Do Female Researchers Increase Female Enrollment in Clinical Trials?*

Harsh Gupta†

December 28, 2022

Abstract

Low female participation in clinical trials is a persistent pattern, potentially resulting in adverse outcomes for female patients. I investigate whether including more female researchers increases female participation. I estimate, using a linear model, that a 1-standard-deviation increase in the share of female researchers is associated with a 0.1-standard-deviation increase in the share of female participants conditional on disease and year fixed effects. Equivalently, moving from a trial with no female PIs to a trial with only female PIs is associated with a 5 percentage point increase in female participation. Key mechanisms include designing inclusive trial criteria, hiring female facility staff, and targeted advertising.

Keywords: Innovation and Inequality; Clinical Trials; Health Innovation

*I am grateful to Matthew Gentzkow and Heidi Williams for their invaluable guidance and advice. I thank Leah Backhus, Lea Bottmer, Dante Domenella, Maya Durvasula, Liran Einav, Nick Grasley, Adam Rosenberg, Helena Roy, Alessandra Voena, and seminar participants at Stanford University for their comments and suggestions. The paper greatly benefited from discussion with Kenneth Mahaffey, Lee James, and Lorena Kuri who provided crucial insights into the trial recruitment process and the role of investigators. Any and all errors are my own.

†Department of Economics, Stanford University. Landau Economics Building, 579 Jane Stanford Way, Stanford, CA 94305. Email: hgupta13@stanford.edu.

1 Introduction

Ambien, a popular drug used to treat sleep disorders, was approved in 1992 by the FDA on the basis of a trial in which only 37% of participants were female (FDA (1992)). In 2013, the FDA cut the recommended dosage of Ambien for female patients in half, after confirming that the dosage was too high for women because they metabolize the drug more slowly than men (FDA (2013)). In the 20 years between Ambien's approval and the 2013 FDA guidance, millions of female patients received an improper dosage of Ambien; this improper dosage has been linked to more than 700 traffic accidents (FDA (2013)).

The Ambien case study illustrates a frequently discussed problem in biomedical innovation. Female participation is low in clinical trials[1], despite strong evidence that drug responses frequently differ by sex[2]. Because of inadequate participation in clinical trials, it is difficult to learn sex differences for novel drugs. If these differences are large it can then result in sub-optimal treatment and adverse health consequences for female patients[3]. Recently, market participants and policymakers like the National Institutes of Health (NIH), Columbia University[4], and Bristol-Myers-Squibb (BMS)[5] have considered engaging more female researchers in clinical trials as a potential solution to this issue. However, there is a lack of evidence on whether more female researchers increases female participation.

In this paper, I investigate whether including more female principal investigators (PIs) is an effective policy for increasing female participation in clinical trials. To answer this question, I utilize comprehensive data on the near universe of clinical trials registered on ClinicalTrials.gov and clinical trials published in biomedical scientific articles. My main result, presented in Figure 1, shows that increasing the share of female PIs from 0 to 1 is associated with a 5 percentage point increase in the share of female participants. Equivalently, a 1 standard deviation (sd) increase in the share of female PIs is associated with an increase of approximately 0.1 sd in the shares of female participants. I use disease and year fixed effects as controls in my analysis to account for recent policy changes and disease selection; this result provides evidence that female PIs increase female participation *within* granular disease categories for a given calendar year. My main result is robust to different methods of controlling for disease selection, to including different trial features as controls, to different methods of predicting PI sex, and across different sample restrictions. I also show robustness to unobserved confounders using the test developed in Oster (2019).

[1] See Nadel (1992); Mastroianni et al. (1994); Ramasubbu et al. (2001); Feldman et al. (2019)
[2] See Office (2001); Yu et al. (2016); Zucker and Prendergast (2020).
[3] See Zopf et al. (2008); Baiu et al. (2021).
[4] For example, the Columbia University and Pfizer Diveristy Initiative.
[5] BMS has committed to train and develop 250 new racially and ethnically diverse clinical investigators that can enroll a diverse patient population in trials conducted across the industry.

I investigate several mechanisms driving the observed positive relationship between female researchers and female participation through a combination The goal is to understand what female PIs do differently from male PIs for a trial within a given disease-year that enables them to recruit female participants more effectively. Previous studies have shown that sex (and racial) concordance between potential patients and staff improves take-up of medical care services (Alsan et al. (2019)). In clinical trials specifically, the relationship between study staff and participants is considered to be critical (CISCRP (2019a)). Female PIs can increase enrollment of female patients because of sex concordance between PIs and participants. Furthermore, female PIs may indirectly improve female enrollment by hiring more female staff: I find that a 1 sd increases in the share of female PIs is associated with approximately a 0.5 sd increase in the share of female staff. Thus, female PIs may also increase female participation by hiring female staff who may themselves better recruit female participants.

Additionally, PIs determine who can participate in clinical trials by designing the exclusion and inclusion criteria. I find that female PIs are more likely to design trials exclusively for females and are less likely to design trials exclusively for males within a given disease-year. Such criteria mechanically increase female participation. Additionally, I find that female PIs are less likely to "reflexively" exclude pregnant women from clinical trials. These results suggest that female PIs are less likely to include criteria that restrict access to female patients for a given disease-year.

This paper contributes to the nascent literature in the economics of innovation and inequality. This literature studies inequality - across dimensions like race, sex, and income - in the innovation process and benefits from innovation. Jaravel (2019) showed that product innovation was more beneficial to high income individuals; Cutler et al. (2012) argued that induced innovation in medical technology for treating at-risk infants contributed about one third of the rise in the black-white infant mortality ratio between 1983 and 1998. This paper focuses on inequalities by sex in the novel and important setting of clinical trial participation. Recently, Michelman and Msall (2022) studied how restrictions on female participation in trials affect sex-specific pharmaceutical innovation.

Additionally, the policy studied in the present paper - the inclusion of more female researchers - is related to the literature which examines how inventor identity (e.g., sex) affects the innovation process and outcomes. Koning et al. (2020) and Koning et al. (2021) find that female inventors are more likely to innovate in areas that serve women's needs. Nielsen et al. (2017) argued that female authors are more attentive to sex analysis in academic papers.

The paper is organized as follows. In Section 2, I provide a brief background on clinical

trials, discuss my data sources, and present baseline descriptive facts. Section 3 explores the central question of this paper and Section 4 discusses possible mechanisms. I conclude in Section 5.

2 Background and Data

2.1 Background on Clinical Trials

Clinical trials[6] are prospective studies comparing the effects and values of novel interventions against a control in human beings (Friedman et al. (2015)). Clinical trials are central to biomedical innovation from both a scientific and regulatory perspective. Clinical trials are the "gold standard" to evaluate a drug (Hariton and Locascio (2018)) and are required for FDA market approval (Junod (2008)). Consequently, a tremendous amount of resources are invested in clinical trials: approximately $30,000$ trials are conducted per year and the median cost per trial is approximately $35 million (Moore et al. (2018)).

Patients volunteer in clinical trials for a variety of reasons — to get access to novel treatments or higher quality of care, to help provide a better scientific knowledge base for future generations, or to receive financial compensation (CISCRP (2019b)). Trial participants bear risk because of possible side effects, receiving a placebo during the trial, and other health risks like adverse interactions with their current medications (CISCRP (2019b)). The decision to participate in clinical trials is thought to be influenced by medical, economic, social, and ethical factors. See NIH (2015) for a detailed discussion.

In investigating the role of female researchers, I focus on PIs - researchers who lead a clinical trials and are responsible for all its activities (Kleppinger (2012)). I focus on PIs because make key decisions about trial design that might potentially affect female enrollment. PIs design and execute the trial protocol; PIs hire and supervise trial staff; and PIs also analyze data and report the results of the trial.

Female Participation in Clinical Trials This issue of low female participation in clinical trials was highlighted in landmark reports by Congress in 1992 (Nadel (1992)) and Institute of Medicine in 1994 (Mastroianni et al. (1994)). These reports highlighted that there was low female participation in both early and late stage trials. As highlighted in the Congressional report (Nadel (1992)), the low participation of female patients implied that many trials did not have the minimum number of female participants needed to detect differences in response

[6]For a more detailed description of clinical trials, see ClinicalTrials.gov (2019).

to novel treatments. Even when there were a sufficient number of female participants, most trials did not report differences by sex.

Several policies have been introduced to increase female participation and female subgroup analysis in clinical trials. For example, the NIH Revitalization Act of 1993 required that "women are included as subjects in each project of such research" unless there is medical reasoning for exclusion. However, these policies have had limited success in increasing female participation in clinical trials. Ramasubbu et al. (2001) performed a systematic review of all clinical trials published between 1994 and 1999 in the New England Journal of Medicine and found no improvement in sex-balance in enrollment. Subsequently, several studies have documented underrepresentation - a lower share of female enrollment relative to female disease incidence - in a multitude of diseases categories: cardiology (Geller et al. (2011), Harris and Douglas (2000)), vascular surgery (Hoel et al. (2009)), emergency medicine (Safdar et al. (2011)), hand surgery (Kalliainen et al. (2018)), and oncology (Murthy et al. (2004), Baiu et al. (2021)).

A recent comprehensive study by Feldman et al. (2019) found that even though representation has improved over time, females are still underrepresented in clinical trials in most major diseases categories. I present further descriptive evidence that documents female underrepresentation in Appendix B. Historical underrepresentation of females is supported by a preponderance of evidence from the existing literature and continues to be a problem even today.

Why is Female Participation Important? Increased female participation would offer limited benefits if drug responses were identical across sex. However, previous studies have documented large differences in drug responses (Yu et al. (2016); Zucker and Prendergast (2020); Nadel (1992)). If there is insufficient female participation in trials, then these differences are difficult to detect and female patients will likely be inappropriately treated (as in the Ambien example). Ambien is not an outlier; for 307 out of 668 commonly prescribed drugs in the US the rate of adverse events is higher for female patients (Yu et al. (2016)).

Further, if evidence on drug response in female patients is not well-understood, high-value treatments may be under-prescribed. Statin prescribing rates illustrate this concern: early trials failed to provide conclusive evidence and, as a result, were considered "underutilized" for female patients (Plakogiannis and Arif (2016)). Moreover, persistent exclusion of female patients may erode trust in medical institutions.

Although suggestive, these studies report various harms associated with low female participation in trials and indicate that increased participation may improve health outcomes.

2.2 Data

2.2.1 Data Sources

AACT For each clinical trial, the key variables are the sex of the Principal Investigator(s) of the trial and the number of male and female participants in the trial, which I obtain from two sources. First, I use the Aggregate Content of ClinicalTrials.Gov dataset (AACT), which is a public dataset that provides information on clinical trials registered on ClinicalTrials.Gov. ClinicalTrials.Gov is the largest clinical trial registry in the world, with over 300,000 clinical trials currently registered. The FDA Administration Amendments Act of 2007 (FDAAA) requires that all trials conducted in the US be registered on ClinicalTrials.Gov and all trials relating to FDA approved drugs provide aggregate outcome data on ClinicalTrials.Gov (FDAAA of 2007). The AACT data provides information on the number of male and female participants, the name(s) of the Principal Investigator(s), the Medical Subject Heading (MeSH)[7] terms associated with the study, and the start and end date of the trial. Despite regulations mandating the registration of clinical trials on ClinicalTrials.Gov, several studies have documented that compliance is low (Geller et al. (2006); Law et al. (2011); Zarin et al. (2017)). I only study trials which were started after the passage of FDAAA in 2007 to limit any selection bias.

PubMed I supplement AACT with data on clinical trials from published biomedical papers in the PubMed dataset. PubMed is a large repository of more than 33 million life sciences articles maintained by the US National Library of Medicine. The PubMed data provides information on names of the authors, MeSH terms, and the year of publication of each scientific paper.

The PubMed data itself does not have information on trial participation by sex. To overcome this limitation, I leverage the PubMed Extract dataset[8] developed in Feldman et al. (2019). Feldman et al. (2019) used a novel extraction algorithm to collect clinical trial data from over 25,000 academic articles published on PubMed between 1996 and 2018. Conceptually, the algorithm scans digital copies of each article, identifies the table with the demographics of trial participants, and then extracts the information from that table into a dataset.

Note that any errors in the PubMed Extract algorithm would result in measurement error in my analysis. To alleviate concerns, Feldman et al. (2019) performed a number of validation exercises, including comparing their measurements with those collected manually

[7] MeSH terms are a controlled and hierarchically-organized vocabulary of biomedical information produced by the National Library of Medicine (NLM).

[8] See PubMed Extract repository for further details on their algorithm and data construction process.

on a random subsample of 100 papers. They found that the mean absolute error in the share of female participants between their algorithmic measurements and manual measurements was 0.008. Additional validation exercises are discussed in detail in Feldman et al. (2019). Their suite of validation exercises broadly suggest that their measurements are accurate. Moreover, there were no systematic differences in the algorithm for male and female participation measurements. Consequently, if any measurement errors are present, they would result in attenuation bias which would lead to conservative estimates in my analysis.

My main sample comprises trials from both AACT and PubMed. I also examine each dataset separately as a robustness exercise in Appendix E.2.

