Gender Differences in Non-Promotable Tasks: The Case of Clinical Note-Taking

Bryan Chu, Ben Handel, Jon Kolstad, and Ulrike Malmendier

April 25, 2022

Abstract

Medicine has a reputation of being a gender-egalitarian profession, but there is also evidence on persistent differences in hours worked as well as procedures and tasks performed. We investigate gender differences in the intensive margin in detail using a unique dataset that contains granular information based on the Electronic Medical Records and Audit Log at a large teaching hospital. Our main sample contains 723 physicians, of which about 48% are women. In this highly standardized environment, we find that even after controlling for a detailed set of physician attributes, women spend about 18.5% more time on notes per shift than men. Preliminary results on the implications for patients suggest that patients who received with more detailed notes because their care team was more female receive fewer medical orders the following day, pointing to faster convergence towards the final diagnosis and thus positive welfare implications of more detailed note taking. These findings have implications not only for understanding gender inequities among physicians but also for understanding variation in patient outcomes.

1 Introduction

Much work has cited medicine as one of the most gender-egalitarian fields (cf. Goldin, 2014). However, it has been documented that women still lag behind in career advancement and salary (Sasser, 2005). For example, the gender gap in Obstetrics and Gynecology disappears after controlling for specialty, private vs. group practice, and procedures performed (Reyes, 2007). Therefore, it appears that differences remain along the intensive margin in medicine, not only by hours worked but also by procedures and tasks performed.

In some respects, these facts should not be surprising. A large literature has documented differences between men and women in domains that are potentially important for the practice of medicine such as competitiveness (Niederle and Vesterlund, 2011), task selection (Gneezy, Niederle and Rustichini, 2003; Buser, Niederle and Oosterbeek, 2014), speaking up (Thomas-Hunt and Phillips, 2004; Coffman, 2014), altruism and cooperation and risk tolerance (Niederle, 2014). In medicine specifically, Currie, et al. (2016) study clinical decision making in treating heart attack patients finds significant differences by gender with male cardiologists systematically making lower quality diagnosis and providing more intensive treatment to inappropriate patients.

Detailed understanding of differences in actions by gender has implications for understanding gender inequities in medicine and how they relate to work flexibility and burn out. Sarsons (2019) shows that surgeon gender impacts the way in which referring physicians interpret quality with implications for inequality in surgeon careers. At a broader level, flexibility, particularly with respect to child care, has been shown to play an important role for women in high skill professions (see Goldin and Mitchell (2017) for a review). The degree to which this flexibility plays a role in medicine, particularly in academic settings, is less well understood. Furthermore, little is known about how such flexibility plays out in day-to-day work.

We investigate these margins in a highly standardized environment at a single top teaching hospital. Leveraging unique Electronic Medical Record (EMR) and audit log data, we assess granular differences by gender in the timing and effort used by physicians. We then relate these to physician outcomes such as salary, promotion, and grants received. Finally, we examine the impact of these differences on clinical outcomes.

We find as in previous studies that women spend more time than men on note-taking (e.g. Gupta, et al. 2019). Even after controlling for a set of physician characteristics such as faculty title, specialty, and publications, we find that women spend 18.5% more time on notes per shift relative to their male counterparts. Categorizing things further, we find that differences are mainly due to note writing rather than note reading. Furthermore, we document hourly patterns in the differences: women spend more time taking notes during their shift, especially between 10am and 4pm, but not less time outside scheduled hours. Despite these findings, we do not find any significant findings for note-taking impact on salary, grants, or publications.

On the other hand, we find interesting implications for patients who are hospitalized. We first establish that there is a linear relationship between time spent writing notes and time spent reading notes by others and that this relationship does not differ by gender. Next, we find that patients who receive more detailed notes written about them because their physician team was more female receive 30% fewer medical orders the following day (although this is imprecisely estimated). We are working on characterizing these differences in more detail and breaking them down into diagnostic vs. therapeutic orders, as well as investigating the temporal distribution of orders. For now, it appears that this is a pure reduction and not simply reallocation or the "pushing forward" of the same set of orders but this is still work in progress.

This project relates primarily to three strands of literature. The first is on gender differences, as detailed above in the introduction. The second is on practice variation across physicians, such as Chandra and Staiger (2007), Molitor (2018), and Finkelstein, et al. (2021). A difference is that while these papers tend to study differences across locations, we focus on differences across physicians within the same hospital. Finally, we contribute to the growing literature, primarily in healthcare, that uses EMR and audit log data, such as Patel, et al. (2018) and Huilgol, et al. (2022).

The remainder of this paper proceeds as follows: Section 2 describes the data used and some descriptive statistics, Section 3 presents methodology and results on note taking, Section 4 presents methodology and results on physician outcomes, and Section 5 presents methodology and results on clinical outcomes. Section 6 concludes.

2 Data

Currently, our primary data consists of ten weeks of EMR and Audit Log data in 2018. This is in the process of being expanded to cover 104 weeks spanning the two year period 2018-2019. The data covers all patients who enter the hospital via the emergency department (ED) and follows them until they are discharged from the hospital[1]. Observations are grouped at the encounter level, which is one discrete hospital visit; the same patient can be present for multiple encounters and we also observe patient identifiers. For each encounter, we observe detailed information on note activity by their physicians, including exactly when and for how long each note was created, viewed, edited, and signed. We also observe the note length in characters for each version of each note. We also observe each Medical Order (medications, procedures) that is prescribed, including orders that are executed as well as orders that are subsequently canceled. When applicable (e.g. labs and imaging), we observe when order results are available, when they are viewed, by who, and for how long. For most labs we also observe a flag for whether the results are abnormal. For each of these actions, we observe provider identifiers.

