Differential Impacts of Online Ratings in the Market for Medical Services

Presenter: Aaron P. Kaye, University of Michigan

Michael Luca, Harvard Business School

Sonal Vats, Boston University

AHEPE, June 23,2023

Motivation: Online Reputation and Medical Services

Online ratings and reviews are an increasingly important driver of economic activity and consumer decision-making

- Top Industries: Restaurants, Hotels, Medical Services (Local consumer Review Survey 2020)

Physician services are a credence good, meaning consumers face ex-ante and ex-post uncertainty about quality

- Ex-ante uncertainty – like experience goods, ratings could provide useful information
- Ex-post uncertainty – unclear what information ratings include

Reputation systems could mitigate or exasperate existing disparities in the medical services industry

Studying a platform with building ratings and booking allows us to better understand important mechanisms in this market

Introduction

Research questions

- What is the impact of ratings on demand for physician services?
- Investigate differential impact of ratings depend on other characteristics?

Context

- Primary Care Physicians on ZocDoc.com in 8 Metropolitan Divisions Feb 2016 - April 2017

Data sources

- Physician Information – Profile information (PCPs) collected by scraping ZocDoc
- Patient volume – Imputed from scraping physician schedules

Methodology

- Regression discontinuity design with multiple cumulative cutoffs (RDMCC)

Differential impact - Repeat analysis for economically interesting subgroups

- Physician Gender
- Number of Ratings (Bayesian learning)
- Hospital Affiliation   (other quality signal)

Background – Recent Literature

Impact of Ratings in Healthcare

- Patients are willing to travel further to receive care from hospitals with higher Yelp ratings (McCarthy, Sanbower, and Sánchez Aragón, 2022)
- Positive ratings increase general practitioner enrollment (Brown, Hansman, Keener, and Veiga, 2023)

Differential Impact of Ratings and Quality Signals

- Impact of ratings could be mediated by private information (Brown, et al ,2023)
- Signals of doctor quality reduce 90% of the racial gaps in willingness to pay (Chan, 2022)
- Women surgeons experience a larger drop in referrals after a patient death (Sarsons, 2017)
- Platform mechanics mediate the impact of ratings (Athey, and Kaye, in progress)

Background on ZocDoc.com: An Online Doctor Reservation Platform

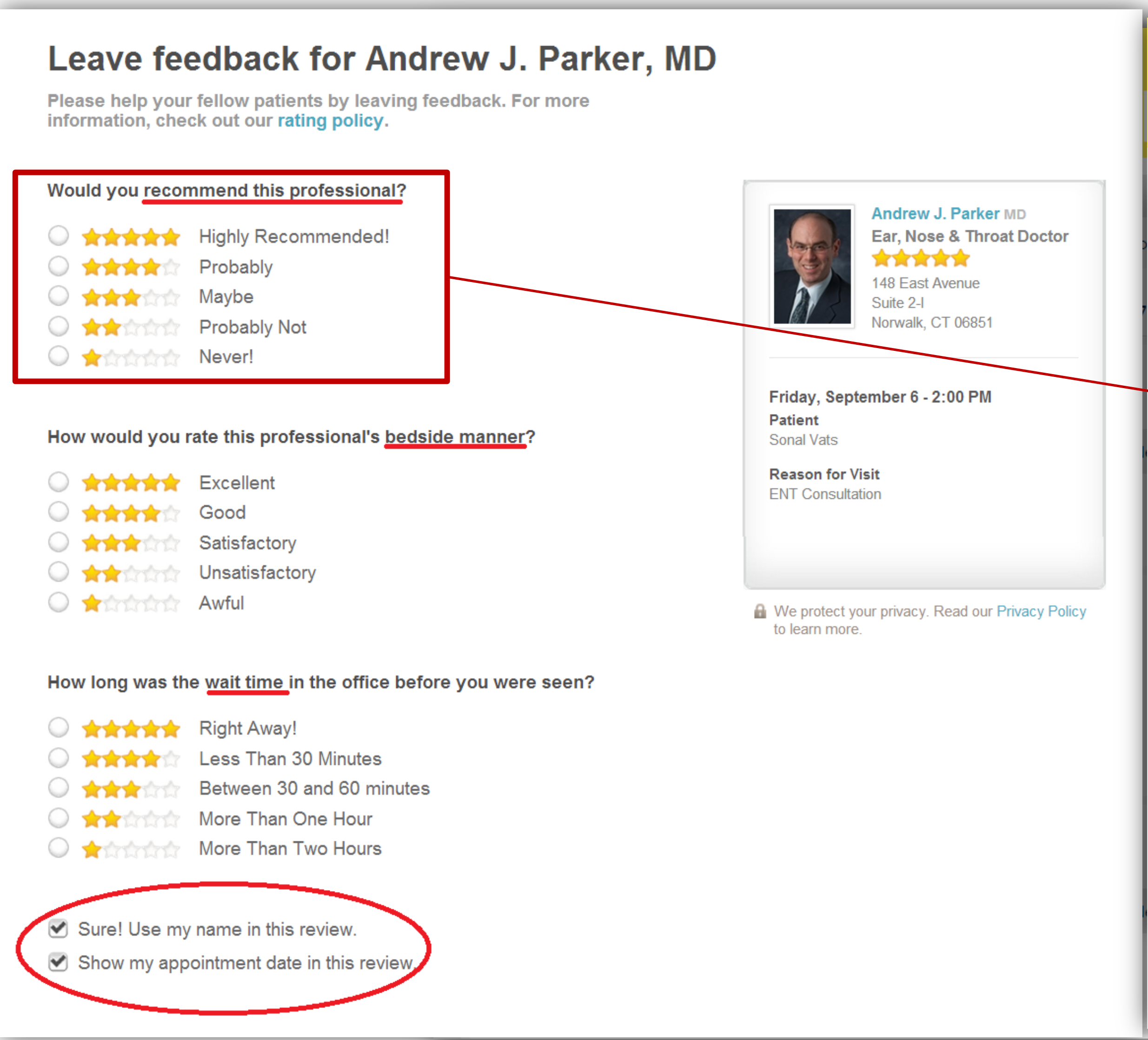
Company timeline:
- 2007: Founded
- 2015: Valued at $1.8 billion