Supplementary Data Sources In addition to the PIs of the trial, the number of male and female participants, the MeSH terms of the trial, and the year of the trial I collect additional information to serve as controls, explore heterogeneity, and/or study mechanisms. I obtain information on the phase of the trial, the location of trial sites, the staff employed at trial sites[9], and the exclusion and inclusion criteria of trials using the AACT dataset.

I use the Unified Medical Language System (UMLS) from the NLM to map MeSH terms and International Classification of Diseases (ICD) - 9 codes. I obtain data on NIH funding associated with clinical trials from the NIH ExPORTER dataset. Data on disease prevalence was obtained from the Center of Disease Control WONDER dataset.

See Appendix A for additional details on data construction.

2.2.2 Predicting PI Sex

The AACT and PubMed datasets do not provide information on the sex of the PIs. To overcome this limitation, I use the first name of the PIs to impute their sex, applying the method developed by Blevins and Mullen (2015).

Briefly, the method uses historical and administrative data to measure the likelihood that a name was associated with a particular sex based on the year and geographical location of birth. The paper uses the Social Security Administration data, IPUMS-USA dataset, and the North Atlantic Population Project to compute the share of female children that had a given name in a given year.

The sex of a female PI is the share of children with their name who were born female. I interpret the sex of a PI as the probability that they are female given their name. In Appendix E.3, I show robustness to alternate definitions: I show that my results are robust

[9]ClinicalTrials.Gov removes information of trial staff once a trial has completed recruitment. Consequently, information on trial staff is limited to trials that were actively recruiting as of Dec 10th, 2021.

to classifying PIs as female if more than 50% or more than 75% of all children born with their name are female.

Share of Female Participants and PIs The primary outcome is the share of female participants, which is the number of female participants divided by the total number of participants.

Similarly, for a given trial the share of female PIs is the mean of the probability that a PI is female across all PIs of that trial. For example, the clinical trial NCT04748341 has four PIs: Mishal, Muhammad, Muhammad, and Amir. The probabilities that the PIs are female are $0.97, 0, 0,$ and 0.54, respectively. The share of female PIs for the trial is then 0.377. Note that I use author and PI interchangeably.

2.2.3 Trial Disease

The disease studied in a trial is determined by three-digit ICD9 codes associated with the trial. ICD9 is the official system of assigning codes to diagnoses and procedures associated with hospital utilization in the United States.

The AACT and PubMed data provide information on MeSH terms associated with the trials. I map MeSH terms to ICD9 using a procedure described in Appendix A.1.

Many trials have multiple ICD9 codes associated with them. For trials with more than 3 ICD9 codes, I only include the three most common ICD9 codes among trials in my sample. Trials with multiple ICD9 codes are treated as separate observations. The intuitive reason is that if a drug is used for multiple diseases, then the trial design should reflect all the diseases. Therefore, a unit of observation in my analysis is a trial-disease combination[10]. However, for simpler exposition I refer to a trial-disease pair as a trial.

2.3 Descriptive Statistics

My main sample consists of 56,915 trials. The average trial has more male PIs than female PIs and has more male participants than female participants: the mean of the share of female PIs is 0.32 and the mean of the share of female participants is 0.46. The mean female underrepresentation[11] is -0.042 and females are underrepresented in majority of clinical trials[12].

[10]Note that for each trial the share of female PIs and share of female participants is identical across diseases.

[11]Recall, female underrepresentation is the difference between the share of female participants and female disease incidence.

[12]See Appendix D.1 for details.

In Appendix C, I show that share of female participants has increased from 0.4 in 2003 to 0.5 in 2018 and that the share of female PIs has steadily increased from 0.27 in 2005 to 0.37 in 2018. These figures highlight that female researchers and female participants are playing an increasingly important role in biomedical innovation.

3 The Effect of Female PIs on Female Participation

I present the relationship between female PIs and female participants graphically in Figure 1. The figure plots the share of female PIs and the share of female participants using a bin scatter after accounting for disease × year fixed effects. Controlling for diseases is important because certain diseases (such as breast cancer) attract more female PIs and more female participants. Controlling for year fixed effects allows us to prevent confounding due to recent policies that might have improved female involvement in medicine as researchers and as participants. Additionally, different diseases may have different trends over time. The disease × year fixed effects control for such confounders. Consequently, one can interpret this figure as the association between share of female PIs and share of female participants within granular diseases for a given calendar year.

In the figure, I observe a strong positive relationship between share of female PIs and share of female participants. This suggests that a higher share of female PIs in a given study increases the share of female participants. To quantify the statistical association between female PIs and participants, I estimate the following linear model:

$$Y_{idt} = \beta F_{idt} + \gamma X_{idt} + \epsilon_{idt}\,, \tag{3.1}$$

where for trial i studying disease d in year t, Y_{idt} denotes the share of female participants and F_{idt} denotes the share of female PIs. The controls are denoted by X_{idt}. My preferred specification includes disease × year fixed effects γ_{dt} and the total participants of the trial as controls. The standard errors are clustered at the disease level. The parameter of interest is β. If $\beta > 0$ then female PIs are associated with a higher share of female participants. I also examine (in separate regressions) the effect of having at least one Female PI on female participants[13].

The results are presented in Table 1. In column (1), I observe that $\beta = 0.08$ when we include no controls, which suggests that there is a positive correlation between the share of female PIs and share of female participants. However, as discussed above, the disease that the trial studies is a crucial confound. In column (2), I include granular disease × year

[13]Here F_{idt} is an indicator for whether the trial has at least one female PI

fixed effects as controls. As expected, the magnitude of the estimate decreases to $\beta = 0.049$. Importantly, even with these granular disease $\times$ year fixed effects, the parameter estimates imply a positive and statistically significant relationship between the share of female PIs and the share of female participants.

In column (3), I examine the effect of having at least one female PI involved in a trial. On average, a trial with at least one female PI has a 0.012 higher share of female participants than trials with no female PI after controlling for disease $\times$ year fixed effects. This suggests that having even one female PI can increase the share of female participants in an economically and statistically significant way.

Further, I show that the share of female PIs is significantly associated with the number of female participants (column (4)) but not the total number of participants (column (5)), suggesting that the increase in the share of female participants is driven by substitution of male participants with female participants, and not at the expense of smaller trials due to exclusion of male participants.

Magnitude To better interpret the magnitude of my results, note that the standard deviation of the share of female PIs and share of female participants is 0.32 and 0.22, respectively. An estimate of $\beta = 0.049$ then implies that a 1 sd increase in the share of female PIs results in an approximately 0.1 sd increase in share of female participants.

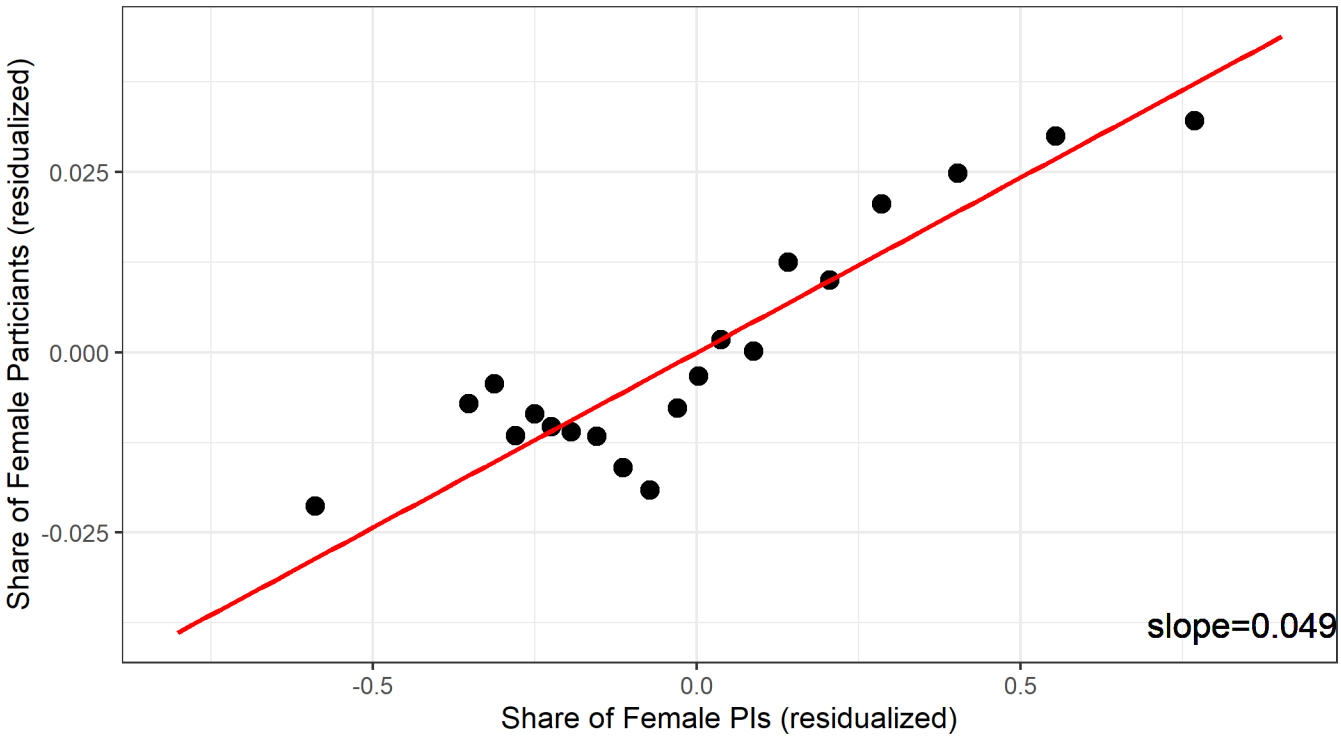
I interpret my estimates in the context of a specific example: trial NCT01510652 which studied cardiac resynchronization therapy (CRT). The share of female participants in this trial was 0.23, which is close to the 25^{th} percentile of the distribution of female participants in cardiovascular trials. All the PIs in the trial were male. If all the PIs were counterfactually female and my estimates are causal, then the share of female participants would be 0.28 which is close to median of the distribution of female participants. That is, counterfactually changing the share of female PIs from 0 to 1 for NCT01510652 would increase the share of female participants from the 25^{th} percentile to the median.

To further contextualize the magnitude of my estimates, I use a back-of-the-envelope exercise. Section 907 of the 2012 Food and Drug Administration Safety and Innovation Act (FDASIA), requires the reporting and analysis of demographic subgroups in applications for drugs, biologics and devices by sex, race and ethnicity, and age (Congress (2012)). The increased transparency mandated by FDASIA increased the scrutiny of sex data in clinical trials by regulators and consumers. Consequently, trial sponsors were incentivized to include more female participants. The FDASIA is often considered to be an important legislative milestone in ensuring sex parity in clinical trials. I compare the observed change in female participation after the FDASIA to my parameter estimates. Trials after the passage of the

FDASIA in 2012 had a 0.009 higher share of female participants than trials conducted before the FDASIA[14]. In comparison, a 1 sd increase in the share of female PIs increases the share of female participants by 0.015 which is almost twice as large as the observed increase after the FDASIA.

Broadly, these results suggest that the association between share of female PIs and female participants is positive and statistically significant, as well as economically significant.

Figure 1: Female Participants and Female PIs



The figure shows the relationship between share of female PIs and share of female participants using a bin scatter plot. Each observation is a clinical trial. The $x-$axis shows the share of female PIs in a given study, and the $y-$axis shows the share of female participants. The values are the residuals of a regression of each variable on disease $\times$ year fixed effects. The red line is a linear best fit line. The bins are chosen such that each bin has the same number of observations. The slope of the fitted line is annotated in the bottom-right corner.

[14]Clearly, these are not causal estimates because trials after the FDASIA might be systematically different from trials before the FDASIA for other reasons.

Table 1: OLS Regression of Female Participation on Female PIs

	Dependent variable:				
	Female Participants			Female Participants	Total Participants
	Share	Share	Share	Number	
	(1)	(2)	(3)	(4)	(5)
Share of Female PIs	0.083	0.049		404.351	3,881.927
	(0.003)	(0.005)		(212.597)	(2,861.600)
Any Female PI			0.012		
			(0.003)		
Total Participants	0.000	0.000	0.000	0.502	
	(0.000)	(0.000)	(0.000)	(0.007)	
Disease x Year F.E.	No	Yes	Yes	Yes	Yes
Observations	56,895	56,895	56,895	56,895	56,895
R^2	0.015	0.310	0.306	0.995	0.003
Adjusted R^2	0.015	0.260	0.255	0.995	−0.003

The table presents the results of an OLS regression of the share of female participants on the share of female PIs (Columns (1) - (3)) and any female PI (Column (4)), as specified in Equation 3.1. In Columns (5) and (6) the outcome is the number of female participants and the total number of participants, respectively. Column (1) includes no controls, Columns (2)-(5) include disease × year fixed effects. SE are clustered at the disease level.

Robustness In Appendix E.1, I show that my results are robust to including even finer disease codes (5-digit ICD9 codes) as controls, restricting to diseases that do not predominantly affect one sex, restricting to a sample matched on disease, year and trial size, and adjusting for disease burden.