We also observe a detailed set of patient characteristics. These include both relatively "typical" measures such as age, sex, and race, but also a set of unique "ex-ante" measures taken prior to physician intervention by the ED's triage nurse. These measures include things such as the patient's chief complaint that brought them to the ED that day, a set of indicators for abnormal vital signs, and the triage nurse's estimation of how urgent the patient's condition is (Emergency Severity Index, or ESI).

For clinical note activity only, we also observe all note activity (but no further details) by the physicians who treat patients who are in the sample. For example, consider a surgeon who operates on both patients admitted via the ED as well as outpatients. We would observe the full details of their note activity for their ED patients (as well as medical orders and everything else detailed above). We also observe the action, time, and duration for all other note activity that occurs during the ten-week sample period, but no details on which patients the notes were for, the note length, or anything else about the patients or procedures performed. Despite the fewer details, these additional observations are important as they allow us to avoid having to consider differential selection by gender into treating ED patients vs. outpatients.

We complement this data with Physician Compare, a dataset provided by CMS, that lists each clinician-group in the country with Medicare enrollments. We use Physician Compare for four items. First, we obtain a second elicitation of gender, which we use to validate our internal data. For the gender project, we only keep physicians for which the Physician Compare gender matches our internal data. Second, we obtain the physician's name, which we use to match physicians to shift scheduling data. Third, we obtain the medical school from which the physician graduated. Fourth, we obtain their medical school graduation year. We also merge to Sacramento Bee's California State Worker Salary Database for salary information, to PubMed for publications, and to NIH RePORTER for grants.

Throughout, we focus on attending physicians. As seen in Table 1, the full sample contain 1703 physicians, which we narrow down to 723 for our Note Taking (Section 3) and Physician Outcome (Section 4) analyses.

[1] Discharge is by far the most common outcome. Alternatives are leaving without being seen, leaving against medical advice, and (rarely) death.

Table 1: Physician Sample Selection for the Main Analysis

Restriction	Total users	Frac. female
All attendings	1661	0.489
...with note actions	1607	0.492
...match Physician Compare	1273	0.465
...internal vs. Physician Compare gender matches	1238	0.467
..merged to job titles	1113	0.465
...specialties with 2+ of each gender	947	0.463
...with at least one inferred shift of 4+ hours	723	0.476

For the Clinical Outcomes analysis (Section 5), we use a different set of encounters and physicians. Because we want to investigate downstream clinical utilization of notes, we focus on the hospital medicine setting. After patients are admitted to the hospital from the ED for inpatient care, they fall under the supervision of a hospital medicine team. Out of the full sample of 8,376 encounters, 1,236 (14.8%) have an inpatient admission. Table 2 summarizes some key encounter-level metrics.

Table 2: Inpatient Encounters for the Clinical Outcomes Analysis

	All Users	Sample Hospitalists
Patient Encounters	1236	
Unique Patients	1165	
Encounter Frac Female	0.46	
Encounter Frac Nonwhite	0.57	
Mean Days Hospitalized	6.21	
Edit Action Count	82.3	13.9
Edit Minutes Spent	376.3	76.6
Per Day: Edit Action Count	12.0	2.4
Per Day: Edit Minutes Spent	56.2	13.3
Orders Authorized	139.7	88.9
Per Day: Orders Authorized	22.2	13.9

Because we are focusing on the hospital medicine team, we limit our analysis to physicians who specialize in internal medicine. This leaves us with 157 physicians, with physician-level averages seen in Table 3.

3 Note Taking

In this section, we describe the differences in note-taking behavior between male and female physicians. We begin by collapsing all data to the physician level and regressing the log of minutes spent viewing and editing notes on an indicator for physician gender. We progressively add more controls, such as the physician's specialty, their faculty title (ex. Clinical Instructor; Assistant/Associate/Full Professor) and grant and publication details (proxies for the degree of research vs. clinical involvement). As seen in Table 4, the coefficient on male is stable across specifications. In our preferred specification, column (12), the value

Table 3: Internal Medicine Physicians for the Clinical Outcomes Analysis

	All	Men	Women
Patient Encounters	1236		
Number (Fraction) Attending Hospitalists	157	0.43	0.57
Mean Patient-Days Worked	40.6	37.8	42.1
Edits Action Count	82.09	75.06	86.88
Edits Hours Spent	7.52	6.10	8.61
Per Patient-Day: Edits Action Count	2.34	2.10	2.53
Per Patient-Day: Edits Hours Spent	0.22	0.18	0.25
Orders Authorized	850	648	996
Per Patient-Day: Orders Authorized	16.6	13.9	18.7
Orders Signed	202	179	224
Per Patient-Day: Orders Signed	5.4	4.6	6.0

of -0.186 indicates that male physicians spend 18.6 log points, or approximately 18.6% less time viewing and editing notes per shift than their female counterparts. This is somewhat surprising since we are comparing physicians who work at the same hospital, in the same division (specialty), with the same job description (faculty title), and clinical involvement (grants and publications).

