks questions regarding the vacancies and enrollment in the residency program, for each program year, as well as information about the compensation, benefits, program hours, policies and faculty member profiles. The survey years used in this paper are 1996-2010. The response rates to the survey range from 82% (in 2001) to 99% (in 1996). The survey questions utilized in this paper pertain to enrollment in the program, by gender, program director reported weekly hours, number of full-time male and female faculty members, resident salary, and the availability of parental leave and onsite childcare for residents.

Two versions of the Masterfile are utilized in this paper. This first has the basic information outlined above for all physicians in the U.S. (AMA Masterfile). The second contains information for physicians who completed medical school or residency training in California or Texas (CA/TX Masterfile), and additionally contains physician first and last name and information on residency training including the name of each residency program attended, the program's specialty and dates attended. The richer residency program information is used to validate the specialty information provided in the basic Masterfile.

American Association of Medical Colleges (AAMC) Matriculating Student Questionnaire (MSQ): The AAMC MSQ is a survey administered to all incoming first-year students in U.S. medical schools. The response rates to this survey range from 64% (in 2005) to 93% (in 1998). The survey asks questions regarding the medical school admission process, career plans in medicine, and medical school financing. The survey years used in this paper are 1998-2006 and 2009-2010. Years 2007 and 2008 were excluded since a specialty preference question was not included in the survey. The survey question utilized in this paper concerns plans for specialization. Prior to 2009, the question states: "Are you planning to become certified in one of the 25 general specialties listed below?" From 2009 onward, the question states: "As part of practicing medicine in the U.S., a licensed physician typically becomes certified in a specialty. Are you planning to become certified in one of the 25 general specialties listed below?" If the respondent answers "Yes" to this question, the respondent is asked a follow-up question regarding their preference among the 25 specialties. For each survey year, I obtained the number of responses associated with each specialty, by gender. I also obtained the number of individuals who responded that they were undecided or did not intended to specialize.

Baldwin et al. (2001): Baldwin Jr et al. (2003) conducts a nationally representative survey of first and second year medical residents during 1998 and 1999. From these responses, Baldwin Jr et al. (2003) computes the average hours worked per week in 21 specialties, 20 of which are used in this paper. I crosswalk the more detailed specialties in the AMA Masterfile to the coarse specialties found in Baldwin Jr et al. (2003) using the Dartmouth Atlas crosswalk. In some cases, there are multiple specialty paths that can lead to a sub-specialty. For example, in order to sub-specialize in Pediatric Radiology, an individual must first complete residency training in either Pediatrics or Radiology, and then complete a fellowship in Pediatric Radiology. Since I only observe an individual's sub-specialty (Pediatric Radiology) in the AMA Masterfile, and not the specialty completed during second year of residency, I must choose whether to assign the individual to Pediatrics or Radiology, which have differing hours requirements during residency. In such cases, I supplement the Dartmouth Atlas crosswalk with an empirical comparison of second year residency specialty and self-reported primary specialty from the CA/TX Masterfile. For physicians who report a sub-specialty as their primary specialty, I assign them the root specialty that is most common among all individuals in their sub-specialty, as observed in the CA/TX Masterfile. I exclude extremely small specialties that have insufficient observations to determine a most common root specialty, such as Nuclear Medicine and Aerospace Medicine. Fewer than one percent of non-missing observations in the AMA Masterfile cannot be classified into one of the 20 specialty categories. Primary specialty is missing in approximately three percent of observations.

References

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.