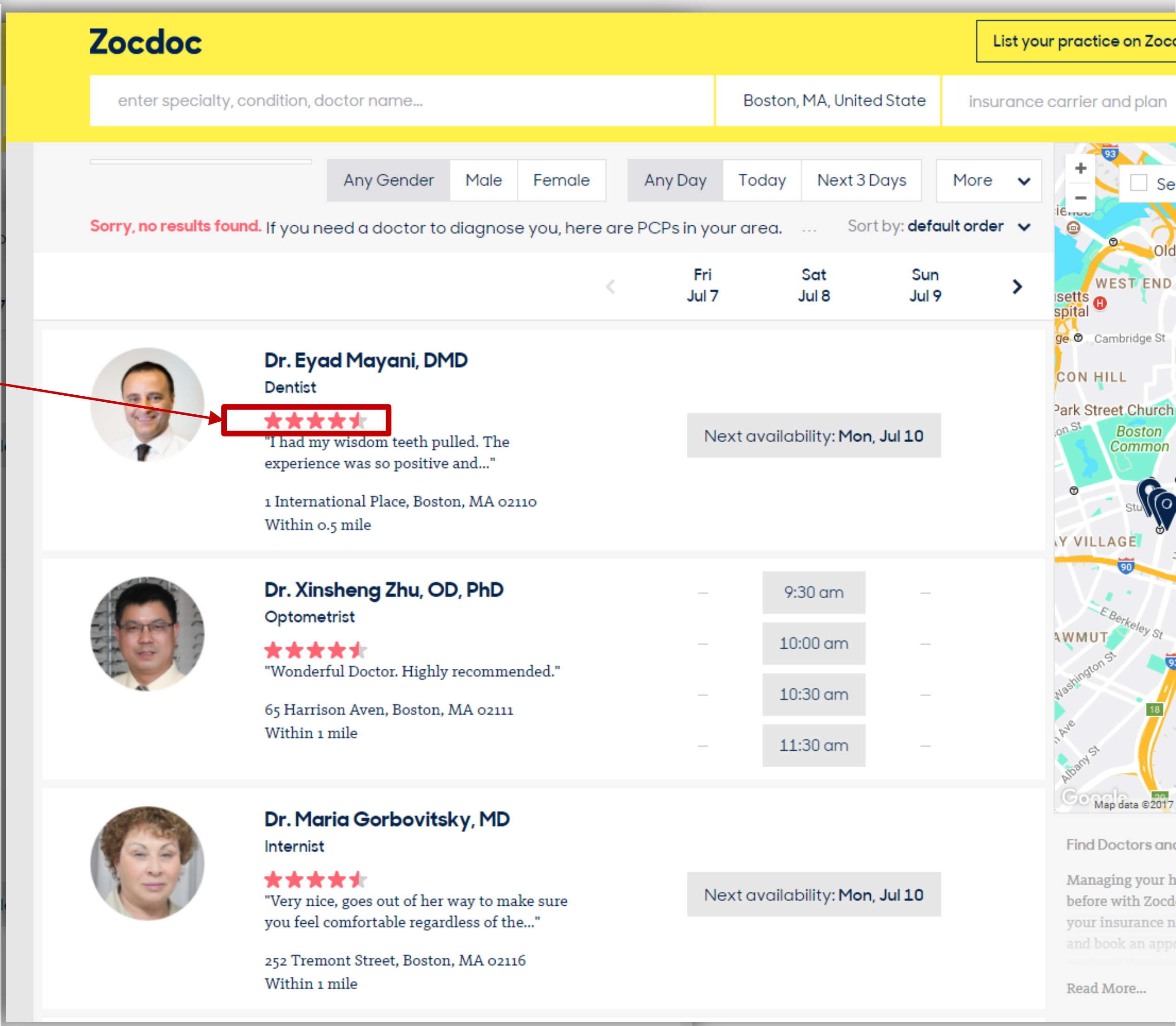
Revenue model charges physicians not patients
- 2015-2018: Physicians subscribe to $300 monthly or $3000 annual contracts
- 2018-2019: Shifted to per-booking fee

Patients can search for physicians by
- insurance, location, specialty etc. and book an appointment

Key features:
- Bundles reviews with appointments
- Verified reviews, less potential for review fraud
- Closed loop review System
- Doctors cannot screen patients

Background on ZocDoc





Preview of Findings

Descriptive Evidence

- Booking likelihood: More likely to be booked
- Booking speed: Booked further in advance

Regression Discontinuity at 5-Stars

- Patient volume via bookings: Approx. twice as many bookings
- Patient volume via vacancies: Approx. half as many vacancies

Differential Impact

- Physician gender: Effect greatest for women physicians
- Number of ratings: Effect increases with number of ratings
- Hospital affiliation: No significant difference

Robustness

- Placebo tests: Effect greatest at true cutoff
- Rating manipulation: Bunching above cutoff

Data

COLLECTION

SAMPLE RESTRICTIONS

Data Collection

Data Collected by Crawling ZocDoc's website

Time Period

- February 25, 2016 – April 17, 2017

Profile photos processed with Microsoft Face API

Region – Coordinates in the following Metropolitan Divisions

Metro Division	*Apts.*	*PCPs*
Boston, MA	86,512	117
Cambridge-Newton-Framingham, MA	71,159	68
Chicago-Naperville-Arlington Heights, IL	795,000	331
Fort Lauderdale-Pompano Beach-Deerfield, FL	184,185	80
New York-Jersey City-White Plains, NY-NJ	3,629,392	1,291
San Francisco-Redwood City-South San Fr., CA	69,384	31
Silver Spring-Frederick-Rockville, MD	232,236	82
Washington-Arlington-Alexandria, DC-VA-MD	774,756	305

Sample Selection

Appointment Sample

- Appointment type: new patient, illness, cross-listed
- Appointments on weekdays between 8am and 6pm
- At least one appointment available three weeks in advance

Physician-week sample

- “Stable” half star rating
  - 90% of observation at this rating
- Remove physicians with deleted reviews
  - More than 4 weeks with a decrease in number of reviews
- At least 8 ratings
- At least one appointment available three weeks in advance