Table 4: Time Spent Viewing and Editing Notes Per Shift

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	log(minutes spent viewing and editing)											
male	-0.225***	-0.238***	-0.222***	-0.192***	-0.233***	-0.221***	-0.227***	-0.265***	-0.204***	-0.200***	-0.195***	-0.186***
	(0.0817)	(0.0849)	(0.0820)	(0.0678)	(0.0822)	(0.0821)	(0.0818)	(0.0818)	(0.0699)	(0.0703)	(0.0704)	(0.0705)
any grant					0.0950	1.035*					-0.144	0.548
					(0.117)	(0.535)					(0.107)	(0.576)
log(award amount)						-0.0695*						-0.0513
						(0.0404)						(0.0426)
Constant	5.428***	6.307***	5.500***	5.334***	5.414***	5.408***	5.588***	6.341***	6.018***	6.153***	6.250***	6.171***
	(0.0571)	(0.0849)	(0.0999)	(0.133)	(0.0582)	(0.0581)	(0.0961)	(0.630)	(0.211)	(0.254)	(0.267)	(0.277)
Faculty Title FE		Y							Y	Y	Y	Y
Publications FE			Y							Y	Y	Y
Specialty FE				Y					Y	Y	Y	Y
Med School Rank FE							Y					
Grad Decade FE								Y				
Observations	723	723	723	723	723	723	723	721	723	723	723	723
R-squared	0.010	0.066	0.021	0.386	0.011	0.016	0.023	0.030	0.419	0.428	0.430	0.432
Adjusted R-squared	0.009	0.024	0.013	0.352	0.009	0.011	0.014	0.022	0.359	0.364	0.365	0.366

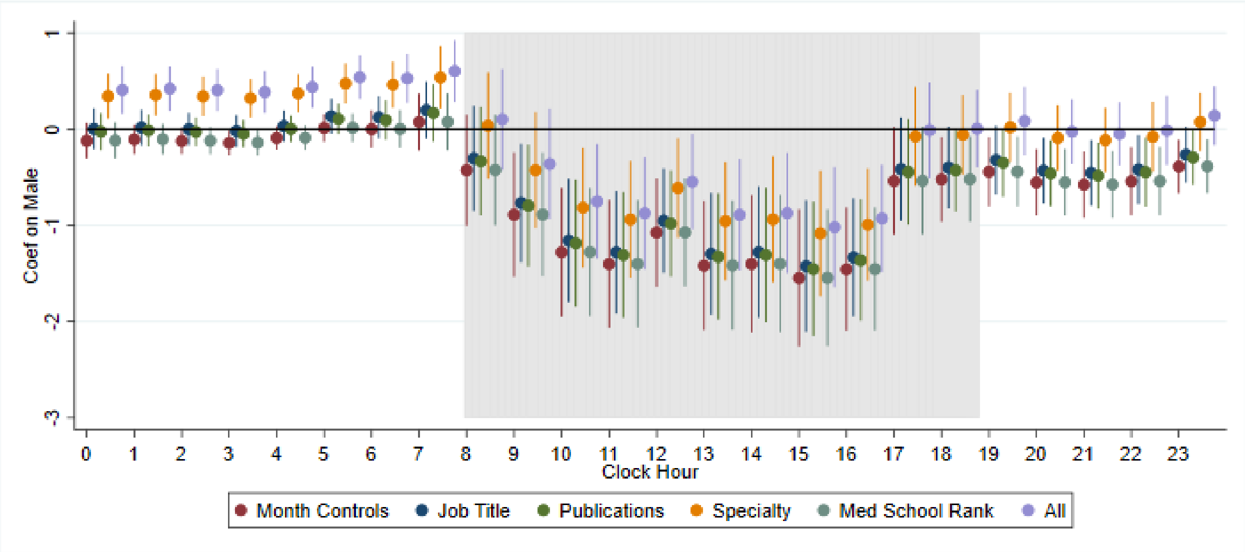
Standard errors in parentheses
* p<0.10 ** p<0.05 *** p<0.01

We next examine differences within the day. We estimate regressions of the form

$$y_{it} = \alpha + \beta_0 male + \sum_{k=1}^{23} [\delta_k hour_k + \beta_k male \cdot hour_k] + X_{it}'\gamma + \varepsilon_{it} \tag{1}$$

Each observation is a physician i hour t. Y_{it} are outcomes (chiefly minutes spent or the number of discrete actions taken). The coefficients of interest are β_k: how much more men work on notes than women, during each hour of the day (k subscript). X_{it} are physician level controls (job title, specialty, publications, med school rank) as well as month controls (January, Apr, June, July, Oct). Plotted are $\beta_0 + \beta_k$, so the figure preserves the overall male-female difference. β_0 is represented in hour 0; the others are the hourly coefficients $\beta_k + \beta_0$. In Figure 1, the LHS is minutes spent, so in hour 0 (0.4) means men spend 0.4 minutes more than women per day at midnight. The shaded area represents the modal shift for attendings derived from our shift data.