Further, to assuage concerns about omitted variable bias due to unobservables, I show robustness using the tests developed in Oster (2019). Under the assumption that unobservable confounders can at most double the R^2, I find that unobservable confounders would need to be at least as important as observable controls for the true association between female PIs and female participants to be null[15]. In other words, for my results to be completely driven by omitted-variable bias there must exist confounders that can both double the predictive power of the linear model and are at least as strongly associated with the share of female

[15]In the notation of Oster (2019), if we assume that $R^2_{max} = 2\widetilde{R^2} = 0.62$ then a $\delta^* > 1.1$ is necessary for $\beta = 0$.

PIs in a trial as the disease the trial is studying and the year of the trial.

The two datasets I use may suffer from data limitations (as discussed in Section 2.2). In Appendix E.2, I show that the estimates are similar when estimated on each dataset separately, suggesting that potential data limitations are not biasing my results in an economically meaningful way. Appendix E.3 shows that my estimates are identical across different methods for classifying PI sex. Last, Appendix E.4 shows robustness of my estimates to the inclusion of additional trial features as controls.

Role of Selection into Disease The underlying thought experiment in my analysis is that the set of trials is fixed and we are counterfactually increasing female PIs for each trial. However, as discussed in Appendix D.1, female PIs are more likely to do research in diseases with a high female disease burden and such diseases are more likely to have a higher share of female participants (Appendix Figure 7). Therefore, female PIs may also increase female participation by selecting into trials for diseases which would have a higher share of female participants *independent* of the share of female PIs in that trial. My analysis estimates the increase in female participation due to an increase in female PIs after controlling for this selection by employing disease×year fixed effects. However, the effect of an overall increase in female PIs on female participation in clinical trials may be higher than I what I estimate because of this potential selection (see Appendix D).

4 Mechanisms

In this section, I investigate what female PIs do differently that enables them to recruit more female patients through empirical and qualitative analysis. The qualitative analysis is based on the existing literature, discussions with clinical researchers at Stanford University, Veterans Health Administration, and Bristol-Meyers-Squibb (BMS) as well as manual reviews of clinical trial protocols.

Recruitment in clinical trials typically involves several key steps: designing the trial protocol, hiring facility staff, advertising, and discussing the trial with volunteers (Chaudhari et al. (2020)). PIs play an important role in each of these steps and each of these steps can influence female participation in clinical trials.

4.1 Trial Staff

Trial staff play an important role in patient recruitment. Communication with patients during the first few visits is critical because patients are often concerned that they will

receive a placebo or suffer from side-effects (CISCRP (2019b)). Clearly explaining different aspects of the trial and appropriately addressing the concerns of participants can be crucial for recruitment. Special attention is often needed for minority patients who are unfamiliar with the medical system or may even distrust medical institutions (NIH (2015)).

A large literature has documented that racial and gender concordance increase participation in medical services and clinical trials. For example, Alsan et al. (2019) showed that racial concordance among African-American patients increased their take-up of preventative services. A survey by the Center for Information and Study on Clinical Research Participation (CISCRP) showed that relationships with and medical attention from study staff was one of the most valued benefits from participating in clinical trials (CISCRP (2019b)). If female staff are more familiar with the needs of female patients and are able to develop better relationships with them, then it might boost female participation.

The existing literature provides several examples that support the idea that sex and/or racial concordance between participants and study staff improves engagement with clinical trials. Maxwell et al. (2005) found that personal invitations from a female project liaison improved female participation in a cancer-screening study. It is worth noting that the information was the same across different outreach channels. Nonetheless, those female participants who were contacted by a female project liaison were much more likely to enroll, independent of the information provided.

Similarly, Aroian et al. (2006) compared different strategies for improving recruitment of Arab Muslim mothers and children for research. They found that hiring data collectors and researchers who are female and speak Arabic was "the single most successful recruitment and retention strategy". One key feature of this example is that Muslim women are often not comfortable talking to men outside their immediate family alone (Aroian et al. (2006)). In the absence of female study staff, it is very unlikely that these women would participate in the trial.

Consequently, female PIs can increase female enrollment when they themselves are recruiting patients. Moreover, female PIs can indirectly improve female enrollment by hiring more female study staff. The latter channel is likely to be particularly important for large trials because PIs have limited capacity to directly recruit patients when trials have many patients and occur in multiple sites.

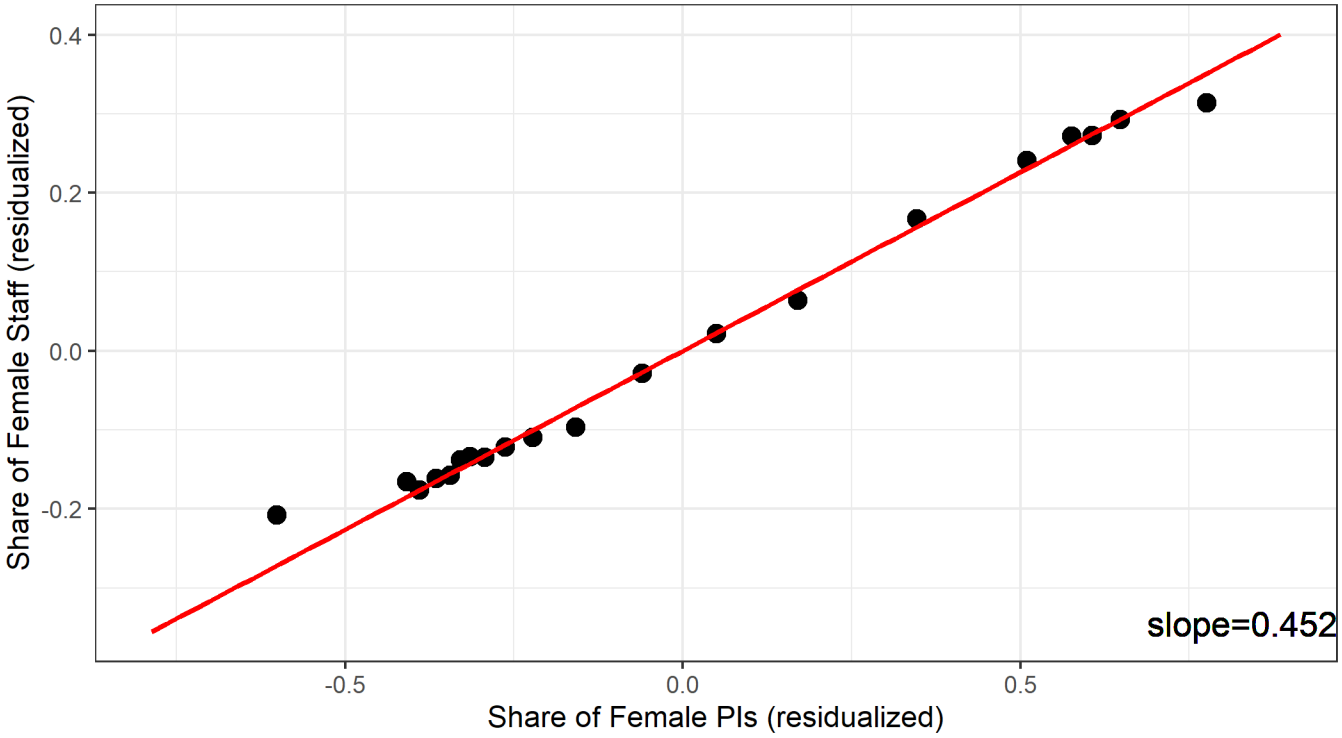
I examine whether a higher share of female PIs increases the share of female staff in clinical trials. Staff here refers to those personnel who were listed as contacts for each site of the trial. I predict the sex of trial staff using the same method I used to predict the sex of PIs (see Section 2.2.2).

Figure 2 presents a bin scatter plot of female PIs and female staff after adjusting for

disease × year fixed effects. Trials with a higher share of female PIs are much more likely to have a higher share of female staff. The slope of the linear relationship between female PIs and female share is 0.45. Equivalently, a 1 sd increases in the share of female PIs is associated with approximately a 0.5 sd increase in the share of female staff.

In summary, an increase in the share of female PIs is associated with an increase in the share of female staff. The existing literature suggest that a higher share of female staff can potentially increase female participation.

Figure 2: Female Staff and Female PIs



The figure shows the relationship between share of female PIs and share of female staff using a bin scatter plot. Each observation is a clinical trial. The $x-$axis shows the share of female PIs in a given study, and the $y-$axis shows the share of female staff. The values are the residuals of a regression of each variable on disease × year fixed effects. The red line is a linear best fit line. The bins are chosen such that each bin has the same number of observations. The slope of the fitted line is annotated in the bottom-right corner.

4.2 Trial Eligibility

PIs determine which volunteers are eligible for participation by setting the inclusion and exclusion criteria. Historically, female participants were excluded from most early stage trials (FDA (1977)) and many trials continue to exclude all female participants. Conversely,

trials may also include *only* female participants to test the safety and efficacy of a drug specifically for the female population.

In Figure 3, I present the share of trials that are female only and male only as a linear function of the share of female PIs after adjusting for disease × year fixed effects. Trials with a higher share of female PIs are less likely to exclude females for a given disease in a given year. Moreover, trials with a higher of a share of female PIs are more likely to be exclusively for female patients. The results suggest that female PIs may increase female participation by ensuring that they are not excluded outright from the trial and by designing trials that are focused exclusively on female patients.

In addition to the outright exclusion of female patients, certain other criteria can disproportionately affect female participation. For example, the symptoms and diagnosis of certain diseases are differ by sex. In the case of Acute Myocardial Infarction (AMI), men are more like to suffer from chest pain while women are more likely to suffer from fatigue and nausea (King and McGuire (2007)). If the inclusion criteria require chest pains as a symptom, then this would hinder female participation in clinical trials.

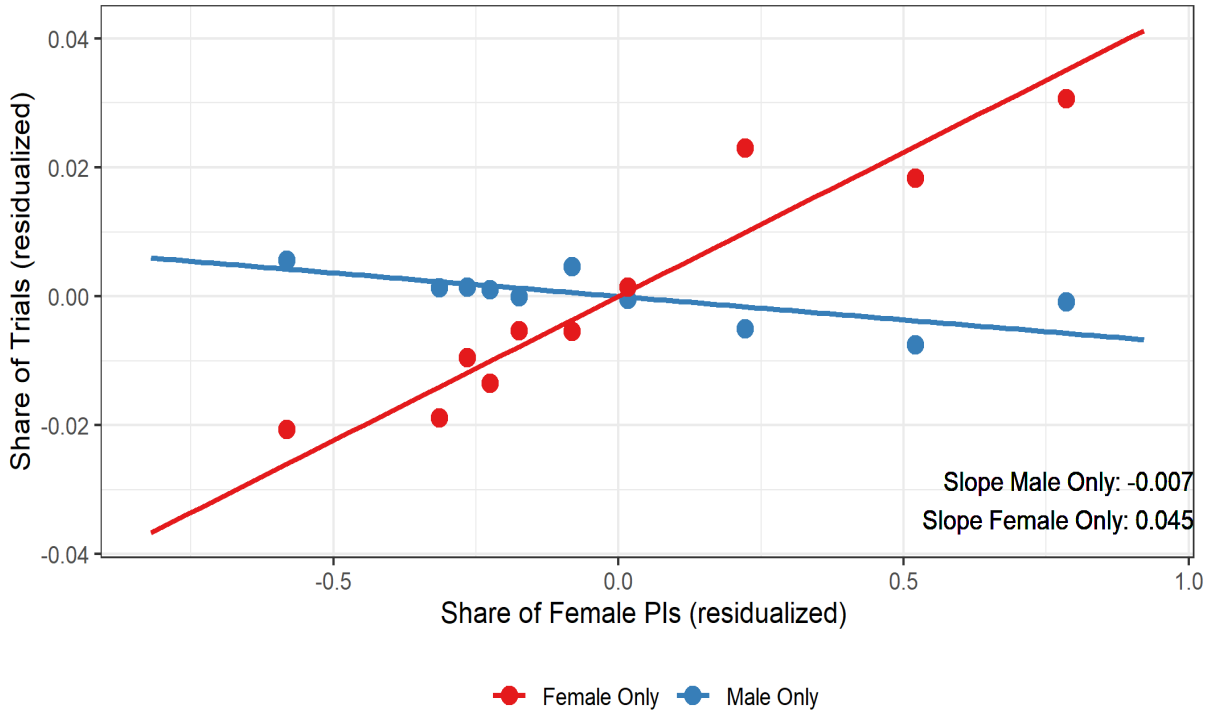
Moreover, certain exclusion criteria are only applicable to female patients. Perhaps the most ubiquitous example is that of pregnancy. Trials often exclude pregnant women: more than 50% of trials in my sample excluded pregnant participants. Although there are medical and ethical justifications for the exclusion of pregnant patients, there is often "reflexive exclusion" of pregnant patients from trials (Shields and Lyerly (2013)). That is, pregnant patients are excluded outright as a rule without efforts to thoughtfully design criteria that would enable them to safely participate in the trial.