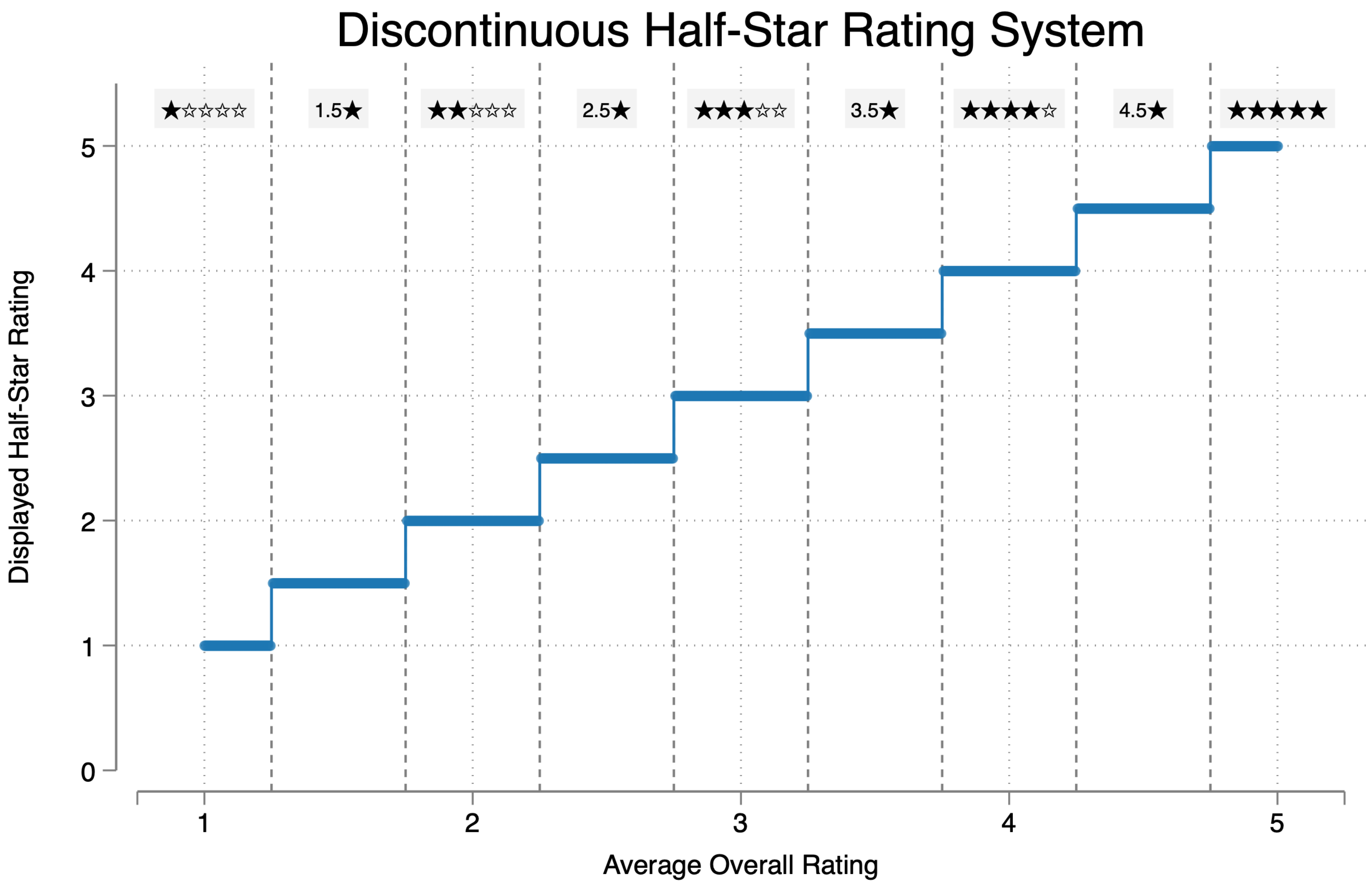
Empirical Strategy

FIRST STAGE: HALF-STAR RATINGS

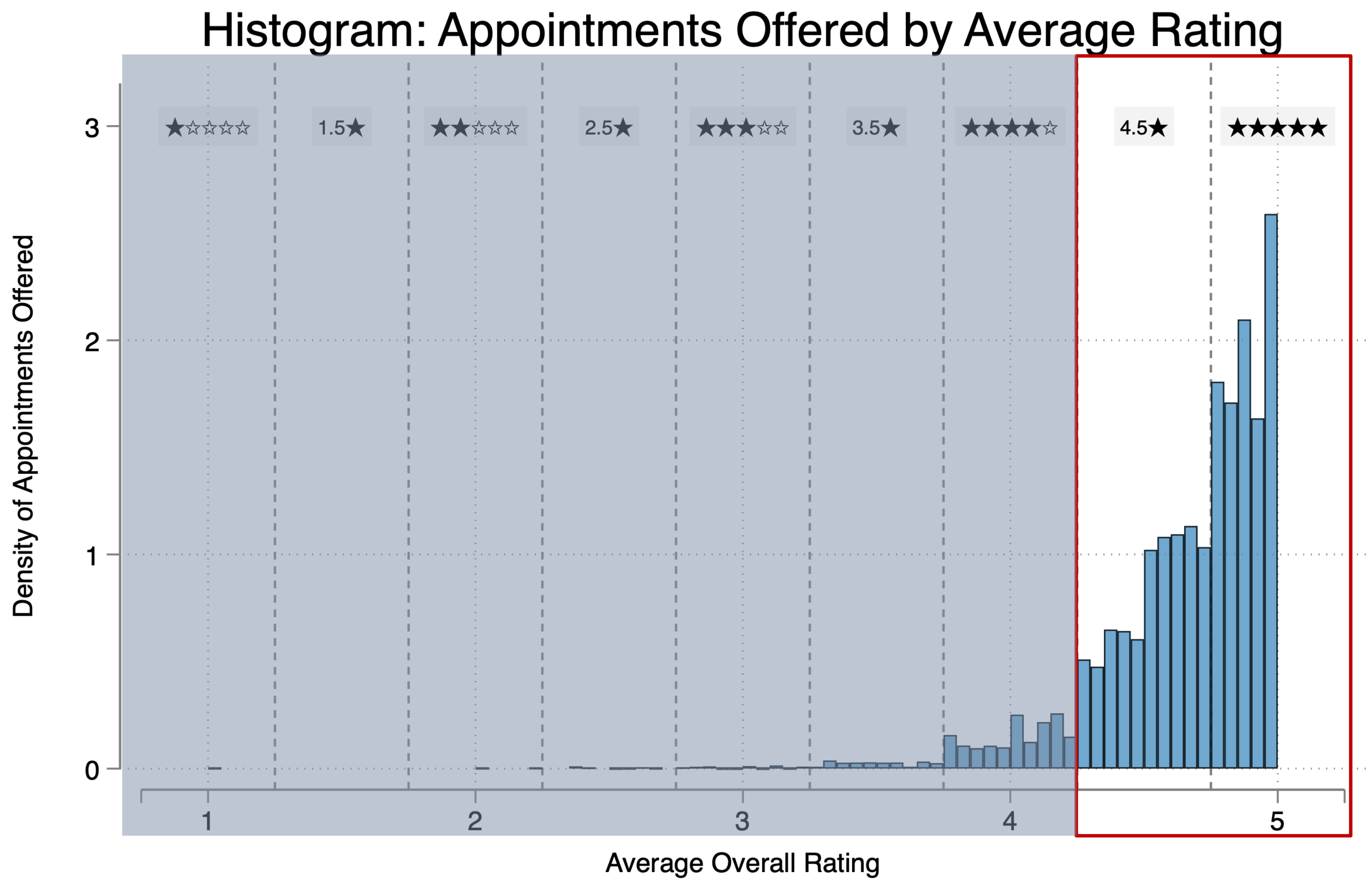
DESCRIPTIVE EVIDENCE