Figure 1: Breakdown of Note-Taking Differences Throughout the Day: Minutes Spent



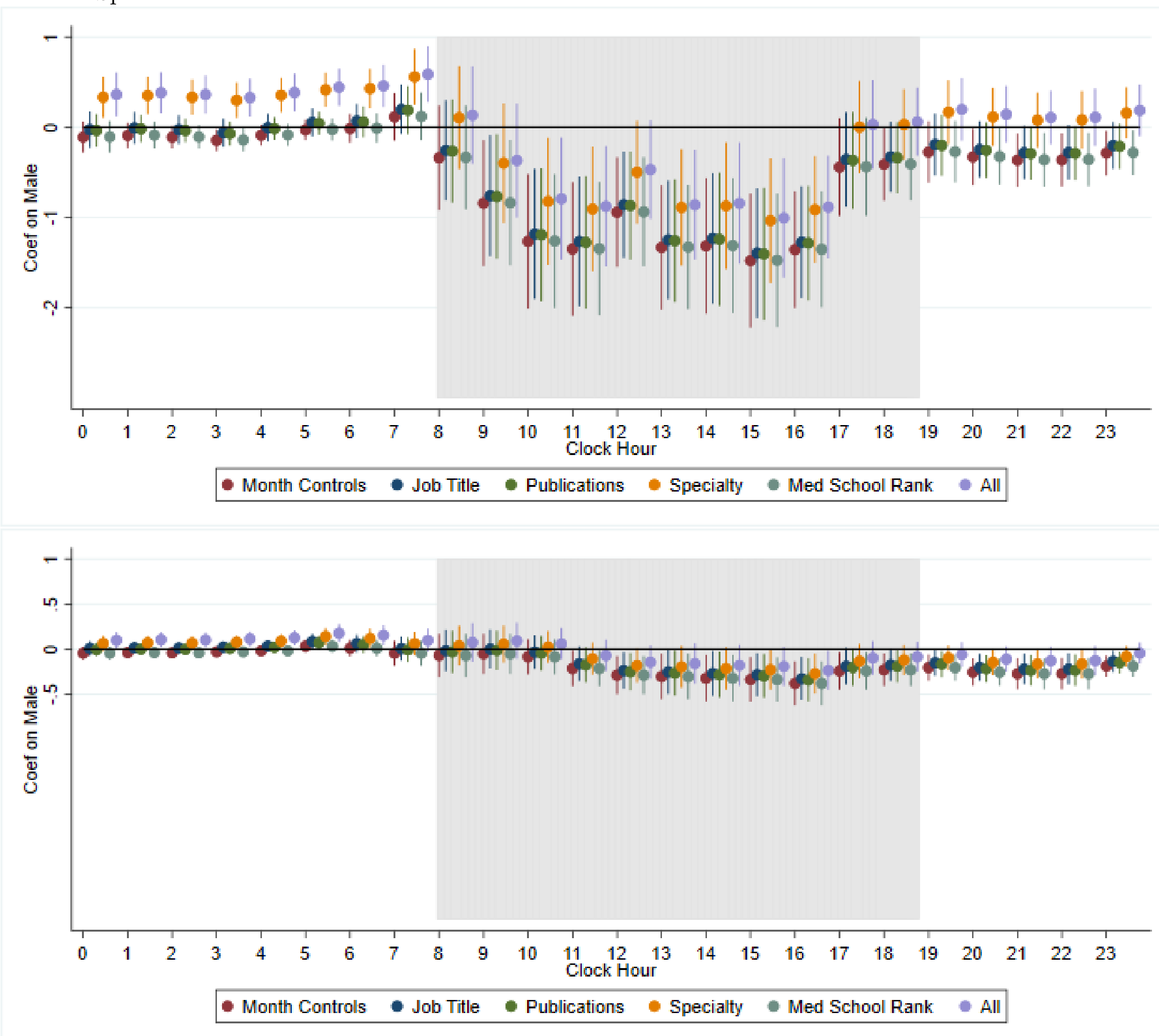
We can further break this down into minutes spent editing (writing) notes compared to time spent viewing notes. As seen in Figure 2, differences are driven by time spent editing notes rather than referring to notes written by others.

Given the additional time women spend writing notes, are there differences in note-writing "efficiency?" We investigate this by regressing the total number of characters added on total time spent editing notes and including an interaction for male times time spent editing. That is, we estimate

$$\log(CharsAdded_i) = \alpha + \beta_1 \log(MinsEditing_i) + \beta_2 male_i + \beta_3 male_i \cdot \log(MinsEditing_i) + X_i'\gamma + \varepsilon_i \quad (2)$$

Each observation is a physician. The coefficients of interest are β_2 and β_3, which represent differences in "characters per minute" between men and women. As can be seen in Table 5, we find (imprecise) zeros on these coefficients, indicating that conditional on the amount of time spent, women and men write similar length notes on average. Therefore, to the extent that note length is a proxy for note content, the additional time women spend writing notes results in more information being conveyed to the physicians who later read these notes. We return to this topic in Section 5.

Figure 2: Breakdown of Note-Editing (top) and Note-Viewing (bottom) Differences Throughout the Day: Minutes Spent



4 Physician Outcomes

In the previous section, we established that women spend more time on documentation than men. Do these differences result in meaningful differences for the physicians? We investigate three key outcomes: salary, grants, and publications. For salary, we investigate both salary levels in the years around our data as well as salary growth over time. Although salary could be viewed as a sufficient statistic for career advancement, we also investigate the research output by the physicians. This is an important margin because research is a key contributor to moving up the academic job ladder. Furthermore, physicians can "buy out" a portion of their clinical teaching responsibilities with grant funding. We measure research output two ways: first by counting the number of publications in 2019 (recall our data is from 2018, and that medical publications have a far faster turnaround time compared to economics publications) and second by tabulating both the

Table 5: Characters Added to Notes vs. Time Spent Editing for Men and Women

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES			log(characters added)			
log(minutes editing)	0.902***	0.847***	0.918***	0.916***	0.904***	0.899***
	(0.0719)	(0.0856)	(0.0604)	(0.0601)	(0.0579)	(0.0584)
male	0.537	0.388	0.780*	0.755	0.698	0.675
	(0.545)	(0.548)	(0.460)	(0.465)	(0.453)	(0.452)
male X log(minutes editing)	-0.0865	-0.0473	-0.106	-0.0976	-0.0839	-0.0824
	(0.0912)	(0.0900)	(0.0755)	(0.0767)	(0.0734)	(0.0733)
Constant	7.937***	8.904***	10.03***	10.23***	8.924***	8.949***
	(0.452)	(0.554)	(0.556)	(0.642)	(1.010)	(1.001)
Observations	439	439	439	439	438	438
R-squared	0.523	0.649	0.685	0.689	0.702	0.703
Specialty FE		Y	Y	Y	Y	Y
Job Title FE			Y	Y	Y	Y
Publications FE				Y	Y	Y
Grad Decade FE					Y	Y
Grant Controls						Y

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

amount of past grants obtained in the past 10 years as well as the number of future grants obtained in 2019 and 2020.