I investigate whether female PIs are less likely to exclude pregnant patients in Figure 3b. The figure shows that female PIs are less likely to exclude pregnant patients for a given disease in a given year, potentially mitigating the "reflexive exclusion" of pregnant patients. However, the decision of whether to include pregnant patients is complex and further research is necessary to evaluate whether the inclusion of more pregnant patients is welfare improving.

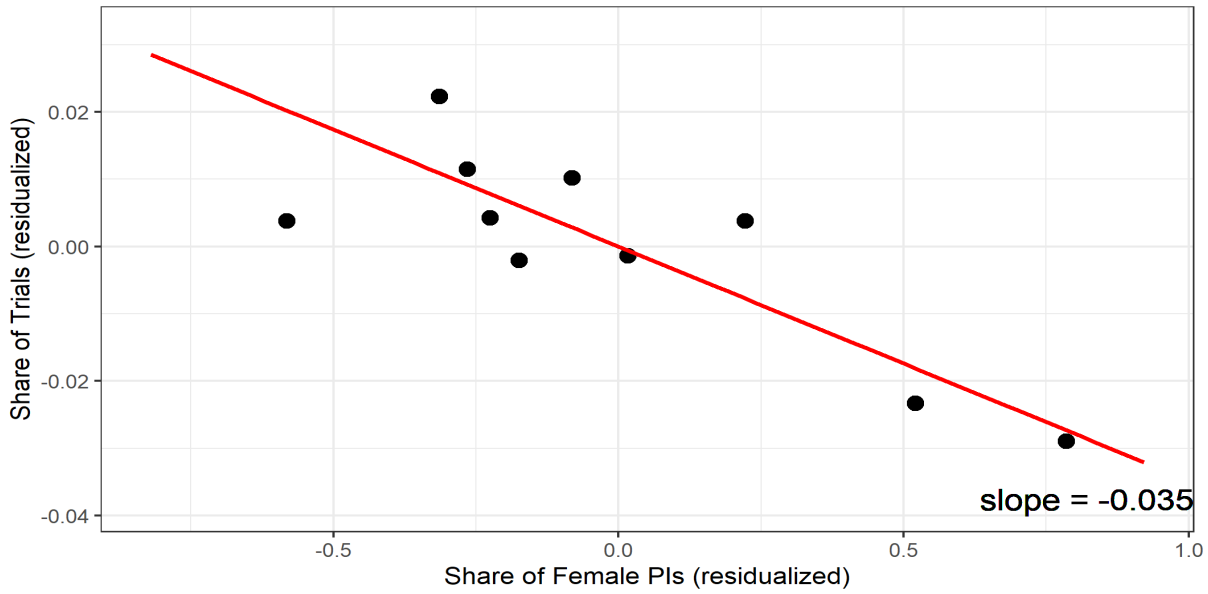
Broadly, female PIs are less likely to exclude female patients from trials and more likely to run female-only trials. Female PIs ensure that criteria that disproportionately exclude women are included only if medically necessary. These actions then increase female participation in clinical trials.

Figure 3: Inclusion and Exclusion Criteria

(a) Sex Specific Criteria



(b) Exclusion of Pregnant Patients



The figures shows the relationship between the share of female PIs and trial criteria using a bin-scatter plot. Each observation is a clinical trial. The $x-$axis shows the share of female PIs in a given study. In the top panel, the y- axis shows the share of trials that include only female patients (in red) and the share of trials that include male patients (in blue). In the bottom panel, the $y-$axis shows the share of trials that exclude pregnant participants. The values are the residuals of a regression of each variable on disease $\times$ year fixed effects.The bins are chosen such that each bin has the same number of observations. The slope of the fitted line is annotated in the bottom-right corner.

4.3 Additional Potential Mechanisms

The pool of potential female participants is influenced by trial features like location of trial sites, network of referring physicians, research funding, advertising, and communication methods. Female PIs are believed to be better aware of the needs of female participants and can design trials so that participation is more appealing to females. I discuss these mechanisms in Appendix F.

5 Conclusion

In this paper, I investigate whether increasing the number of female PIs would increase female participation in clinical trials. I find that a 1 sd increase in share of female PIs increases the share of participants by 0.1 sd within a given disease for a calendar year, highlighting that more female researchers are likely to boost female participation. Additionally, my estimates may be a lower bound because they do not account for the fact that female PIs select into diseases with a high female disease burden. Female PIs appear to increase female participation by hiring more female staff, designing trial criteria that do not superfluously exclude female patients, obtaining more research funding, and via targeted advertising.

As remarked in JAMA Internal Medicine, "Enrollment of women in clinical trials of heart failure and acute coronary syndrome still lag behind disease prevalence in the general population, which likely reflects the paucity of female leadership in these trials" (Wang and DesJardin (2020)). A back-of-the-envelope calculation based on my estimates suggests that an increase in the share of female PIs by 0.5 would be to sufficient to close the gap between enrollment and disease prevalence for cardiovascular diseases[16]. My results are applicable beyond cardiovascular diseases, supporting the idea that increasing female leadership can help increase female enrollment for many diseases. Increased female enrollment will facilitate a better understanding of the safety and efficacy of novel drugs for female patients, potentially improving treatment decisions and health outcomes.

[16] In my sample, I find the mean underrepresentation in trials studying blood diseases is approximately 0.03. Equivalently, an increase in the share of female participants by 0.03 would close the gap between female enrollment and female disease burden. Assuming my estimates are causal, an increase in share of female PIs by 0.5 would increase share of female participants by 0.03.

References

Alsan, Marcella, Owen Garrick, and Grant Graziani, "Does Diversity Matter for Health? Experimental Evidence from Oakland," *American Economic Review*, 2019, *109* (12), 4071–4111.

Angrist, Joshua D and Jörn-Steffen Pischke, *Mostly Harmless Econometrics*, Princeton University Press, 2008.

Aroian, Karen J, Anne Katz, and Anahid Kulwicki, "Recruiting and Retaining Arab Muslim Mothers and Children for Research," *Journal of Nursing Scholarship*, 2006, *38* (3), 255–261.

Baiu, Ioana, Ashley L Titan, Linda W Martin, Andrea Wolf, and Leah Backhus, "The Role of Gender in Non-Small Cell Lung Cancer: a Narrative Review," *Journal of Thoracic Disease*, 2021, *13* (6), 3816.

Blevins, Cameron and Lincoln Mullen, "Jane, John... Leslie? A Historical Method for Algorithmic Gender Prediction.," *Digital Humanities Quarterly*, 2015, *9* (3).

Boca, Daniela Del, Noemi Oggero, Paola Profeta, and Mariacristina Rossi, "Women's and Men's Work, Housework and Childcare, Before and During COVID-19," *Review of Economics of the Household*, 2020, *18* (4), 1001–1017.

Bohensky, Megan A, Damien Jolley, Vijaya Sundararajan, Sue Evans, David V Pilcher, Ian Scott, and Caroline A Brand, "Data Linkage: a Powerful Research Tool with Potential Problems," *BMC health services research*, 2010, *10* (1), 1–7.

Chaudhari, Nayan, Renju Ravi, Nithya J Gogtay, and Urmila M Thatte, "Recruitment and Retention of the Participants in Clinical Trials: Challenges and Solutions," *Perspectives in Clinical Research*, 2020, *11* (2), 64.

CISCRP, "2019 Perceptions and Insights Study," 2019, *https://www.ciscrp.org/wp-content/uploads/2019/12/General-Perceptions-and-Awareness-Report-04DEC.pdf*.

_ , "Survey of Those of Participated in Clinical Trials," 2019, *https://www.ciscrp.org/education-center/charts-statistics/*.

ClinicalTrials.gov, "Learn About Clinical Studies," *National Library of Medicine*, 2019, *https://clinicaltrials.gov/ct2/about-studies/learn*.

Congress, "Food and Drug Administration Safety and Innovation Act (FDASIA)," *Public Law*, 2012, *"https://www.congress.gov/112/plaws/publ144/PLAW-112publ144.pdf"*, 112–144.

Crane, Melissa M, Elisabeth M Seburg, Rona L Levy, Robert W Jeffery, and Nancy E Sherwood, "Using Targeting to Recruit Men and Women of Color into a Behavioral Weight Loss Trial," *Trials*, 2020, *21* (1), 1–9.

Cutler, David M, Ellen Meara, and Seth Richards-Shubik, "Induced Innovation and Social Inequality Evidence from Infant Medical Care," *Journal of Human Resources*, 2012, *47* (2), 456–492.

Einiö, Elias, Josh Feng, and Xavier Jaravel, "Social Push and the Direction of Innovation," *SSRN Working Paper 3383703*, 2019.

FDA, "General Considerations for the Clinical Evaluation of Drugs," in "General Considerations for the Clinical Evaluation of Drugs" 1977, pp. 11–11.

_ , "Review of Ambien," *Drugs@FDA*, 1992, *https://www.accessdata.fda.gov/drugsatfda_docs/nda/pre96/019908_S000_STATR.pdf*.

_ , "Risk of Next-Morning Impairment after Use of Insomnia Drugs; FDA Requires Lower Recommended Doses for Certain Drugs Containing zolpidem (Ambien, Ambien CR, Edluar, and Zolpimist)," *FDA Drug Safety Communication*, 2013, *https://www.fda.gov/drugs/drug-safety-and-availability/questions-and-answers-risk-next-morning-impairment-after-use-insomnia-drugs-fda-req*

Feldman, Sergey, Waleed Ammar, Kyle Lo, Elly Trepman, Madeleine van Zuylen, and Oren Etzioni, "Quantifying Sex Bias in Clinical Studies at Scale with Automated Data Extraction," *JAMA network open*, 2019, *2* (7), e196700–e196700.

Friedman, Lawrence M, Curt D Furberg, David L DeMets, David M Reboussin, and Christopher B Granger, *Fundamentals of Clinical Trials*, Springer, 2015.

Geller, Stacie E, Abby Koch, Beth Pellettieri, and Molly Carnes, "Inclusion, Analysis, and Reporting of Sex and Race/Ethnicity in Clinical Trials: Have we Made Progress?," *Journal of Women's Health*, 2011, *20* (3), 315–320.

_ **, Marci Goldstein Adams, and Molly Carnes**, "Adherence to Federal Guidelines for Reporting of Sex and Race/Ethnicity in Clinical Trials," *Journal of Women's Health*, 2006, *15* (10), 1123–1131.

Hariton, Eduardo and Joseph J Locascio, "Randomised Controlled Trials the Gold Standard for Effectiveness Research," *BJOG: an International Journal of Obstetrics and Gynaecology*, 2018, *125* (13), 1716.

Harris, David J and Pamela S Douglas, "Enrollment of Women in Cardiovascular Clinical Trials Funded by the National Heart, Lung, and Blood Institute," *New England Journal of Medicine*, 2000, *343* (7), 475–480.

Hoel, Andrew W, Ahmed Kayssi, Soma Brahmanandam, Michael Belkin, Michael S Conte, and Louis L Nguyen, "Under-Representation of Women and Ethnic Minorities in Vascular Surgery Randomized Controlled Trials," *Journal of Vascular Surgery*, 2009, *50* (2), 349–354.

Jaravel, Xavier, "The Unequal Gains from Product Innovations: Evidence from the US Retail Sector," *The Quarterly Journal of Economics*, 2019, *134* (2), 715–783.

Jones, Rachel, Lorraine J Lacroix, and Eloni Porcher, "Facebook Advertising to Recruit Young, Urban Women into an HIV Prevention Clinical Trial," *AIDS and Behavior*, 2017, *21* (11), 3141–3153.

Junod, Suzanne White, "FDA and Clinical Drug Trials: a Short History," *A Quick Guide to Clinical Trials*, 2008, pp. 25–55.

Kalliainen, Loree K, Ian Wisecarver, Allison Cummings, and Jennifer Stone, "Sex Bias in Hand Surgery Research," *The Journal of Hand Surgery*, 2018, *43* (11), 1026–1029.

King, Kathleen B and Mary Ann McGuire, "Symptom Presentation and Time to Seek Care in Women and Men with Acute Myocardial Infarction," *Heart & Lung*, 2007, *36* (4), 235–243.

Kleppinger, Cynthia F, "Investigator Responsibilities–Regulation and Clinical Trials," *FDA*, 2012.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson, "Inventor Gender and the Direction of Invention," *AEA Papers and Proceedings*, 2020, *110*, 250–54.

_ , _ , and _ , "Who do we invent for? Patents by women focus more on women's health, but few women get to invent," *Science*, 2021, *372* (6548), 1345–1348.

Law, Michael R, Yuko Kawasumi, and Steven G Morgan, "Despite Law, Fewer Than One in Eight Completed Studies of Drugs and Biologics are Reported on Time on ClinicalTrials.gov," *Health Affairs*, 2011, *30* (12), 2338–2345.

Legge, F, D Eaton, R Molife, G Ferrandina, I Judson, J de Bono, and S Kaye, "Participation of Patients with Gynecological Cancer in Phase I Clinical Trials: Two Years Experience in a Major Cancer Center," *Gynecologic Oncology*, 2007, *104* (3), 551–556.