PRIMARY SPECIFICATION

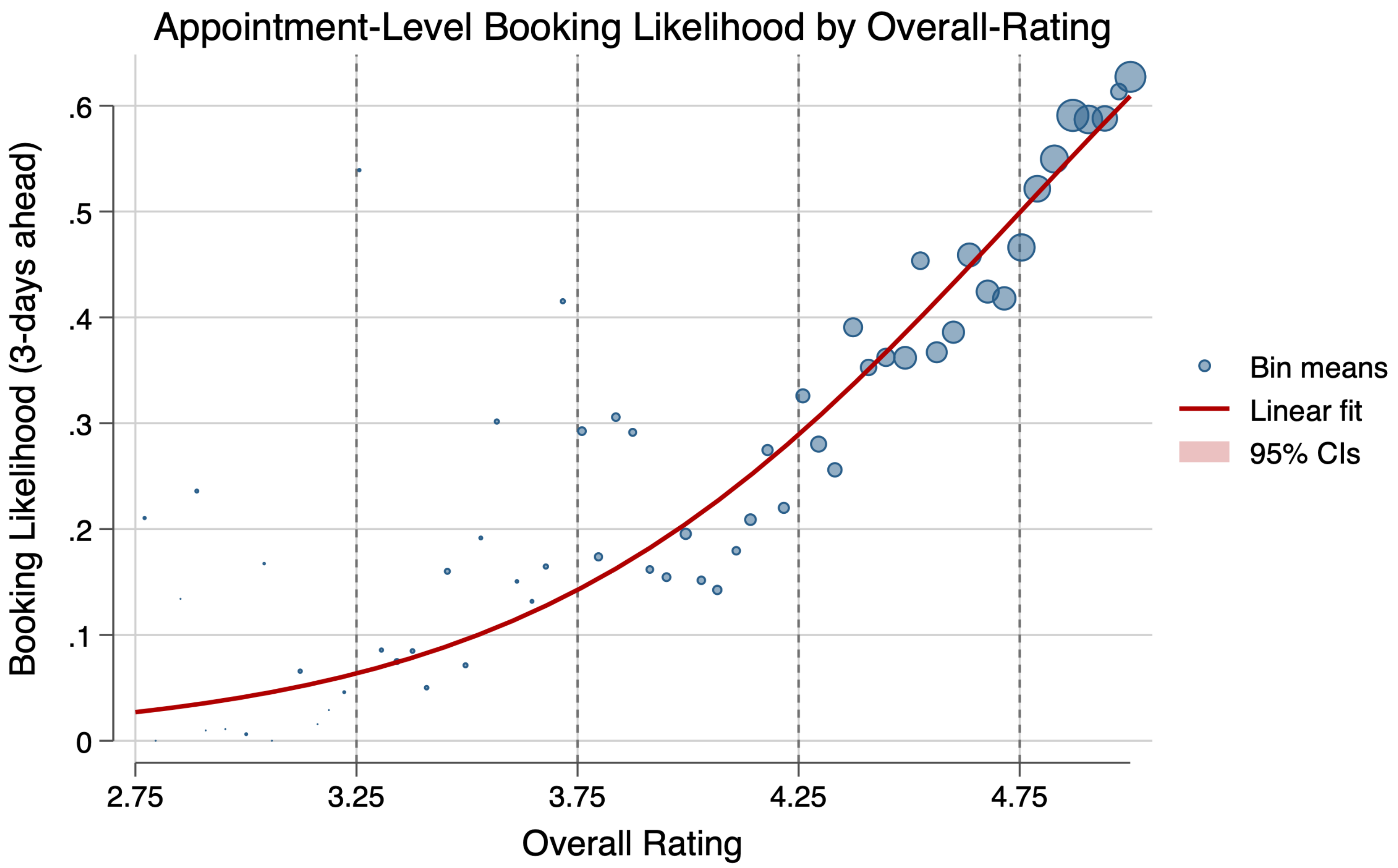
Empirical Strategy:
regression discontinuity w/ multiple cumulative cutoffs