We take two complimentary approaches to investigating these margins. First, we estimate OLS models of the form

$$y_i = \alpha + \beta_1 male + \beta_2 \log(MinsPerShift_i) + \beta_3 male_i \cdot \log(MinsPerShift_i) + X_i'\gamma + \varepsilon_i \tag{3}$$

Each observation is a physician, and y_i are our salary, grant, and publication outcomes. The coefficients of interest are β_1, which is the gender gap, as well as β_3, which represents differential returns to note taking realized by men vs. women.

Because there may still be endogeneity issues even after we include our detailed physician-level controls X_i, we also pursue an "IV Wald" specification. Here, we instrument for the log of note minutes per shift with an indicator for male in order to isolate the difference in note taking due to gender. As such, the first stage is

$$\log(MinsPerShift_i) = \pi_0 + \pi_1 male_i + X_i'\pi_3 + v_i \tag{4}$$

and the second stage is

$$y_i = \alpha + \beta \log(\widehat{MinsPerShift_i}) + X_i'\gamma + u_i \tag{5}$$

We are currently rebuilding our salary data so we do not currently have results on it. However, we do have results for grants and publications.

We begin with grants. Our two outcomes of interest are (1) whether physicians have obtained any grant in the past 10 years, and (2) whether the physicians receives a "future grant" in the next two years after we observe their activity. Physicians are eligible for this analysis if they belong to the academic medicine job ladder (e.g. they are Assistant, Associate, or Full Professors). We begin with the OLS model in Table 6, which in this case are linear probability models. Although we find a small gender gap in past grants (columns 1-4) and a statistically insignificant gap in future grants (columns 5-8), there is little evidence that men and women experience a differential return to note-taking. If anything, men experience a greater penalty for note taking than women (negative coefficient on the interation of male and log(notes per shift)).

Table 6: Gender Differences in Grants, OLS Specification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Any Grant in the past 10 years				Any Grant in the next two years			
male	0.463*	0.384	0.436*	0.445*	0.247	0.202	0.225	0.234
	(0.231)	(0.241)	(0.247)	(0.254)	(0.196)	(0.210)	(0.211)	(0.212)
log(note minutes per "shift")	0.0485	-0.00629	0.0117	0.0224	0.00953	-0.0407	-0.0395	-0.0298
	(0.0426)	(0.0559)	(0.0575)	(0.0552)	(0.0319)	(0.0425)	(0.0433)	(0.0409)
male X log(notes per "shift")	-0.0667	-0.0521	-0.0646	-0.0674	-0.0361	-0.0285	-0.0325	-0.0350
	(0.0452)	(0.0483)	(0.0482)	(0.0491)	(0.0379)	(0.0409)	(0.0410)	(0.0409)
log("shifts")				-0.0650***				-0.0592***
				(0.0166)				(0.0172)
Constant	-0.0686	0.142	-0.0684	-0.0200	0.0949	0.287	0.0475	0.0916
	(0.221)	(0.275)	(0.271)	(0.270)	(0.170)	(0.210)	(0.211)	(0.206)
Specialty FE		Y	Y	Y		Y	Y	Y
Grad Decade FE			Y	Y			Y	Y
Observations	607	607	607	607	607	607	607	607
R-squared	0.018	0.173	0.198	0.212	0.007	0.142	0.149	0.164
Adjusted R-squared	0.013	0.109	0.129	0.143	0.002	0.076	0.075	0.091

Standard errors in parentheses
* p<0.10 ** p<0.05 *** p<0.01

Moving on to the IV Wald analysis, we fail to find significant differences in past or future grants via this specification. Part of the issue may be that our first stage is relatively weak, complicating the interpretation of the IV coefficients and standard errors. However, if we are willing to take the signs of the coefficient as correct, this would indicate that women do experience a penalty in grants due to the additional time they spend on notes relative to men.

Next, we replicate this analysis for the log of the number of publications in 2019. We include in this sample any "actively publishing" physicians. We define these as physicians who have publications in 2019 or had any publications in 2015-2018. We deal with zeros by taking log of the number of publications plus one[2]. In the OLS specification (Table 9), none of our coefficients of interest are statistically significant, although the point estimates suggest a gender gap in favor of men as well as a larger penalty for note taking for men. On the other hand, the IV Wald results (Table 10) suggest a large penalty for women. We are in the process of investigating drivers behind the differences between our OLS and IV Wald results.

[2]Replicating this using the inverse hyperbolic sine (IHS) is in progress.

Table 7: Gender Differences in Past Grants, IV Wald Specification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	log(note minutes per "shift")	Any Grant 2008-2018		log(note minutes per "shift")	Any Grant 2008-2018		log(note minutes per "shift")	Any Grant 2008-2018		log(note minutes per "shift")	Any Grant 2008-2018	
	1S	OLS	IV	1S	OLS	IV	1S	OLS	IV	1S	OLS	IV
male	-0.235***			-0.149***			-0.120*			-0.113*		
	(0.0630)			(0.0548)			(0.0630)			(0.0619)		
log(note minutes per "shift")		0.000437	-0.447*		-0.0473	-0.749*		-0.0338	-0.789		-0.0240	-0.779
		(0.0322)	(0.229)		(0.0389)	(0.440)		(0.0377)	(0.583)		(0.0355)	(0.624)
log("shifts")										0.0709	-0.0670***	-0.0104
										(0.0443)	(0.0163)	(0.0563)
Constant	5.420***	0.250	2.619**	5.014***	0.415**	3.875*	4.805***	0.263	3.796	4.695***	0.313	3.767
	(0.113)	(0.161)	(1.204)	(0.0316)	(0.192)	(2.171)	(0.505)	(0.183)	(2.564)	(0.471)	(0.187)	(2.680)
Specialty FE				Y	Y	Y	Y	Y	Y	Y	Y	Y
Grad Decade FE							Y	Y	Y	Y	Y	Y
Observations	607	607	607	607	607	607	607	607	607	607	607	607
R-squared	0.024	0.000	.	0.376	0.157	.	0.412	0.186	.	0.418	0.201	.
Adjusted R-squared	0.023	-0.002	.	0.331	0.096	.	0.363	0.119	.	0.369	0.134	.
F	13.92	0.000184		.	.		.	.		.	.	