Mastroianni, Anna C, Ruth Faden, and Daniel Federman, "Women and Health Research: a Report from the Institute of Medicine," *Kennedy Institute of Ethics Journal*, 1994, *4* (1), 55–62.

Maxwell, Annette E, Roshan Bastani, Perlaminda Vida, and Umme S Warda, "Strategies to Recruit and Retain Older Filipino–American Immigrants for a Cancer Screening Study," *Journal of Community Health*, 2005, *30* (3), 167–179.

Michelman, Valerie and Lucy Msall, "Sex, Drugs, and R&D: Missing Innovation from Regulating Female Enrollment in Clinical Trials," *Working Paper*, 2022, *https://uchicago.app.box.com/s/yuljg219q3ar536npkew1x18q7419280*.

Moore, Thomas J, Hanzhe Zhang, Gerard Anderson, and G Caleb Alexander, "Estimated Costs of Pivotal Trials for Novel Therapeutic Agents Approved by the US Food and Drug Administration, 2015-2016," *JAMA Internal Medicine*, 2018, *178* (11), 1451–1457.

Murthy, Vivek H, Harlan M Krumholz, and Cary P Gross, "Participation in Cancer Clinical Trials: Race-, Sex-, and Age-based disparities," *JAMA*, 2004, *291* (22), 2720–2726.

Nadel, Mark V, "Womenś Health: FDA Needs to Ensure More Study of Gender Differences in Prescription Drug Testing," *Washington, DC: US House of Representatives General Accounting Office*, 1992.

Nielsen, Mathias Wullum, Jens Peter Andersen, Londa Schiebinger, and Jesper W Schneider, "One and a Half Million Medical Papers Reveal a Link between Author Gender and Attention to Gender and Sex Analysis," *Nature Human Behaviour*, 2017, *1* (11), 791–796.

NIH, "Review of the Literature: Primary Barriers and Facilitators to Participation in Clinical Research," *National Institutes of Health*, 2015, *https://orwh.od.nih.gov/sites/orwh/files/docs/orwh_outreach_toolkit_litreview.pdf*.

Office, United States General Accounting, "Most drugs withdrawn in recent years had greater health risk for women," 2001.

Oster, Emily, "Unobservable selection and coefficient stability: Theory and evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Plakogiannis, Roda and Sally A Arif, "Women versus men: is there equal benefit and safety from statins?," *Current atherosclerosis reports*, 2016, *18* (2), 1–5.

Ramasubbu, Kumudha, Hitinder Gurm, and David Litaker, "Gender Bias in Clinical Trials: Do Double Standards Still Apply?," *Journal of Women's Health & Gender-Based Medicine*, 2001, *10* (8), 757–764.

Safdar, Basmah, Alyson J McGregor, Sherry A McKee, Asad Ali, Radu Radulescu, Nadine T Himelfarb, Matthew R Klein, and Carolyn M Mazure, "Inclusion of Gender in Emergency Medicine Research," *Academic Emergency Medicine*, 2011, *18* (2), e1–e4.

Shields, Kristine E and Anne Drapkin Lyerly, "Exclusion of Pregnant Women from Industry-Sponsored Clinical Trials," *Obstetrics & Gynecology*, 2013, *122* (5), 1077–1081.

Steinberg, Jecca R, Brandon E Turner, Brannon T Weeks, Christopher J Magnani, Bonnie O Wong, Fatima Rodriguez, Lynn M Yee, and Mark R Cullen, "Analysis of Female Enrollment and Participant Sex by Burden of Disease in US Clinical Trials Between 2000 and 2020," *JAMA Network Open*, 2021, *4* (6), e2113749–e2113749.

Wang, Tracy Y and Jacqueline Talea DesJardin, "Time to End 'Manels' in Clinical Trial Leadership," *JAMA Internal Medicine*, 2020, *180* (10), 1383–1384.

Yu, Yue, Jun Chen, Dingcheng Li, Liwei Wang, Wei Wang, and Hongfang Liu, "Systematic analysis of adverse event reports for sex differences in adverse drug events," *Scientific reports*, 2016, *6* (1), 1–9.

Zarin, Deborah A, Tony Tse, Rebecca J Williams, and Thiyagu Rajakannan, "Update on Trial Registration 11 Years After the ICMJE policy was Established," *New England Journal of Medicine*, 2017, *376* (4), 383–391.

Zopf, Y, C Rabe, A Neubert, KG Gassmann, W Rascher, EG Hahn, K Brune, and H Dormann, "Women Encounter ADRs More Often than Do Men," *European Journal of Clinical Pharmacology*, 2008, *64* (10), 999–1004.

Zucker, Irving and Brian J Prendergast, "Sex differences in pharmacokinetics predict adverse drug reactions in women," *Biology of sex differences*, 2020, *11* (1), 1–14.

A Data Construction and Sample Restrictions

A.1 MeSH to ICD9 Mapping

MeSH terms are a controlled and hierarchically-organized vocabulary produced by the National Library of Medicine used for indexing, cataloging, and searching of biomedical and health-related information. I only look at MeSH headings which are considered to be "Major Topics" and exclude generic MeSH terms like "disease" or "pain" manually.

I map MeSH terms to ICD9 codes using the UMLS from the NLM and the NBER ICD10 to ICD9 crosswalk. For example, the MeSH terms associated with trial NCT00292162 are "Heart Failure" and "Atrial Fibrillation". The mapped ICD9 codes are 429 ("Complications of Heart Disease") and 997 ("Complications Affected Specified Body System").

Disease Burden For my descriptive analysis, I compute the female disease burden for a trial from the CDC Wonder dataset. The dataset provides the number of deaths in the United States for each 3 digit ICD9 code in 2018. For example, approximately $23,000$ females died due to colon cancer (ICD9 code 153) in 2018. The share of female deaths for an ICD9 code is the number of female deaths divided by the total number of deaths due to that ICD9 code in 2018. The female disease burden for a given trial is the mean share of female deaths across the ICD9 codes associated with the trial.

Disease mortality is an imperfect proxy for disease burden or incidence because mortality depends on the available treatments which is likely endogenous to the composition of clinical trials. I use disease mortality as a measure of disease burden because there is no comprehensive dataset on incidence that spans all diseases. I emphasize that disease burden is used only for descriptive analysis. In my main regressions, I use disease fixed effects which does not rely on this imperfect measure of disease burden.

A.2 Sample Restrictions

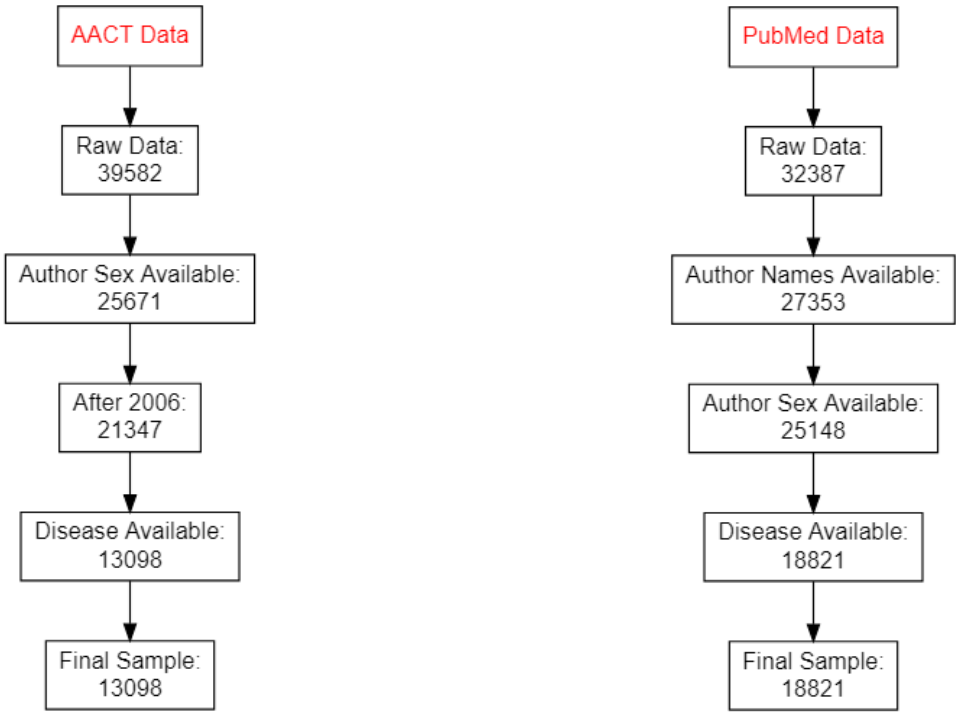
The raw data from ClinicalTrials.Gov consists of all interventional studies that were completed as of September 1st, 2021. As previously discussed, I include only trials which started on or after Jan 1st, 2007. Trials which started before 2007 did not fall under the purview of the FDAAA and were not mandated to register on ClinicalTrials.gov. Consequently, the sample of trials registered on ClinicalTrials.gov prior to 2007 is likely to be a selected sample. To limit selection bias, I only include trials registered after 2007.

The raw data for the PubMed sample consists of all published articles from the search engine Semantic Scholar and PubMed Central from 1996 to 2018. Note that although the

PubMed Extract data consists of trials from 1996, the coverage prior to 2002 is limited. In my main analysis, I account for this using year fixed effects. I only include trials from the PubMed Extract data that provided the full first name of at least one author. Further, I only include trials for which I can determine the sex of at least one author.

For both datasets, I only include trials for which I can determine the sex of at least one of the principal investigator(s). Last, I exclude trials which are not associated with at least 1 ICD-9 code. The number of unique trials after each of these restrictions is presented in Figure 4.

Figure 4: Sample Restrictions



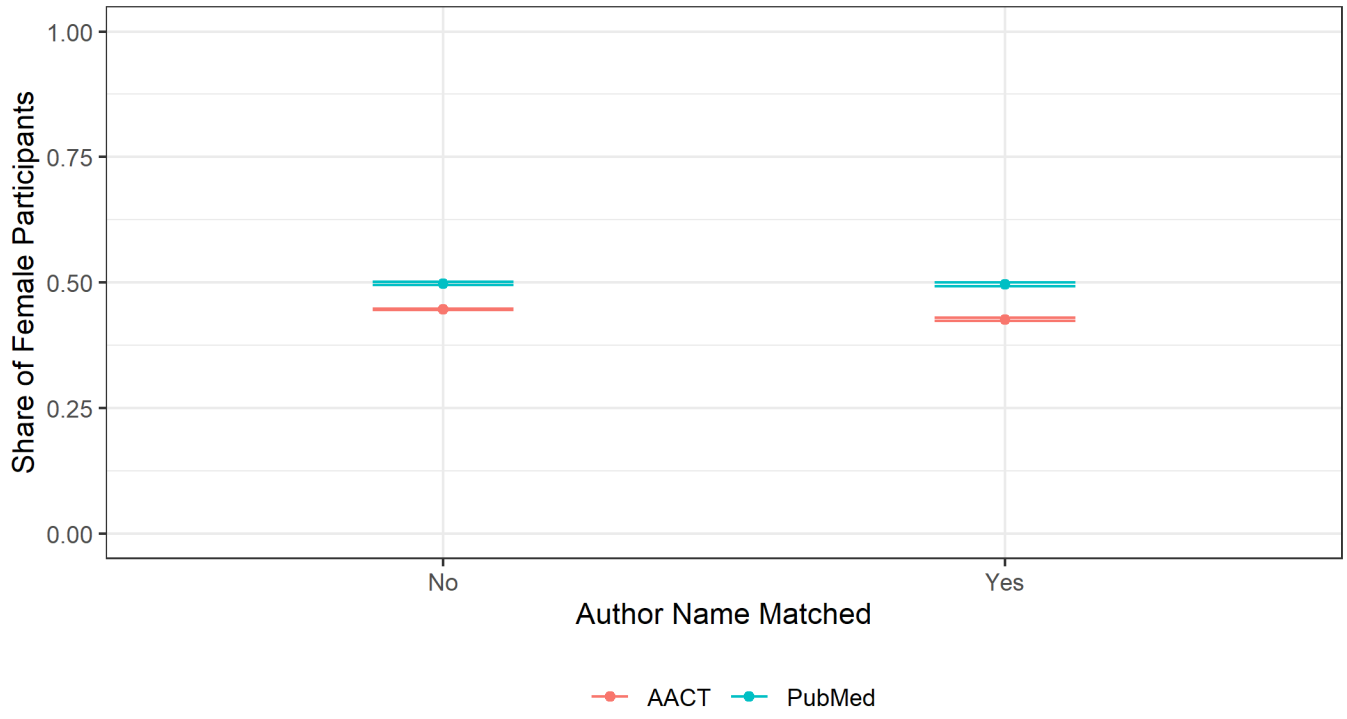
The plot presents the sample restrictions and the resulting number of unique trials from each of the restrictions for the AACT dataset (left panel) and the PubMed dataset (right panel). Each observation is a unique clinical trial. "Author Names Available" restricts the sample to studies where the names of the authors/PIs were available. "Author Sex Available" restricts the data to studies where that I could correctly identify the sex of at least one author based on their first name. "Disease Available" restricts the sample to studies which had at least one ICD9 code associated with the trial. "After 2006" restricts the sample to include trials conducted after December 31st, 2016.