Empirical Strategy



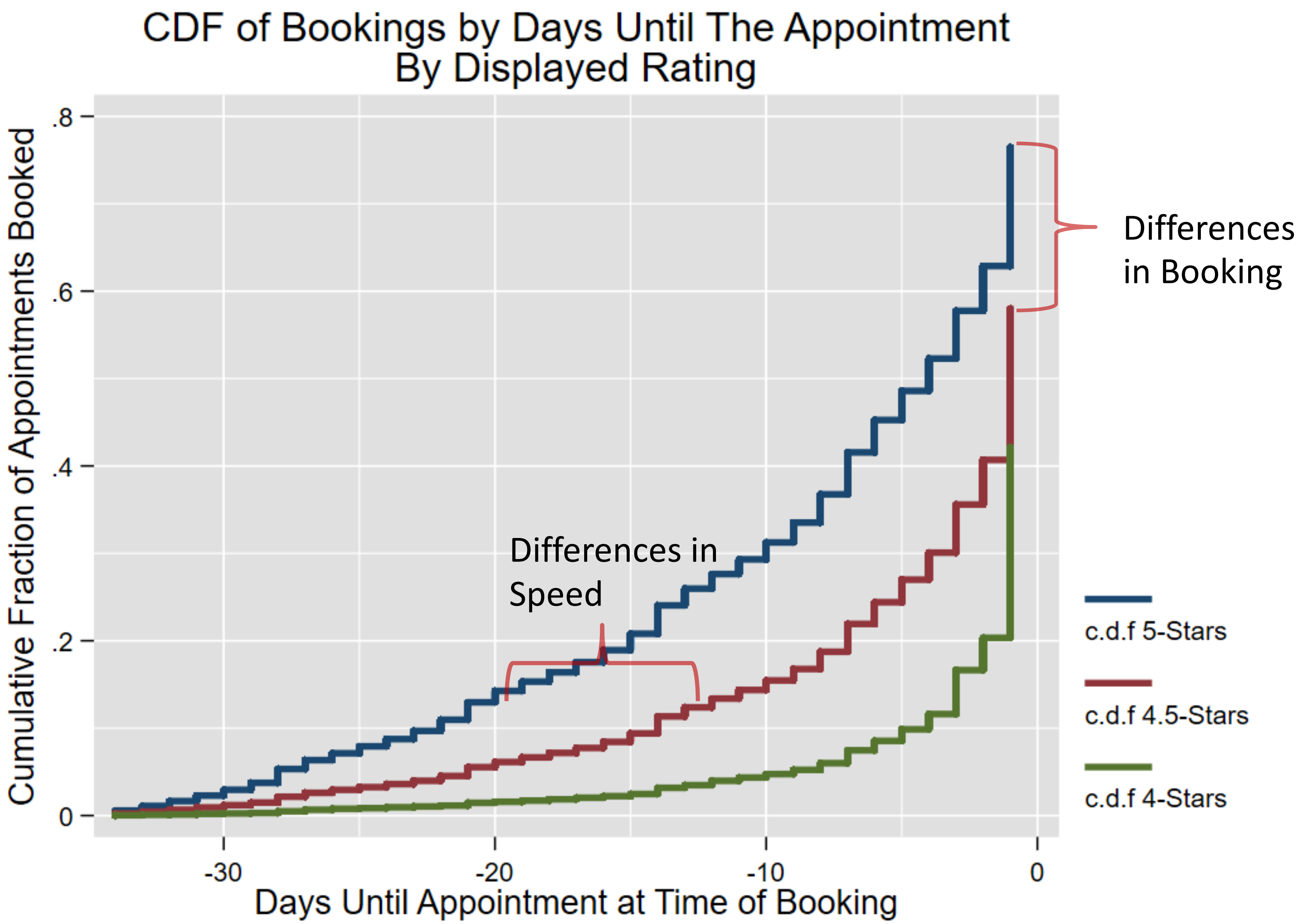
Booking Likelihood by Rating



Sample: 2/24/2016-4/17/2017, primary care, min 8 ratings, with apts offered during business hours
Controls: None

Preliminary Results: Booking by Time (CDF Comparison)

- **The vertical difference:** difference in percent of appointments booked at a given number of days in advice.
- **The horizontal differences:** The difference in how many days in advanced the same percent of appointments were booked.



Empirical Strategy: Primary Specification

Observation level: Physician-week

Dependent variable: Weekly patient volume based

- Inverse Hyperbolic Sign (IHS) of bookings
- IHS vacant appointments (alternative)

Running variable: Average overall rating

Covariates: Market-week, IHS(offered appointments), number of location, appt length and type no. reviews, hospital affiliation

Methods:

- Asymmetric data-driven MSE-optimal bandwidth selectors
- Triangular kernel
- Mass point adjustments
- Bias-corrected RD estimates with robust variance estimator
- Cluster-robust nearest neighbor variance estimation clustered on physician (panel data)