Standard errors in parentheses
* p<0.10 ** p<0.05 *** p<0.01

Table 8: Gender Differences in Future Grants, IV Wald Specification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	log(note minutes per "shift")	Any Grant 2019-2020		log(note minutes per "shift")	Any Grant 2019-2020		log(note minutes per "shift")	Any Grant 2019-2020		log(note minutes per "shift")	Any Grant 2019-2020	
	1S	OLS	IV	1S	OLS	IV	1S	OLS	IV	1S	OLS	IV
male	-0.235***			-0.149***			-0.120*			-0.113*		
	(0.0630)			(0.0548)			(0.0630)			(0.0619)		
log(note minutes per "shift")		-0.0161	-0.243		-0.0624**	-0.395*		-0.0629**	-0.495		-0.0541**	-0.469
		(0.0198)	(0.148)		(0.0262)	(0.228)		(0.0264)	(0.303)		(0.0235)	(0.320)
log("shifts")										0.0709	-0.0603***	-0.0292
										(0.0443)	(0.0171)	(0.0314)
Constant	5.420***	0.263**	1.461*	5.014***	0.429***	2.067*	4.805***	0.223	2.246*	4.695***	0.268**	2.167
	(0.113)	(0.107)	(0.791)	(0.0316)	(0.129)	(1.125)	(0.505)	(0.133)	(1.365)	(0.471)	(0.123)	(1.400)
Specialty FE				Y	Y	Y	Y	Y	Y	Y	Y	Y
Grad Decade FE							Y	Y	Y	Y	Y	Y
Observations	607	607	607	607	607	607	607	607	607	607	607	607
R-squared	0.024	0.001	.	0.376	0.137	.	0.412	0.144	.	0.418	0.160	.
Adjusted R-squared	0.023	-0.001	.	0.331	0.075	.	0.363	0.073	.	0.369	0.089	.
F	13.92	0.658		.	.		.	.		.	.	

Standard errors in parentheses
* p<0.10 ** p<0.05 *** p<0.01

5 Clinical Outcomes

In Section 3, we established that there are economically significant differences in the note-taking behavior of men and women, and that this difference is driven chiefly by women spending more time writing notes. Although we failed to find much evidence of differences in outcomes for the physicians themselves in Section 4, we now examine whether there are implications for their patients. Before we begin, we should note that this section is very much work in progress. First, we examine whether other physicians spend a commensurate amount of time reading longer notes (predominantly written by women) compared to shorter notes (predominantly written by men). Next, we turn to the hospital medicine setting and investigate whether the additional note content "yesterday" affects what is done "today."

Previously, we established that conditional on the amount of time spent writing notes, men and women write notes of similar length. However, it still may be that these notes are differentially "useful." As a first

Table 9: Gender Differences in 2019 Publications, OLS Specification

	(1)	(2)	(3)	(4)
	log(publications)			
male	0.497	0.510	0.655	0.656
	(0.575)	(0.557)	(0.590)	(0.584)
log(note minutes per "shift")	-0.116	-0.127*	-0.0415	-0.0479
	(0.0808)	(0.0724)	(0.0870)	(0.0879)
male X log(notes per "shift")	-0.0393	-0.0568	-0.0901	-0.0899
	(0.1000)	(0.0988)	(0.107)	(0.106)
log("shifts")				0.0409
				(0.0491)
Constant	3.046***	2.475***	1.768**	1.745**
	(0.437)	(0.394)	(0.761)	(0.742)
Specialty FE		Y	Y	Y
Grad Decade FE			Y	Y
Observations	481	481	481	481
R-squared	0.043	0.233	0.283	0.284
Adjusted R-squared	0.037	0.157	0.203	0.203

Standard errors in parentheses
* p<0.10 ** p<0.05 *** p<0.01

Table 10: Gender Differences in 2019 Publications, IV Wald Specification

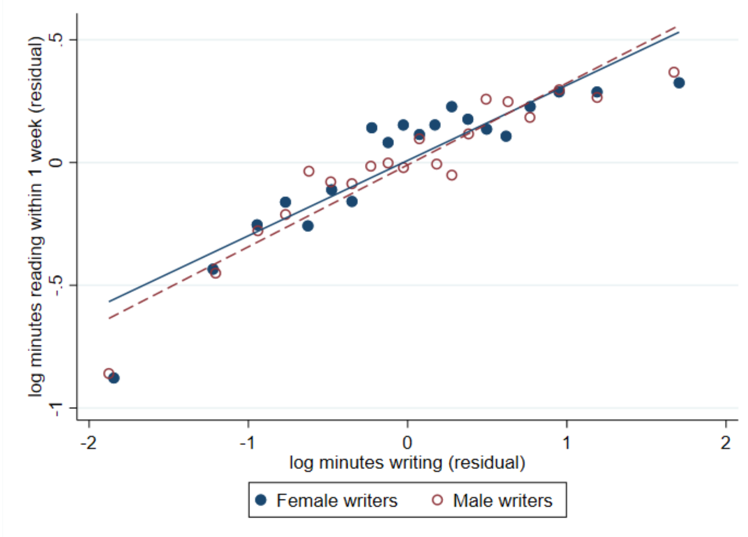
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	log(note minutes per "shift")	log(publications)		log(note minutes per "shift")	log(publications)		log(note minutes per "shift")	log(publications)		log(note minutes per "shift")	log(publications)	
	1S	OLS	IV	1S	OLS	IV	1S	OLS	IV	1S	OLS	IV
male	-0.206***			-0.104**			-0.0892**			-0.0842*		
	(0.0511)			(0.0411)			(0.0431)			(0.0425)		
log(note minutes per "shift")		-0.167**	-1.322***		-0.184***	-1.355**		-0.110*	-1.303*		-0.116*	-1.377*
		(0.0666)	(0.495)		(0.0527)	(0.621)		(0.0575)	(0.751)		(0.0586)	(0.792)
log("shifts")										-2.137*	-0.00109	-0.00478
										(1.201)	(0.00378)	(0.00490)
Constant	5.466***	3.477***	9.539***	5.131***	2.873***	8.365***	4.952***	2.297***	7.592**	4.857***	2.279***	7.759**
	(0.121)	(0.348)	(2.595)	(0.0226)	(0.247)	(2.914)	(0.410)	(0.540)	(3.042)	(0.371)	(0.526)	(3.110)
Specialty FE				Y	Y	Y	Y	Y	Y	Y	Y	Y
Grad Decade FE							Y	Y	Y	Y	Y	Y
Observations	918	481	481	918	481	481	916	481	481	916	481	481
R-squared	0.018	0.019	.	0.339	0.221	.	0.381	0.274	.	0.386	0.275	.
Adjusted R-squared	0.017	0.017	.	0.308	0.148	.	0.349	0.197	.	0.353	0.196	.
F	16.27	6.316		.	.		.	.		.	.	