C For Online Publication: Methods of Statistical Inference

Standard Errors for Average Marginal Effects

The standard errors for the conditional logit *coefficients* are computed by the Stata command cmclogit, using the delta method. I manually compute the standard errors for the conditional logit *average marginal effects* via the delta method, as outlined in Greene (2012). The baseline conditional logit model is:

$$\Pr(C_{it} = s) = \frac{\exp(\lambda_1(\text{Hours}_{s,1999} \times \text{Transition}_t) + \lambda_2(\text{Hours}_{s,1999} \times \text{Post}_t) + \alpha_s)}{\sum_{s' \in S} \exp(\lambda_1(\text{Hours}_{s',1999} \times \text{Transition}_t) + \lambda_2(\text{Hours}_{s',1999} \times \text{Post}_t) + \alpha_{s'})} \tag{7}$$

I estimate the model using maximum likelihood. The coefficients of interest are λ_1 and λ_2 on the interaction terms ($\text{Hours}_{s,1999} \times \text{Transition}_t$) and ($\text{Hours}_{s,1999} \times \text{Post}_t$), respectively. I use the following procedure to compute average marginal effects and standard errors:

1. For each individual i in the estimation sample, I use the estimated coefficients to predict the probability $\hat{p}_{is}$ individual i chooses specialty s. Each individual i will have 20 predicted probabilities associated with the 20 specialty alternatives.

2. For each individual i in the estimation sample, I compute the marginal effects associated with the interaction terms ($\text{Hours}_{s,1999} \times \text{Transition}_t$) and ($\text{Hours}_{s,1999} \times \text{Post}_t$). The marginal effect associated with the term ($\text{Hours}_{s,1999} \times \text{Post}_t$) for individual i and specialty alternative s is: $ME_{is} = \frac{\partial p_{ist}}{\partial(\text{Hours}_{s,1999} \times \text{Post}_t)} = \hat{\lambda}_2 \times \hat{p}_{is}(1 - \hat{p}_{is})$.[44]

3. For each specialty s, I compute the average marginal effects associated with the interaction terms ($\text{Hours}_{s,1999} \times \text{Transition}_t$) and ($\text{Hours}_{s,1999} \times \text{Post}_t$) by averaging the individual-specific marginal effects from (2). $ME_s = \hat{\lambda}_2 \times \frac{1}{N} \sum_{i=1}^{N} \hat{p}_{is}(1 - \hat{p}_{is})$

4. To compute the average marginal effect, I average the 20 specialty-specific marginal effects from (3). This is equivalent to: $AME_{\lambda_2} = \hat{\lambda}_2 \times \frac{1}{N \times S} \sum_{s=1}^{S} \sum_{i=1}^{N} \hat{p}_{its} \times (1 - \hat{p}_{its})$

5. Standard errors are heteroskedastic robust. The Stata command *cmclogit* computes standard errors for coefficients via the delta method. I manually compute the standard errors associated with the average marginal effects also using the delta method. The manually programmed standard errors are identical to those computed by the Stata command *margins,* but are less computationally burdensome due to my usage of analytical derivatives instead of numerical derivatives. Following Green (2012), to compute standard errors associated with AME for λ_2, let $G_{is}(\lambda_2) = ME_{is}$. $G'_{is}(\lambda_2)$ is a vector of partial derivatives with respect to all model parameters, computed for each individual-specialty alternative in the data set. The asymptotic variance of AME_{λ_2} is estimated using:

$$Asy.Var(AME_{\lambda_2}) = [\bar{G}(\lambda_2)]\mathbf{Var}(\lambda_2)[\bar{G}(\lambda_2)]'$$

   where $\mathbf{Var}(\lambda_2)$ is the variance-covariance matrix for λ_2 and $\bar{G}(\lambda_2)$ is the sample average of the vector of partial derivatives, that is, $\bar{G}(\lambda_2) = \frac{1}{N \times S} \sum_{s=1}^{S} \sum_{i=1}^{N} G'_{is}(\lambda_2)$.

[44] An alternative method for computing the average marginal effect of interaction terms in nonlinear models has been proposed by Ai and Norton (2003). Their approach uses the cross-partial derivative of the choice probability rather than the partial derivative of the choice probability with respect to the interaction term. I reproduced the marginal effects using the Ai and Norton (2003) approach and they are almost identical to the marginal effects reported in the paper.

References

Ai, Chunrong and Edward C. Norton, "Interaction terms in logit and probit models," *Economics Letters*, 2003, *80* (1), 123–129.