A.3 Bias Due to Name Linkage

I exclude trials where I am unable to determine the sex of at least one PI. Recall, that I determine the sex of a PI by linking their name to administrative records (see Section 2.2.2 for details). Previous studies have documented that name linkage may be less accurate for names that are non-Western (eg., Asian names) or rare (Bohensky et al. (2010)). Consequently, the exclusion of trials where I am unable to link the name of authors may bias my results. Note that the direction of bias is ambiguous.

In Figure 5, I compare the share of female participants of trials in my sample to that of the trials that I exclude. The figure shows that the share of trials is economically indistinguishable across these two samples, suggesting that this exclusion criterion does not result in meaningful selection on my main outcome variable. I acknowledge that there may be selection on other variables induced by this criterion and this potential selection is a limitation of the present paper.

Figure 5: Balance on Share of Female Participants



The $y-$ axis represents the share of female participants and the $x-$ axis represents whether I am to able to match at least one author of the trial. The figure shows the mean share of female participants with 95% confidence intervals for each set of trials. AACT data and PubMed datasets are presented in red and blue, respectively.

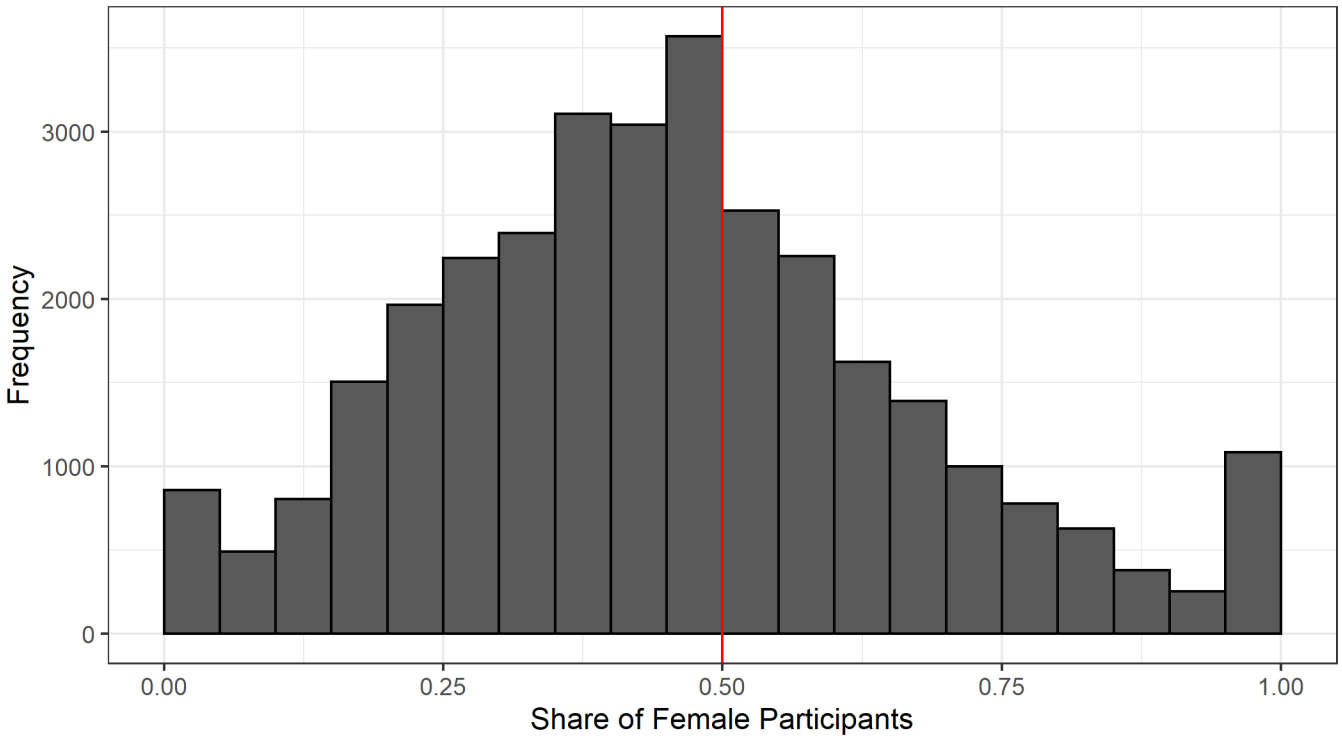
B Female Underrepresentation in Clinical Trials

Female representation for a trial is defined to be the difference between the share of female participants and female disease burden. A trial is underrepresenting females if female representation is negative.

Despite improvements in female participation over time, the existing literature has documented that female underrepresentation remains an issue (Feldman et al. (2019)). I explore basic facts about the female underrepresentation for clinical trials in my sample.

In Figure 6, I present a histogram of the share of female participants across trials [17]. The figure shows that a majority of trials have a female share of less than 50%, highlighting that females are underrepresented relative to the general population in a majority of trials.

Figure 6: Histogram of Female Participation



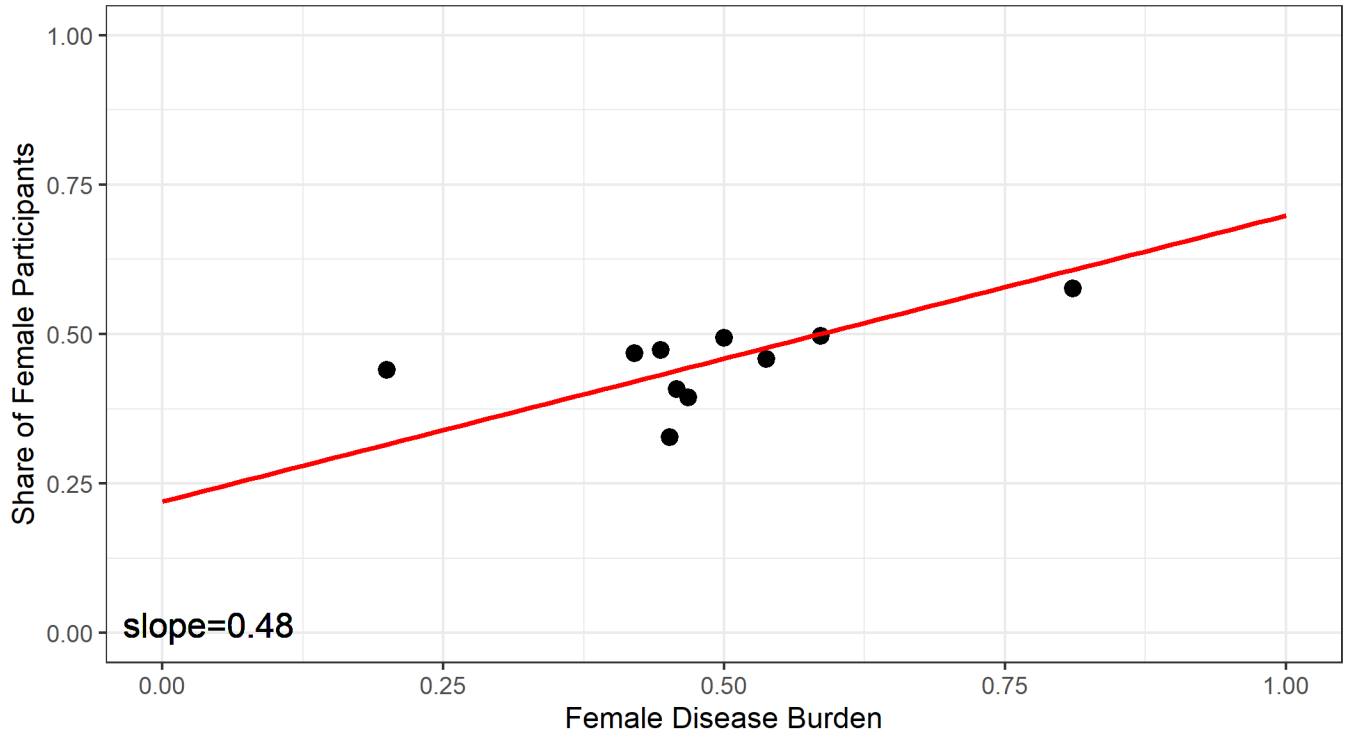
The figure presents a histogram of the share of female participants across clinical trials. Each observation is a clinical trial. The share of female participants is the number of female participants divided by the total number of participants. The red vertical line shows the point at which there is an equal number of male and female participants in clinical trials.

I also observe that there is a large mass of trials that enroll no female participants and a slightly smaller mass of trials that enroll no male participants. These trials primarily study

[17] The share of female participants for a given trial is the number of female participants divided by the total number of participants.

diseases that are sex specific like breast cancer or prostate cancer. These extremes highlight two keys points. First, disease prevalence is likely to influence female participation in clinical trials. I plot the female disease burden and share of female participants in Figure 7. The figure confirms the fact that the share of female participants is higher for trials that study diseases that affect a higher share of females.

Figure 7: Female Trial Participation and Disease Burden



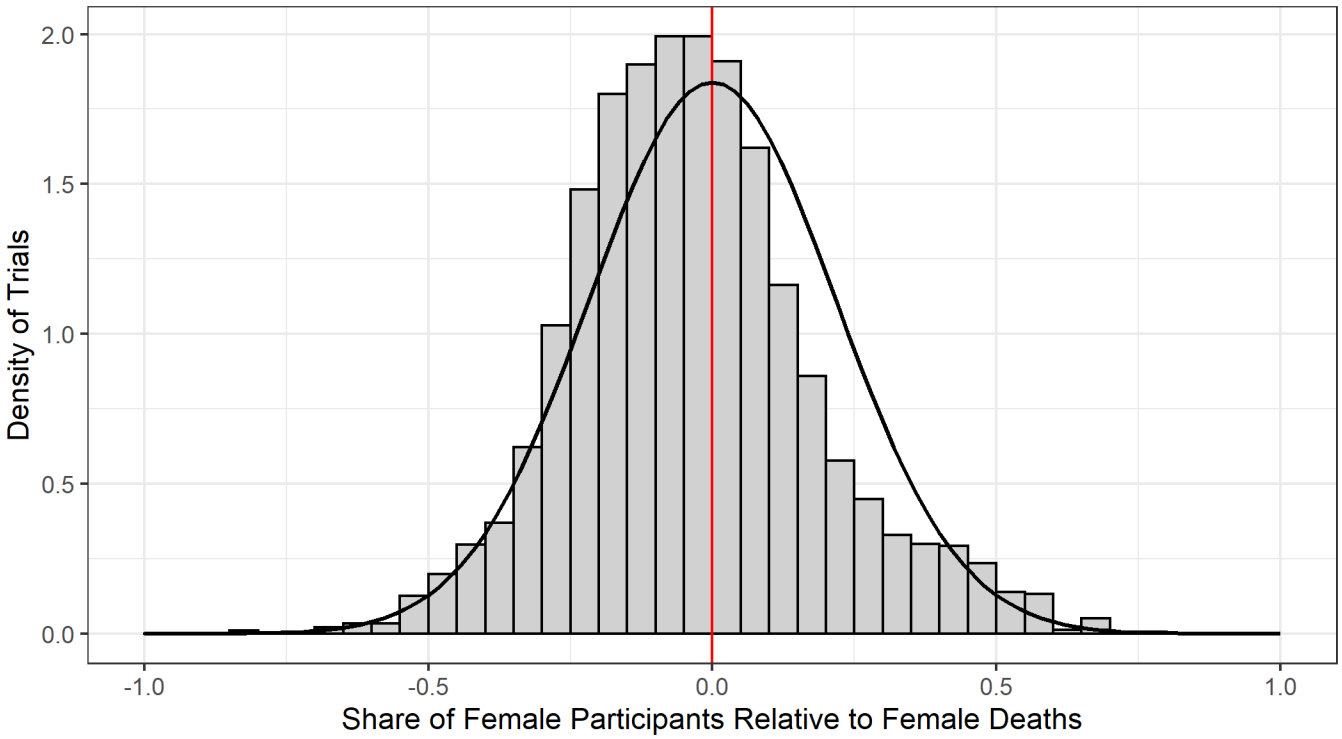
The plot shows a bin-scatter plot of the disease burden among female patients and female participation in clinical trials. Each observation is a clinical trial. The $x-$axis presents the share of female deaths for the disease (ICD9 codes) and the $y-$axis presents the share of female participants. The red line is the line of best fit from a linear regression. The bins are chosen such that each bin has the same number of observations. The slope of the fitted line is annotated in the bottom-left corner.

Second, these figures highlight that comparing female participation in trials to the general female population may not be appropriate. Conceptually, it makes little sense to say females are underrepresented in prostate cancer clinical trials if less than 50% of women are enrolled because approximately 0.01% of those who get prostate cancer are women. Motivated by this idea, in Figure 8, I present the histogram of the share of female participants after adjusting for disease prevalence. In this figure, a negative value implies female underrepresentation while a positive value implies female overrepresentation. Even after controlling for disease burden at the individual trial level, I continue to see that there is female underrepresentation

in a majority of clinical trials.