Standard errors in parentheses
* p<0.10 ** p<0.05 *** p<0.01

pass, we proxy for usefulness by the amount of time others spend reading the note in the next week by other providers. We restrict to providers in order to exclude non-medical staff such as coders who may also read the note, but not for clinical purposes. The temporal restriction to the next week is also to restrict to clinical purposes, as some notes are read by other providers much later as a teaching example. The results are summarized in Figure 3, which plots the log of minutes reading notes by others vs. log of minutes spent writing the note. Both are residualized for note type. The relationship is surprisingly linear, and notably there appear to be little difference between notes written by women (solid blue line and filled circles) vs.

notes written by men (dashed maroon line and hollow circles).

Figure 3: Minutes Reading Notes vs. Minutes Writing Notes for Men and Women



The OLS regression results in Table 11 confirm our visual interpretation, as both the coefficient on male and the coefficient on male interacted with time spent writing are close to zero. Column 2 controls for note type and is what is plotted in Figure 3. Column 3 further adds indicators for final diagnosis group in order to control for ex-post patient type. There is a large difference in the coefficient on log minutes writing when including the fixed effects for note type, but that is not a concern for us. While some notes take much longer to write and are also more frequently read (History and Physical Examination notes in particular), we would only be concerned for our purposes if controlling for note type changes the coefficient on the interaction of male and minutes writing. However, that is not the case.

The above results provide suggestive evidence that longer notes written by women are not differentially less useful and that longer notes by women are not read differently. However, we have more direct measures of the clinical value of the notes: the clinical orders that are prescribed by physicians. We follow the IV Wald strategy outlined in Section 4 and instrument for note edits with fraction of the care team that is female. We then ask the following question: if a patient had a longer note written about them on day $t-1$ because their care team was more female, then is there an impact on orders (medication and procedures) on day t? If longer notes have clinical value, then we might see fewer orders the following day.

Our first stage is

$$\log(edits_{it}) = \alpha_0 + \alpha_1 FracFemale_{it} + \alpha_2 MaleAttending_{it} + X_{it}'\alpha_3 + v_{it} \tag{6}$$

and the corresponding IV is

$$\log(orders_{it}) = \beta_0 + \beta_1 \log(\widehat{edits}_{it-1}) + \beta_2 MaleAttending_{it} + X_{it}'\beta_3 + u_{it} \tag{7}$$

Each observation is an encounter-day (i-t). $X_i t$ contains controls such as final diagnosis group, day of stay FE, consults, and lagged log(orders). The coefficient of interest is β_1: how much orders on day t change

Table 11: Minutes Reading Notes vs. Minutes Writing Notes for Men and Women

	(1)	(2)	(3)
	log(minutes reading within 1 week)		
male	-0.359	0.0287	0.0382
	(0.354)	(0.185)	(0.136)
log(minutes writing)	0.541***	0.326***	0.266***
	(0.0367)	(0.0353)	(0.0293)
male X log(minutes writing)	0.109	-0.0190	-0.000757
	(0.102)	(0.0525)	(0.0432)
Constant	1.162***	2.511***	1.751***
	(0.111)	(0.116)	(0.203)
Note Type FE		X	X
Dx group FE			X
Observations	4827	4827	4433
R-squared	0.244	0.513	0.564
Adjusted R-squared	0.243	0.511	0.551

Standard errors in parentheses
* p<0.10 ** p<0.05 *** p<0.01

by given differences in edits on day $t-1$. Because we instrument, this "Wald Statistic" gives the change in orders caused by additional notes written about the patient solely because the care team was more female.

In Table 12 we find suggestive evidence that this is true. In our preferred specification, Column (12), we see that a 10% change in note edits on day t-1 is associated with a 2.9% reduction in orders on day t.