Results

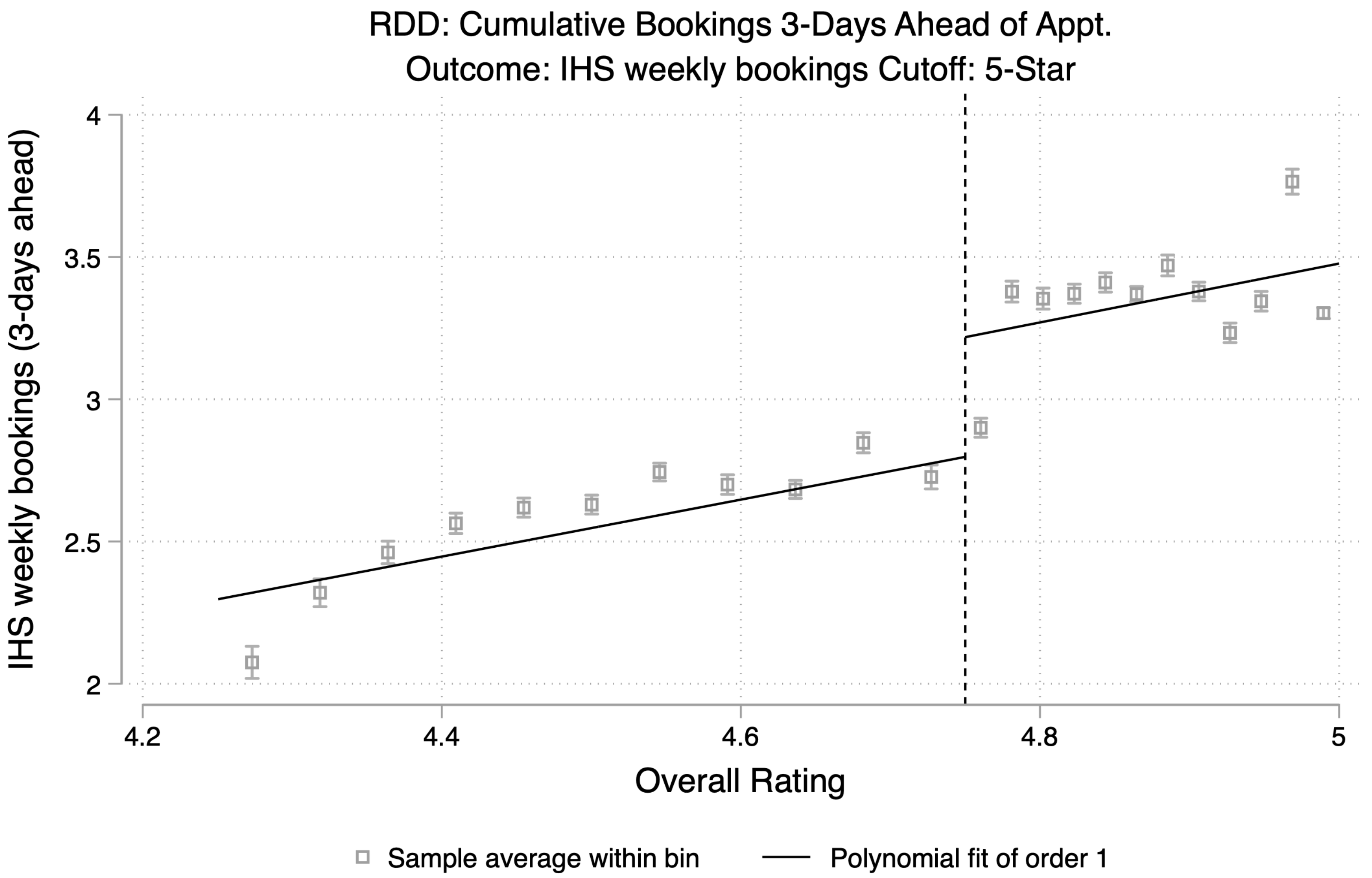
PRIMARY SPECIFICATION

ALTERNATIVE SPECIFICATIONS

DIFFERENTIAL IMPACTS

ROBUSTNESS CHECKS

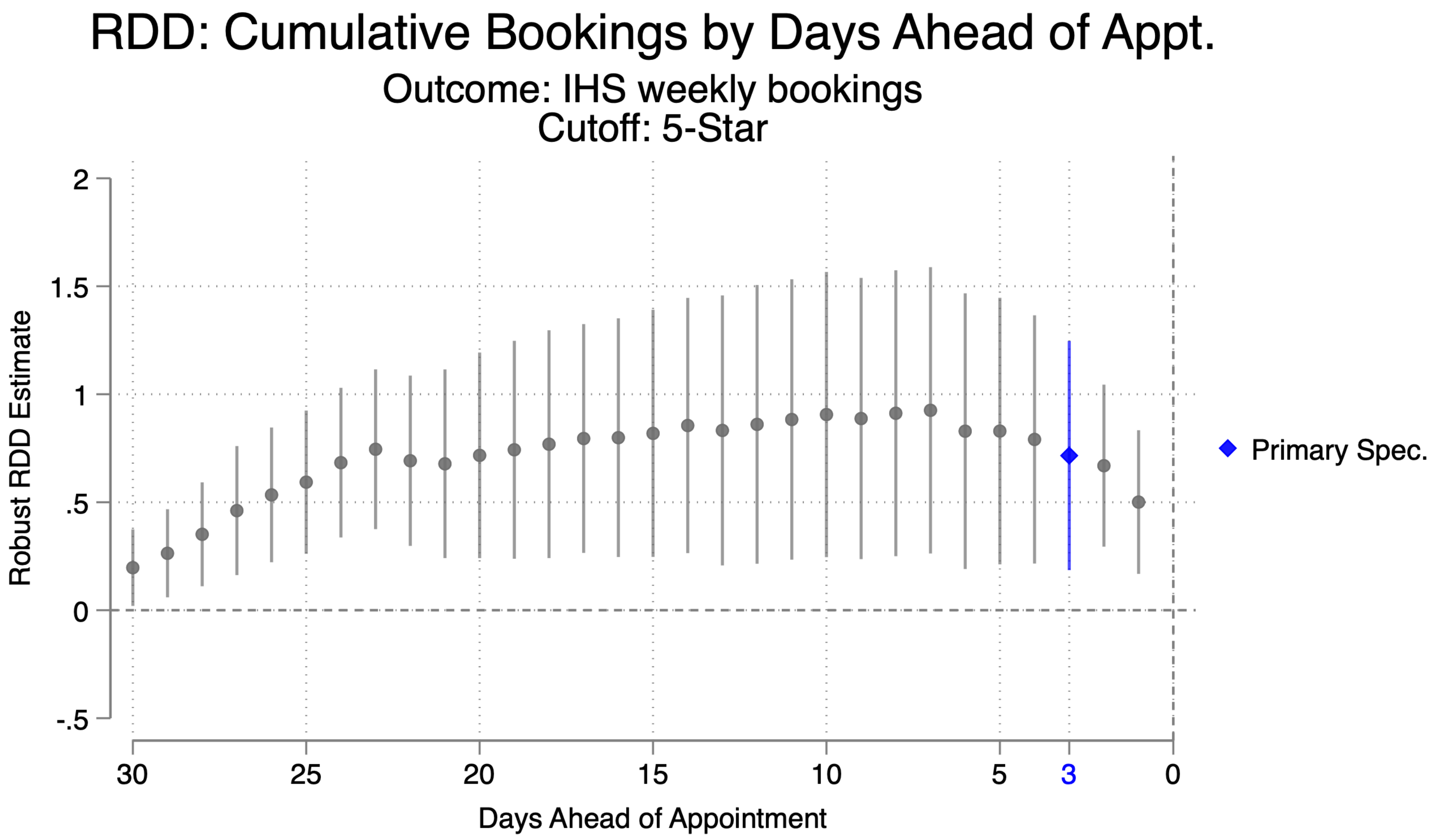
Impact of Ratings on Patient Volume: Approx. Doubling of Bookings



Method	(1) IHS(Weekly Bookings)
Conventional	0.686***
	(0.228)
Bias-corrected	0.716***
	(0.228)
Robust	0.716***
	(0.271)
Observations	54263
Controls	X