Greene, William H., *Econometric Analysis, 7th Edition*, Prentice Hall, 2012.

D For Online Publication: Role of Pre-Existing Specialty Trends

In this Appendix, I provide descriptive evidence on which specialties experienced increased entry of women and men prior to the reform, 1993-2002. For each of the 20 specialties, I estimate a specialty-specific linear time trend, using either the conditional logit or the OLS specification (omitting Pathology as the outside option specialty for conditional logit, and including it for OLS). Figure D.1, plots the specialty-specific trends against a specialty's pre-policy hours. For women, we observe a positive relationship between specialty-specific linear time trends and a specialty's pre-policy hours. For men, we observe a negative relationship. It is this correlation between the linear time trends and the exposure to the reform that causes the coefficients of interest in my main specification to change with the inclusion of specialty-specific linear time trends. The relationship between pre-policy specialty-specific trends and pre-policy hours is the slope of the line of best fit in the figures. Using the OLS specification, from a regression of the pre-policy specialty-specific trends on pre-policy hours, $\text{Hours}_{s,1999}$, the coefficient on pre-policy hours is 0.018 (standard error: 0.040). The relationship is positive but statistically insignificant.

In order to assess whether there is overlap in the specialties experiencing increased female entry pre- and post-reform, I contrast the pre-trends with the change in female entry before and after the reform (Figure 6), which I have reproduced in the right panel Figure D.2 below. For women, high hours specialties such as Orthopedic Surgery and Urology exhibit positive growth both before and after the reform. However, specialties with either no trend or a negative pre-period trend, such as General Surgery, Otolaryngology, and Neurological Surgery, also show positive growth after the reform. For men, after the reform there is a continuation of the negative trend in Ob/Gyn, and continued positive growth in Anesthesiology. In contrast to the pre-period, there is increased entry into Pathology, Radiology and Physical Medicine. The negative trend in General Surgery appears to abate, and there is a positive growth in Neurological Surgery and Orthopedic Surgery.

To assess the extent to which specialty pre-trends can account for the increased entry of women after the reform, I control for extrapolated specialty-specific linear pre-trends. Specifically, separately for each specification (conditional logit and OLS) and data set (GME Census Track and AMA Masterfile):

1. Estimate specialty-specific linear pre-trends over the time period 1993-2002. Save the estimates in a new variable $\hat{\nu}$.