Although female underrepresentation has been previously documented, the above analysis extends the literature in two key ways. First, these figures look at a large set of trials across many diseases over time. Most papers in the literature only study a specific disease for a short period of time [18]. Second, the paper is able to adjust for disease burden at the individual trial level using granular ICD9 codes. Most previous studies either assume that the appropriate benchmark for equal representation is 0.5 or use very coarse benchmarks like the number of patients afflicted with coarse diseases categories like "heart disease"[19].

Figure 8: Relative Female Participation across Diseases



The figure presents a histogram of the share of female participation relative to disease burden. Each observation is a clinical trial. The $x-$axis represents the difference between the share of female participants and the female death share across clinical trials; see Section 2.2 for detailed definitions. A negative value implies female underrepresentation while a positive value implies female overrepresentation. The vertical red line at $x = 0$ represents diseases for which the share of female participants equals the share of female deaths.

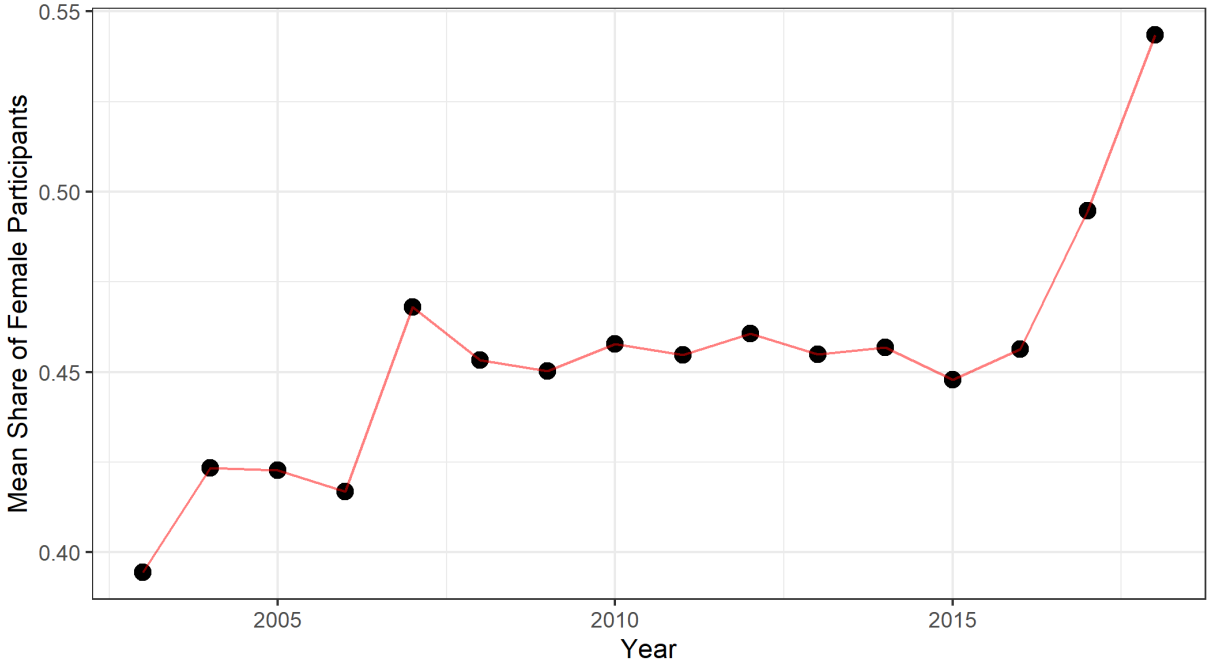
[18] Feldman et al. (2019) provides a compehensive analysis and is a notable exception.
[19] Even recent papers like Feldman et al. (2019); Steinberg et al. (2021) suffer from this limitation.

C Trends in Female PIs and Female Participants

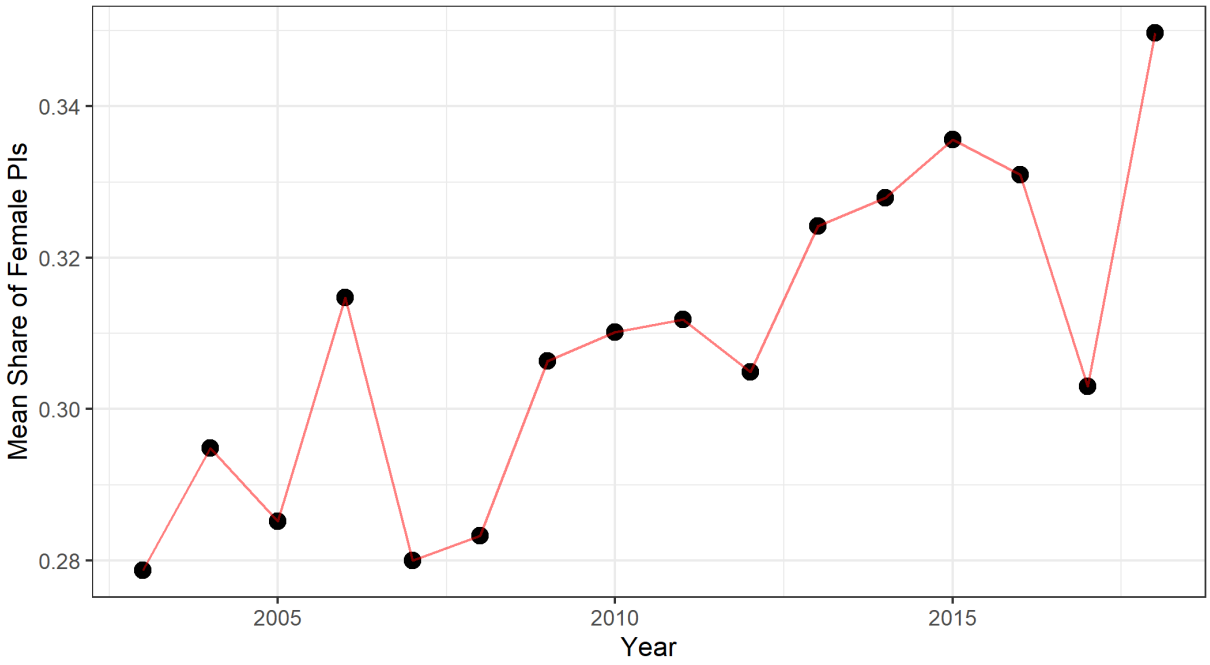
Over time, the share of female PIs and share of female participants have both increased. In Figure 9, I present the mean of the share of female PIs and share of female participants by year. The average share of female participants at the start of 2003 is approximately 0.40. Female participation has steadily increased over time and was slightly above 0.50 in 2018. Similarly, the share of female PIs has steadily increased from 0.27 in 2005 to 0.37 in 2018. The figure highlights there has been an improvement in aggregate female participation in clinical trials. Moreover, female researchers are playing an increasingly important role in biomedical innovation.

Figure 9: Female PIs and Female Participants Over Time

(a) Female Participants



(b) Female PIs



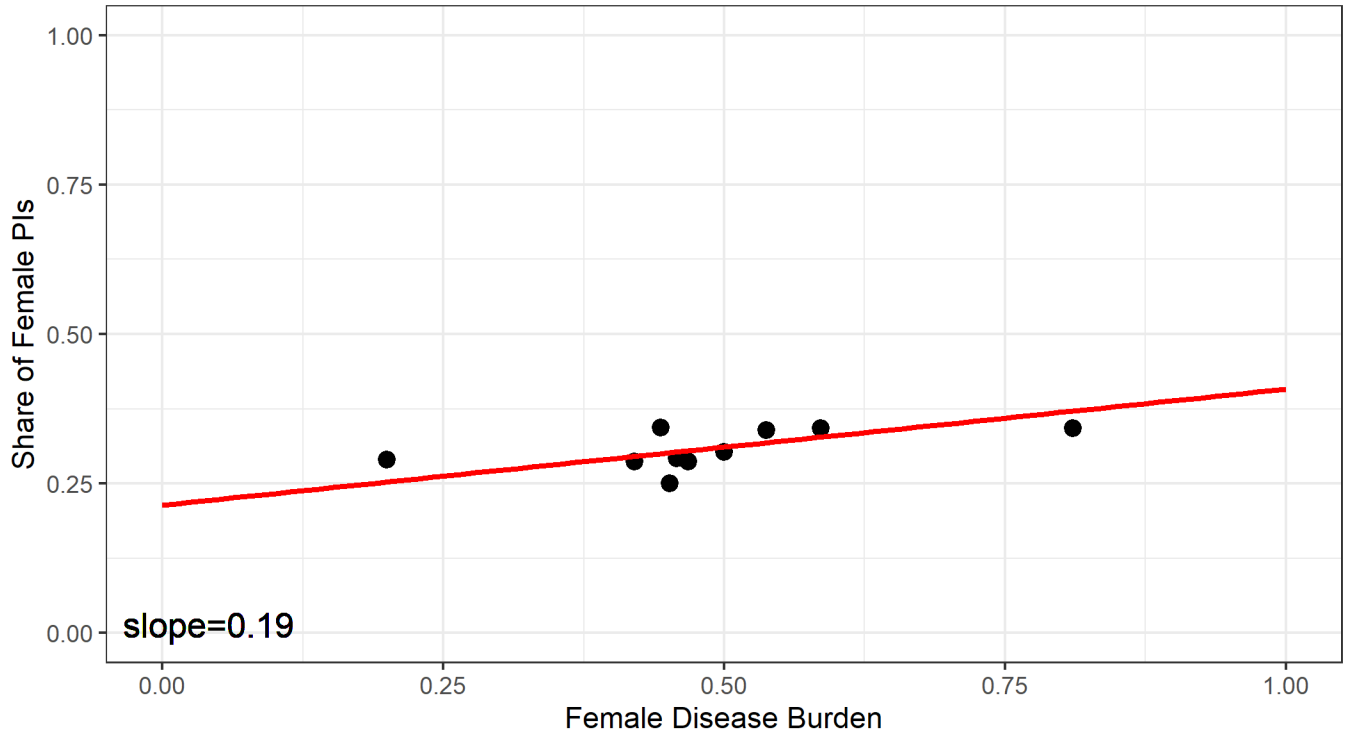
The figure presents a plot of the mean of the share of female participants (top panel) and the mean share of female PIs (bottom panel) for different years in my sample. The $x-$axis represents the calendar year and the $y-$axis represents the mean of the share of female participants/PIs for the corresponding year.

D Role of Selection into Disease

D.1 Female PIs and Disease Burden

I examine whether female PIs select into trials for disease that afflict more female patients in Figure 10. In this figure, I present a bin scatter of the share of female PIs and the female disease burden across trials. I observe that the share of female PIs is higher in trials for diseases that have a higher female disease burden. The slope of the linear relationship between disease burden and female PIs is 0.11; this relationship is very similar to the relationship between female researchers and female patents found in Koning et al. (2020). The result implies that female PIs are more likely to be researchers in diseases more likely to benefit females, in line with the recent literature on inventor gender and direction of innovation (Koning et al. (2020); Einiö et al. (2019)).

Figure 10: Female PIs and Disease Burden



The plot shows a bin-scatter plot of the disease burden among female patients and female PIs in clinical trials. An observation is a trial. The $x-$axis presents the female disease burden for the trial and the $y-$axis presents the share of female PIs. Female disease burden is defined in Section 2.2. The red line is the line of best from a linear regression of share of PIs on female disease burden. The bins are chosen such that each bin has the same number of observations. The slope of the fitted line is annotated in the bottom-left corner.

D.2 Selection and Female Participation

The discussion above highlights that female PIs are more likely to do research in diseases with a high female disease burden and such diseases more likely to have a higher share of female participants (Figure 7). Therefore, female PIs may also increase female participation by selecting into trials for diseases which would have a higher share of female participants *independent* of the share of female PIs in that trial. I briefly discuss the role of this selection in improving female participation in clinical trials.

Consider a simple example that focuses on two diseases - a diseases that afflicts only females (ICD9 625: "diseases associated with female genital organs") and a disease that afflicts only males (ICD9 600 "Hyperplasia of prostate"). The share of female participants is 1 for the female-only disease and is 0.03 for the male-only disease. Because these diseases are sex specific, the share of female participants is likely to be very similar irrespective of the share of female PIs. Among these two diseases, if there is at least one female PI in a trial then the probability that the trial studies the female-only disease is 0.8. On the other hand, if the team has only male PIs then the probability that the trial studies the female-only disease is 0.12. In other words, the probability that a trial studies a female-only disease increases by 0.68 when all trials included at least one female PI compared to when all PIs are male. Consequently, the aggregate share of female participants would increase by 0.63 solely due to selection into the female-only disease by female PIs.

The diseases chosen for this example are, by construction, not representative. I emphasize that this example should be viewed solely as illustrative; I do not make any general claims about the role of selection into diseases in the present paper. A rigorous and complete analysis of this potential mechanism remains an avenue for future research.