Details

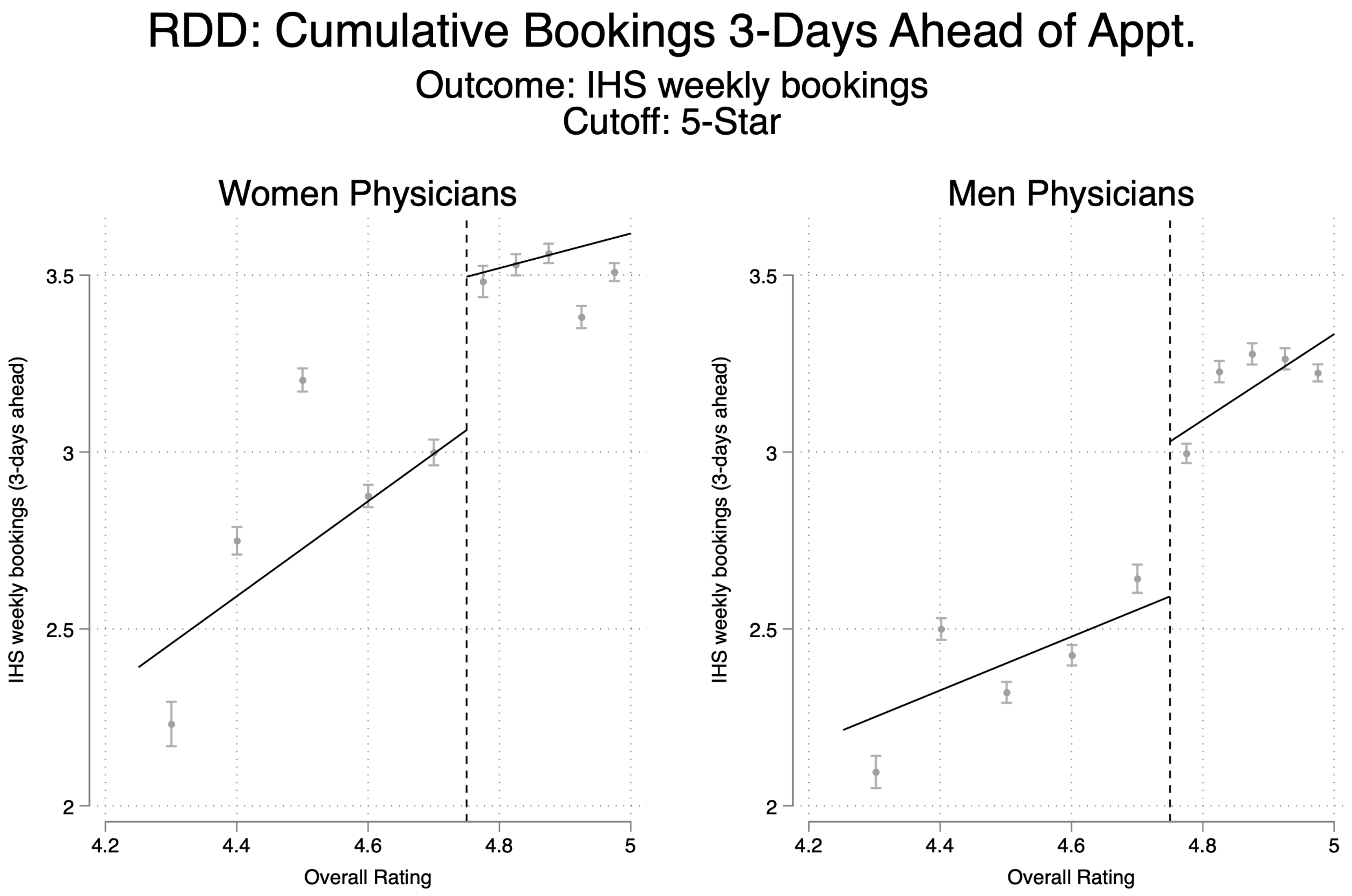
Impact of Ratings on Patient Volume: Approx. Doubling of Bookings



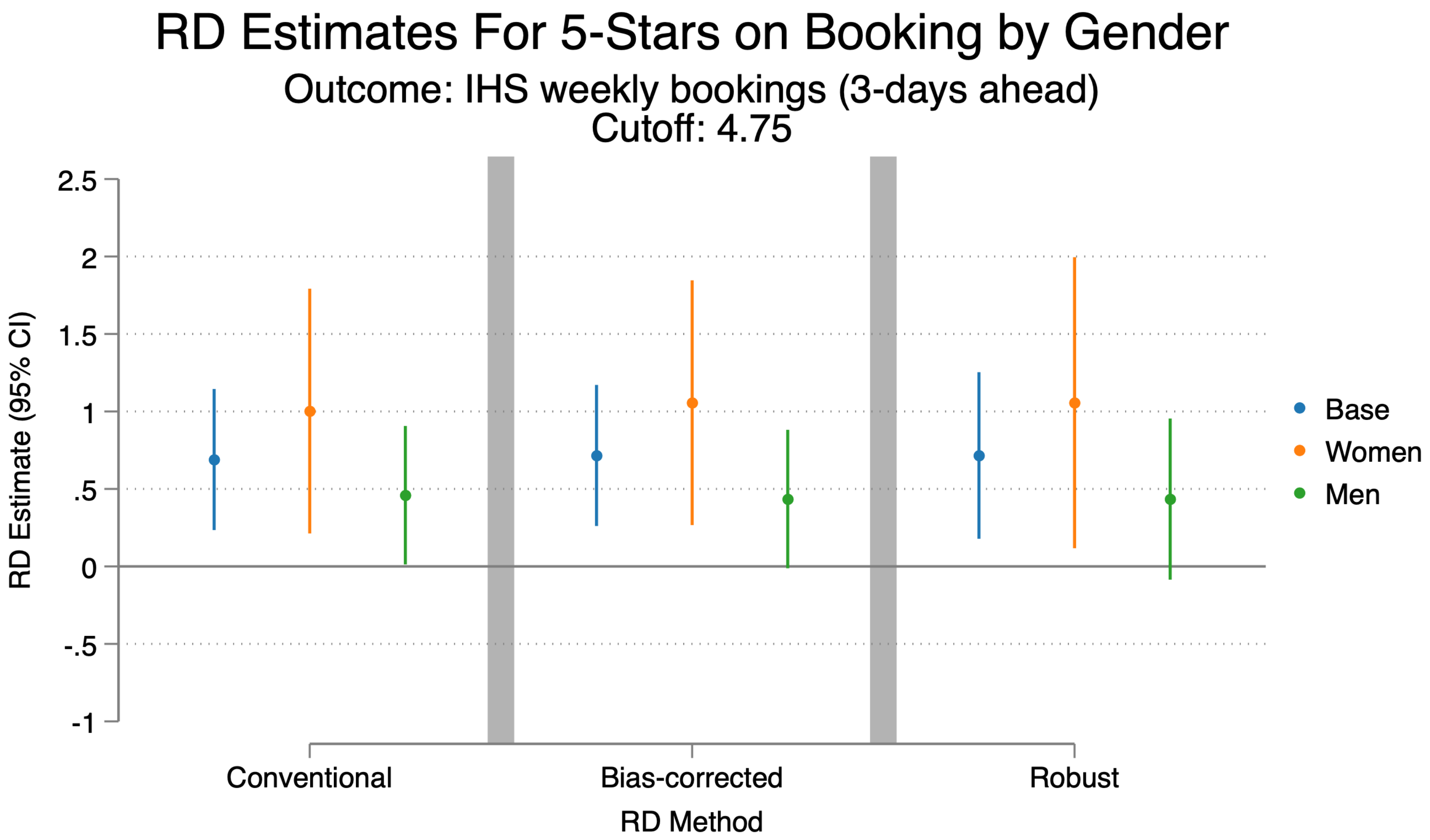
Controls: market-week, no. locations, appt length, appt type, no. reviews, hospital affiliation
Sample: 2/24/2016-4/17/2017, primary care, min 8 ratings, with apts offered during business hours 21 days in advance, stable ratings, excludes profiles with >4 rating removals
Specification: data-driven asymetric bandwidth, triangular kernal, NNcluster on physician