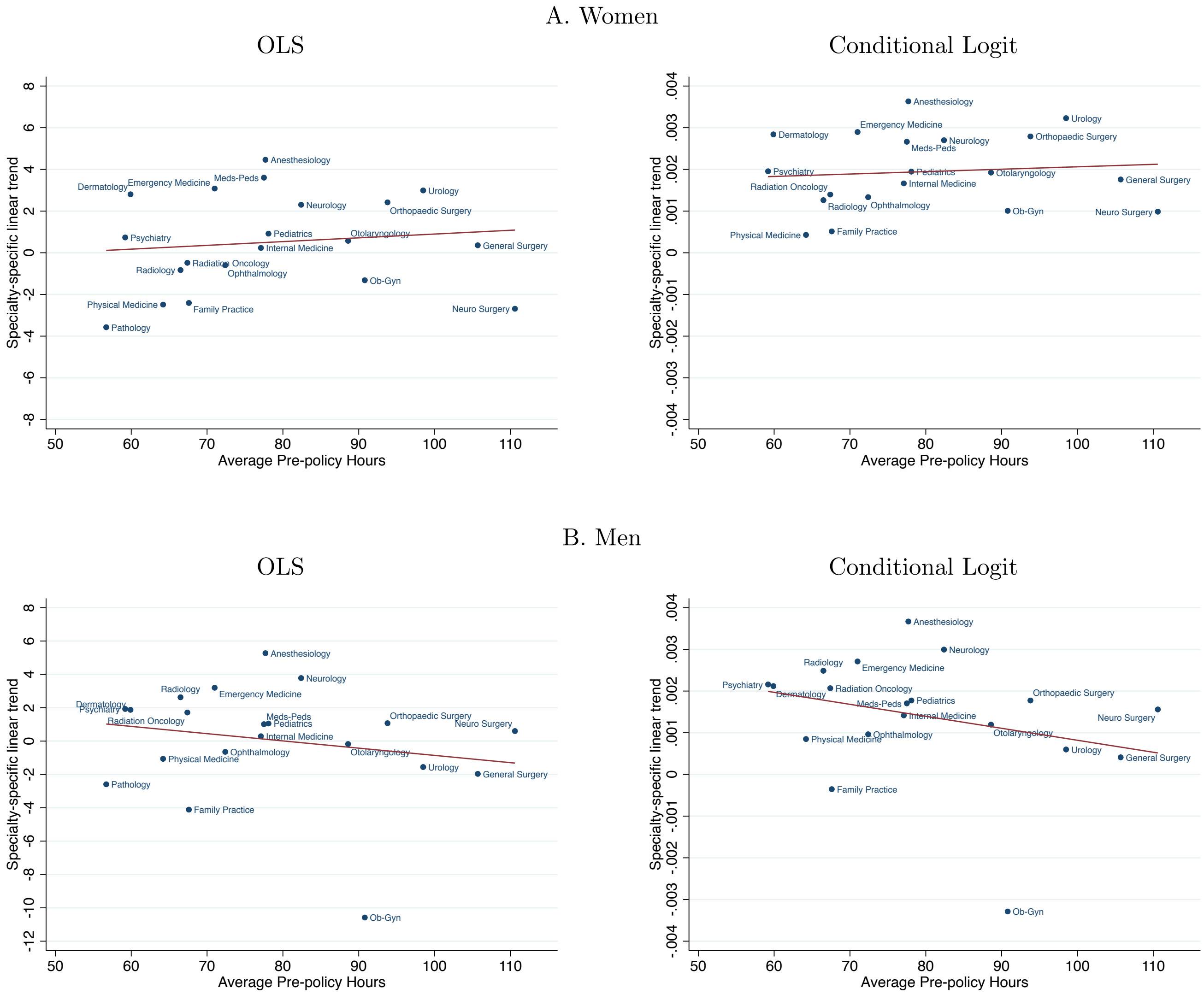
2. In specifications that estimate the effect of the reform on specialty choice, control for the interaction of $\hat{\nu}$ and a linear trend. This permits the specialty pre-trends to evolve linearly over the post-reform period.

Other papers that use a similar approach include Bhuller et al. (2013). The results are reported in in Table 2 and Appendix Table A.6. Note that the inclusion of extrapolated pre-trends (column 7) is a weaker test than directly including specialty-specific linear trends over the entire sample period (column 8). Since the direct inclusion of these specialty-specific trends over the entire sample period leaves little residual variation and causes the standard errors to rise substantially, it prevents drawing conclusions on the role of pre-existing specialty trends. Based on these estimates, a continuation of specialty-specific pre-trends can account for 39 percent of the increase in their medium-run specialty choices and 27 percent of the increased entry of women after the reform. The OLS results show that the pre-trends can explain 13 percent of eventual specialty choice and cannot explain any of the initial entry (the coefficient increases with their inclusion). Note that the inclusion of specialty-specific trends flips the sign of the coefficient for men. Prior to the reform, there was declining male entry into time-intensive specialties and controlling for this substantial negative trend reverses the sign of the coefficients of interest.