Table 12: Clinical Impact of Longer Notes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	ln(edits_t-1)		ln(all orders)_t		ln(edits)_t		ln(all orders)_t		ln(edits)_t		ln(all orders)_t	
	1S	OLS all	OLS nonzero	IV	1S	OLS all	OLS nonzero	IV	1S	OLS all	OLS nonzero	IV
fraction edits by women_t	0.151***				0.164***				0.169***			
	(0.0522)				(0.0528)				(0.0520)			
ln(edits)_t-1		-0.00194	0.0982***	-0.629		0.000120	0.00939	-0.321		0.00235	0.000208	-0.293
		(0.00856)	(0.0196)	(0.398)		(0.0104)	(0.0167)	(0.254)		(0.00973)	(0.0164)	(0.233)
Male Attending_t	-0.00292	-0.00287	0.0299	-0.0318	-0.00325	-0.0154	-0.000183	-0.0290	-0.00604	-0.0104	-0.00866	-0.0356
	(0.0491)	(0.0212)	(0.0346)	(0.0564)	(0.0488)	(0.0176)	(0.0312)	(0.0425)	(0.0485)	(0.0171)	(0.0312)	(0.0412)
ln(all orders)_t-1					0.168***	0.215***	0.218***	0.273***	0.162***	0.198***	0.208***	0.255***
					(0.0302)	(0.0142)	(0.0257)	(0.0514)	(0.0302)	(0.0138)	(0.0249)	(0.0455)
Constant	2.496***	2.138***	1.835***	3.742***	1.548***	1.523***	1.542***	1.603***	1.469***	1.417***	1.370***	1.412***
	(0.0508)	(0.0182)	(0.0609)	(1.038)	(0.517)	(0.140)	(0.146)	(0.416)	(0.521)	(0.134)	(0.147)	(0.374)
Dx group FE					X	X	X	X	X	X	X	X
Day of stay FE					X	X	X	X	X	X	X	X
Lag Specialty any create/edit FE					X	X	X	X	X	X	X	X
Specialty any create/edit FE									X	X	X	X
Observations	2306	7173	2306	2306	2306	7173	2306	2306	2306	7173	2306	2306
R-squared	0.007	0.000	0.014	.	0.174	0.206	0.247	0.114	0.180	0.256	0.300	0.196
Adjusted R-squared	0.006	-0.000	0.013	.	0.119	0.188	0.197	0.055	0.121	0.237	0.250	0.138
F	3.594	0.0248	8.631									

Standard errors in parentheses
* p<0.10 ** p<0.05 *** p<0.01

We are still investigating the particulars of this relationship. It appears to be driven by medication orders, which is a bit strange given that our narrative and priors are that notes are a substitute for diagnostic orders. We also do not see much evidence of shifts in time of our inferred final set of diagnostic orders (via Kaplan-Meier Survival figures or Cox Proportional Hazard models, not shown).

6 Discussion and Conclusion

We have leveraged detailed data to show that even within a single, highly standardized environment, male and female attending physicians exhibit meaningful differences in how they carry out their jobs. Women spend about 18.5% more time on notes per shift than men with the same specific job description as them. Although we do not find that these behavioral differences lead to differential outcomes on their salary, grants, or publications, we cannot currently rule out earlier exit by women, potentially related to burnout (cf. Patel et al., 2018). This complements existing work showing differences along the intensive margin in medicine and may be important in jobs across the economy. We are continuing to investigate a promising margin in which this differential note activity, primarily driven by women spending more time to write longer notes, has positive effects on patients. If true, this may shed light on another source of variation in variation in both patient outcomes and in treatment styles (e.g. Chandra and Staiger 2007). Our findings have implications not only within medicine: they may speak to the gender gap across the economy more generally especially if things such as job titles are insufficient to fully describe heterogeneity in tasks performed by men and women, as well as potentially relating to productivity differences across otherwise observationally similar firms.

References

Buser, Niederle, and Oosterbeek. 2004. “Gender, Competitiveness, and Career Choices.” *The Quarterly Journal of Economics.*

Chandra and Staiger. 2007. “Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks.” *Journal of Political Economy.*

Coffman. 2014. “Evidence on Self-Stereotyping and the Contribution of Ideas.” *The Quarterly Journal of Economics.*

Currie, MacLeod, and Van Parys. 2016. “Provider Practice Style and Patient Health Outcomes: The Case of Heart Attacks.” *Journal of Health Economics*

Finkelstein, Gentzkow, Li, and Williams. 2021. ”What Drives Prescription Opiod Abuse? Evidence from Migration.” *NBER Working Paper.*

Gneezy, Niederle, and Rustichini. 2003. “Performance in Competitive Environments: Gender Differences.” *The Quarterly Journal of Economics*

Goldin, Claudia. 2014. “A Grand Gender Convergence: Its Last Chapter.” *American Economic Review*

Goldin and Mitchell. 2017. “The New Life Cycle of Women’s Employment: Disappearing Humps, Sagging Middles, Expanding Tops.” *Journal of Economic Perspectives.*

Gupta, Murray, Sarkar, Mourad, and Adler-Milstein. 2019. “Differences in Ambulatory EHR Use Patterns for Male vs. Female Physicians.” *NEJM Catalyst.*

Huilgol YS, Adler-Milstein J, Ivey SL, Hong JC. 2022. “Opportunities to use electronic health record audit logs to improve cancer care.” *Cancer Med.* 2022 Mar 29.

Molitor, David. 2018. “The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration.” *American Economic Journal: Economic Policy.*

Niederle and Vesterlund. 2011. “Gender and Competition.” *The Annual Review of Economics.*

Patel, R.S.,Bachu, R., Adikey, A., Malik, M., and Shah, M. (2018). Factors Related to Physician Burnout and Its Consequences: A Review. *Behav. Sci.*

Reyes, Jessica Wolpaw. 2007. “Gender Gaps in Income and Productivity of Obstetricians and Gynecologists.” *Obstetrics & Gynecology*

Sarsons. 2019. “Interpreting Signals in the Labor Market: Evidence from Medical Referrals.” *Working Paper.*

Sasser, Alicia C. 2005. “Gender Differences in Physician Pay: Tradeoffs between Career and Family.” *The Journal of Human Resources*

Thomas-Hunt and Phillips. 2004. “When What You Know Is Not Enough: Expertise and Gender Dynamics in Task Groups.” *Personality and Social Psychology Bulletin.*