E Robustness

E.1 Selection Into Diseases

As discussed in Section 3, a crucial confounder is the disease that a trial in investigating. In my main analysis, I attempt to account for this selection using disease $\times$ year fixed effects, where disease is measured using ICD9 codes at the 3-digit level. In this section, I provide further evidence to show that my results are not driven by selection into disease. The results are presented in Table 2.

One potential concern is that 3-digit ICD9 codes are not sufficiently granular and there may be selection within these codes. In column (1), I include even finer 5-digit ICD9 codes as controls and find that the parameter estimates of β with these controls are statistically

indistinguishable and qualitatively similar to those reported in my main analysis.

I further show that my main results are not driven by diseases that affect predominantly one sex (such as breast cancer and prostate cancer) by excluding trials that study such diseases. I define a sex-specific trial using three different criteria and columns (2)-(4) show that parameter estimates from these restricted samples are virtually identical to those of my main results. In column (5), I match trials based on disease, year, and decile of the total number of participants and estimate the parameters within these matched sets. The advantage of the matching method is that trials that are in sets with limited overlap are excluded. The results are essentially unchanged.

Last, I control for disease selection using disease burden. Recall, disease burden is measured by mortality which is an imperfect proxy for incidence. Additionally, the use of disease burden as a linear control makes stronger parametric assumptions compared to using disease fixed effects. The parameter estimate from this regression is 0.083, which suggests that my main results are not inflated due to disease selection.

Table 2: Robustness to Selection Into Diseases

	Dependent variable:					
	Share of Female Participants					
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Female PIs	0.037	0.045	0.049	0.045	0.049	0.083
	(0.005)	(0.006)	(0.007)	(0.006)	(0.006)	(0.013)
Female Disease Burden						0.353
						(0.125)
Disease x Year F.E.	No	Yes	Yes	Yes	Yes	No
ICD9-5 x Year F.E.	Yes	No	No	No	No	No
Exclude Sex Specific	No	Yes	Yes	Yes	No	No
Criteria	-	Restrictions	Burden	Enrollment	-	
Matched Sample	No	No	No	No	Yes	No
Observations	63,258	48,170	38,634	48,170	40,705	38,838
R^2	0.445	0.274	0.307	0.274	0.249	0.050
Adjusted R^2	0.362	0.233	0.265	0.233	0.227	0.049

The table presents parameter estimates of an OLS regression of the share of female participants on the share of female PIs for different samples and specifications. In column (1) include ICD9 5-digit codes × year fixed-effect. In columns (2)-(4) I exclude trials that study diseases that affect predominantly one sex, defined in three ways. A trial is sex specific if more than 90% of trials studying the same disease completely exclude male or female patients (column (2)); second, a trial is sex specific if the female disease burden of the trial is greater than 0.9 or less than 0.1 (column (3)); third, a trial is sex specific if other trials studying the same disease enroll more than 90% female patients or less than 0.10% female patients (column (4)). In column (5), I estimate the parameters within sets of trials matched on disease, year, and decile of the total number of participants. In column (6), I use disease burden as a control.

E.2 Different Datasets

The two datasets - AACT and PubMed - potentially suffer from data limitations (as discussed in Section 2). The PubMed data is susceptible to measurement error because it employs a machine extract algorithm to collect the number of participants. Additionally, the PubMed sample may suffer from selection bias because it only includes trials that are published. The AACT may suffer from selection bias because of non-compliance with the FDAAA. To evaluate the extent to which each of these datasets affects my main result, I estimate the

regression of the share of female participants on the share of female PIs for each dataset separately.

The results are reported in Table 3. The estimate for the AACT dataset is $\beta = 0.038$ while the estimate for the PubMed dataset is $\beta = 0.069$. These estimates are quantitatively similar to my main estimates of $\beta = 0.049$ and have the same qualitative implications. The estimates are statistically significant for both datasets at all conventions levels. The robustness of the estimates suggests that potential data limitations are not biasing my results in any economically meaningful way.

Table 3: OLS Regression of Female Participation on Female PIs for Different Datasets

	Dependent variable:	
	Share of Female Participants	
	(1)	(2)
Share of Female PIs	0.037	0.069
	(0.006)	(0.006)
Disease x Year F.E.	Yes	Yes
Dataset	AACT	PubMed
Observations	22,803	34,105
R^2	0.322	0.338
Adjusted R^2	0.221	0.295

The table presents the results of an OLS regression of the share of female participants on the share of female PIs while controlling for disease × year fixed effects. Each column corresponds to a separate dataset. Column (1) presents results with the AACT data, while column (2) presents results with the PubMed data. SE are clustered at the disease level.

E.3 Cut-off for PI Sex

As discussed in Section 2.2.2, the sex of a PI is the share of births with their name that were female. The share of female PIs for a study is the mean of the probability that a PI is female across its PIs. In this section, I show robustness to classifying the sex of a PI to be female if more than 50% of births with their name were female and male otherwise. Separately, I classify the sex of a PI to be female if more than 75% of births with their name were female and male if less than 25% of birth were female. I then re-estimate the linear model defined in Equation 3.1. The results of the regression are presented in Table 4. The estimate of β is essentially identical across different methods for classifying the sex of PIs.

Table 4: Robustness to PI Sex Cut-offs

	Dependent variable:		
	Share of Female Participants		
	(1)	(2)	(3)
Share of Female PIs	0.049	0.043	0.044
	(0.005)	(0.005)	(0.005)
Cut-Off	-	0.5	0.75
Disease x Year F.E.	Yes	Yes	Yes
Observations	56,908	56,908	56,805
R^2	0.309	0.309	0.309
Adjusted R^2	0.259	0.259	0.259

The table presents the results of an OLS regression of the share of female participants on the share of female for different methods of predicting PI sex. The probability that a PI is female is the share of female births among all birth with that name . The sex of PI is the probability that PI is female, classified female if the probability is greater than 50%, and female if the probability is greater than 75% in Column (1), Column (2), and Column (3), respectively. The regressions include disease $\times$ year fixed effects. SE are clustered at the disease level.

E.4 Trial Features

I show that my main result is robust to the inclusion of different trial features as controls. Specifically, I examine if trial size, whether a trial is a multi-center trial, whether a trial is a late stage trial, NIH funding, and location(s) of the study affect the positive association between female PIs and female participants. To do so, I estimate a linear model of the form

$$Y_{idt} = \beta F_{idt} + \beta_1 X_{idt} + \gamma_{dt} + \epsilon_{idt}\,, \tag{E.1}$$

where X_{idt} is the number of participants in the trial, an indicator for multi-center trial, an indicator for late stage trials, log of NIH funding, and state fixed effects for trial i, treating disease d in year t. All other variables are as previously defined. I estimate a separate linear model for each trial feature.

Note that several of these trial features could also be mechanisms. However, I cannot empirically verify whether these are mechanisms or confounders without assuming the direction of causality. If these trial features were mechanisms, then my robustness analysis would be invalid due to the issue of "bad controls" (Angrist and Pischke (2008)).

Consider the case of research funding and assume that research funding is a mechanism rather than a confounder. In this case, the above regression which compares the effect of female participation on female PIs conditional on NIH funding is biased *even if* the share of female PIs is randomly assigned. The intuition is that because female PIs increase research funding and I randomly assign a trial to have more female PIs then that trial would have higher research funding. When I compare two trials with the same amount of research funding with different shares of female PIs I am not comparing apples to apples. This is because the fact that a trial with a low share of female PIs received the same amount of research funding as a trial a the high share of female PIs suggests that the former trial has some unobservable features that makes it superior to the latter trial[20].

Nonetheless, I perform this analysis to assuage concerns about omitted-variable bias but recommend caution while interpreting these results. The results of these regressions are presented in Table 5. The estimate of β remains positive and statistically significant across all specifications. For all trial features except late stage trials, the point estimate ranges from 0.035 to 0.048 which is close to my main result. The point estimate is lower when I include an indicator for Late Stage trials as a control. One reason for this might be that only late stage trials have the incentive to include more female participants. Another reason could be the "bad controls" issue I discuss above. Overall, my main result is qualitatively robust and quantitatively similar when including different trial features as controls.

[20]See Angrist and Pischke (2008) for a detailed discussion.

Table 5: Robustness to Trial Features

	Dependent variable:			
	Share of Female Participants			
	(1)	(2)	(3)	(4)
Share of Female PIs	0.048	0.035	0.025	0.046
	(0.00003)	(0.006)	(0.007)	(0.005)
Trial Size	0.000			
	(0.000)			
Multi-center		−0.003		
		(0.006)		
Late Stage			0.027	
			(0.005)	
Log of NIH Funding				0.002
				(0.0004)
Disease x Year F.E.	Yes	Yes	Yes	Yes
PubMed Data	Yes	No	No	Yes
State F.E.	No	No	No	No
Observations	56,887	21,671	15,195	56,908
R^2	0.309	0.322	0.371	0.312
Adjusted R^2	0.259	0.217	0.253	0.262

The table presents the results of an OLS regression of the share of female participants on the share of female while controlling for different trial features, as specified in Equation E.1. Column (1), (2), (3), (4), and (5) presents results for regressions which uses the number of participants in the trial ("Trial Size"), an indicator for a multi-center trial, an indicator for a late-stage trial, log of NIH funding, and state fixed effects, respectively as controls. All regressions include disease × year fixed effects. SE are clustered at the disease level. Some of the trial features are not available in the PubMed data, as noted in the "PubMed Data" row.

F Additional Mechanisms

Research Funding A commonly cited barrier to the inclusion of female (and other minority) participants is that more resources are necessary to recruit from these populations

(NIH (2015)). Costs of recruitment are higher because of challenges in advertising to minorities, providing care to uninsured patients, and higher attrition rates among minorities. Research funding can help cover these costs and increase female participation. If female PIs are better at obtaining research funding they can improve female participation by utilizing their resources.

I empirically examine whether NIH funding is positively associated with the share of female PIs and the share of female participants. I do so by estimating the following linear model:

$$Y_{idt} = \beta_1 \log\left(\text{Funding}_{idt} + 1\right) + \gamma_{dt} + \epsilon_{idt}\,, \tag{F.1}$$

where Y_{idt} is the share of female participants and (in a separate regression) the share of female PIs; the total amount of NIH funding is denoted by Funding_{idt} and all other variables are as previously defined. The results of these regressions are reported in Table 6.

The results show that funding amount is positively and statistically significantly associated with the share of female PIs and the share of female participants. Doubling the funding amount[21] modestly increases the share of female PIs and share of female participants by 0.005 and 0.002, respectively.

These associations imply that NIH funding is either a mechanism or confounder for the positive association between share of female PIs and share of female participants. If female PIs are more successful in obtaining research funding and this increased research funding increases the share of female participants, then research funding is a mechanism.

However, if having more funding results in more female PIs and independently results in more female participants, then research funding is a confounder. Without assuming the direction of causality, I cannot firmly conclude whether research funding is a mechanism or confounder. Institutional knowledge suggests that PIs of a study are determined prior to funding applications and awards. If this is generally true then research funding is a mechanism. However, I am unable to empirically test this premise. I show robustness to research funding as a confounder in Appendix E.4.

[21] that is, a 100% increase.

Table 6: OLS Regression of Female PIs and Female Participants on NIH Funding

	Dependent variable:	
	Share of Female PIs	Share of Female Participants
	(1)	(2)
Log of Total NIH Funding	0.005	0.002
	(0.0003)	(0.0004)
Disease x Year F.E.	Yes	Yes
Observations	56,928	56,908
R^2	0.131	0.308
Adjusted R^2	0.067	0.257

The table presents the results of an OLS regression of the share of female PIs and share of female participants on the log of total NIH research funding, as specified in Equation F.1. The outcome variable is the share of female PIs and the share of female participants in Column (1) and Column (2), respectively. Both regressions include disease × year fixed effects. SE are clustered at the disease level.

Advertising The communication and advertising of trial can also play a key role in female participation. Having faces or names of female PIs on advertisements and fliers may improve take-up due to increased familiarity. Obviously, this is only possible if there are female researchers involved with the trial. A recent study showed that targeted advertisements to female participants has shown to boost participation (Crane et al. (2020)). Another study showed that advertising on online platforms used more frequently by women (like Facebook) improved recruitment of young women (Jones et al. (2017)). To test whether advertising is a potential mechanism, one would need information on advertising to female patients in clinical trials and how much PIs influence advertising decisions. This analysis is outside the scope of the present paper and remains an avenue for future research.

Trial Logistics One major burden for participants is transportation to the research facility and/or health care facility. For example, Legge et al. (2007) observed that the probability a patient enrolls in a gynecological clinical trial decreases in their distance from the hospital. Female PIs could potentially select sites for the trial in regions that are more accessible by female patients. For example, sites that are serviced with safe and reliable public transit.

Another common reason why people do not participate in clinical trials is because of childcare responsibilities. It is well known that on average the female parent bears a greater share of childcare burden than the male parent (Del Boca et al. (2020)). Consequently, any

policies that help in reducing childcare burdens as providing day care services at the trial site, fewer clinical visits, and appropriate timing of clinical trials, will likely increase female enrollment. Female PIs can adjust the frequency, duration, and time of clinical visits to help alleviate childcare burden.