References

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

Figure D.1: Specialty-Specific Linear Time Trends, 1993-2002

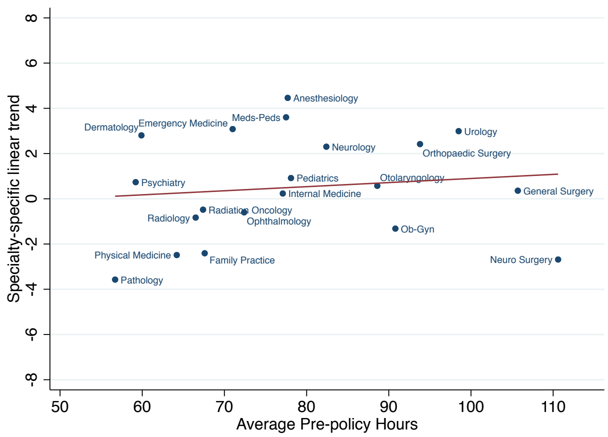


Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This figure plots the specialty-specific linear trends, 1993-2002, estimated using an OLS log share specification or a conditional logit specification. The estimated trends are plotted against a specialty's average pre-policy hours. The line of best fit is from a regression of the estimated coefficients on pre-policy hours.

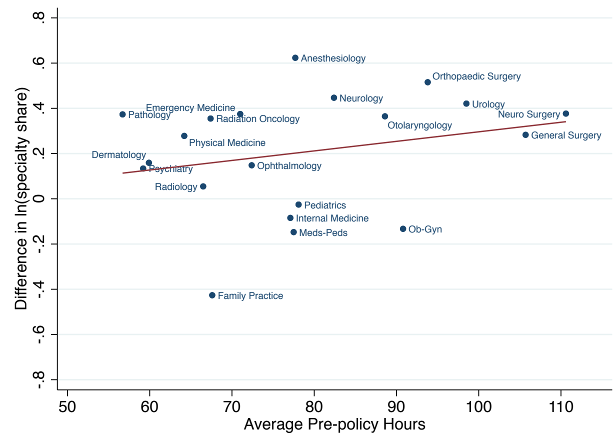
Figure D.2: Relationship between Change in Specialty Shares and Pre-Policy Specialty Hours

A. Women

Pre-period Trends (1993-2002)

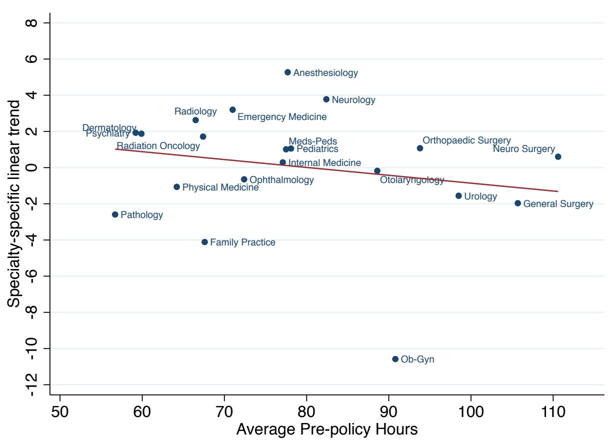


Pre/Post Difference

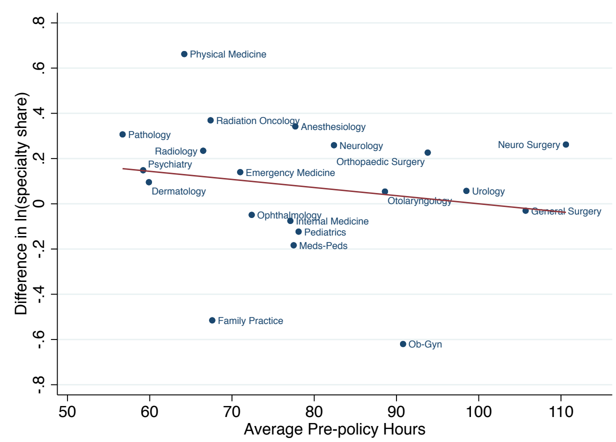


B. Men

Pre-period Trends (1993-2002)



Pre/Post Difference



Source: AMA Physician Masterfile, Baldwin Jr et al. (2003). Note: This figure plots the change in the natural log share of women and men entering a medical specialty before (1993-2002) and after (2006-2010) the reform against a specialty's average pre-policy hours, for the sample of U.S. medical school graduates, 1993-2010. The line of best fit is from a regression of the difference in log shares pre/post reform on pre-policy hours.