Vacancy Results

RD Plot by Physician Gender: Women Have More Bookings at 4.5 and 5 Stars



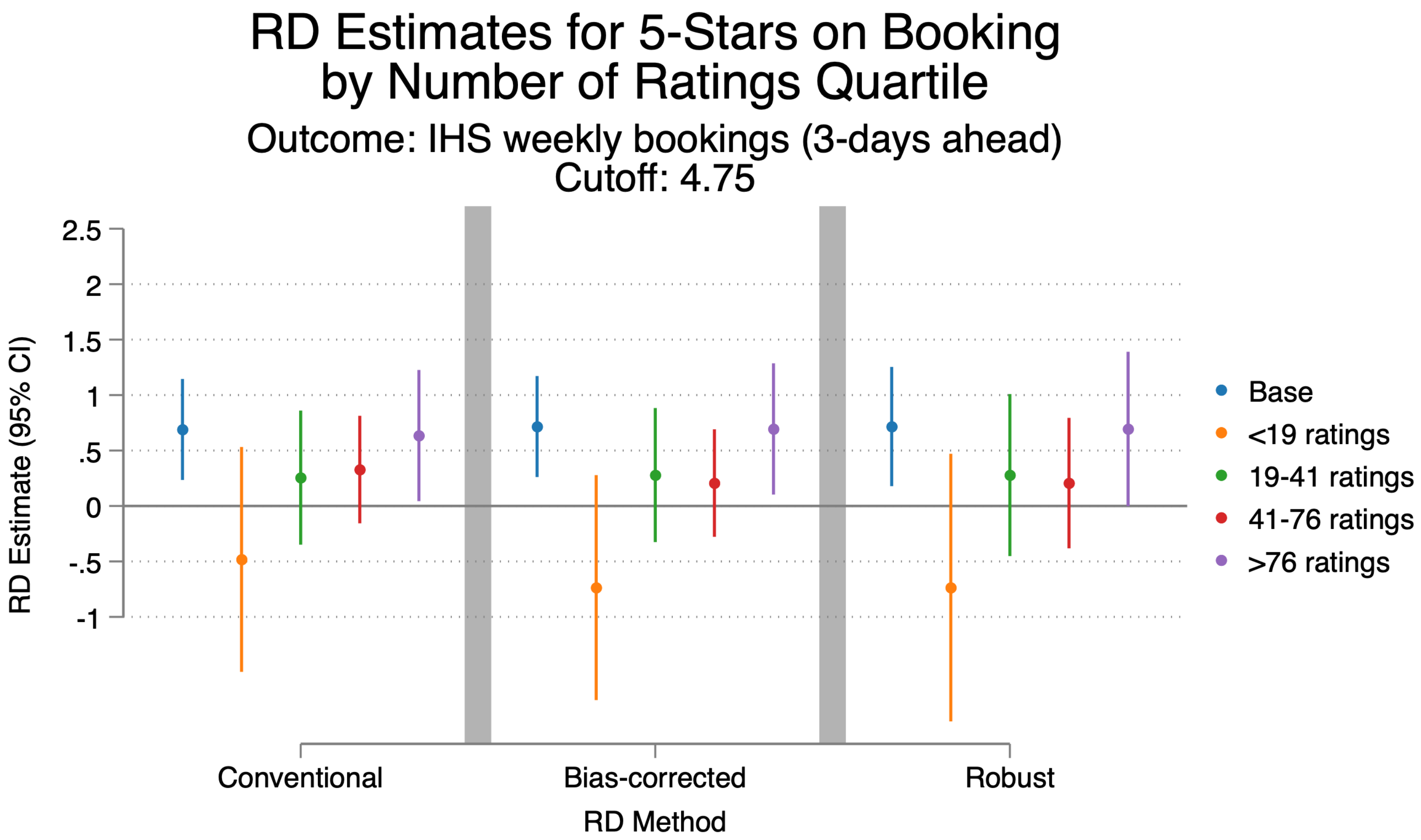
Differential Impact of by Gender: Effect Greatest for Women



Controls: market-week, no. locations, appt length, appt type, no. reviews, hospital affiliation, offered appts
Sample: 2/24/2016-4/17/2017, primary care, min 8 ratings, with apts offered during business hours 21 days in advance, stable ratings, excludes profiles with >4 rating removals
Specification: data-driven asymetric bandwidth, triangular kernal, NNcluster on physician

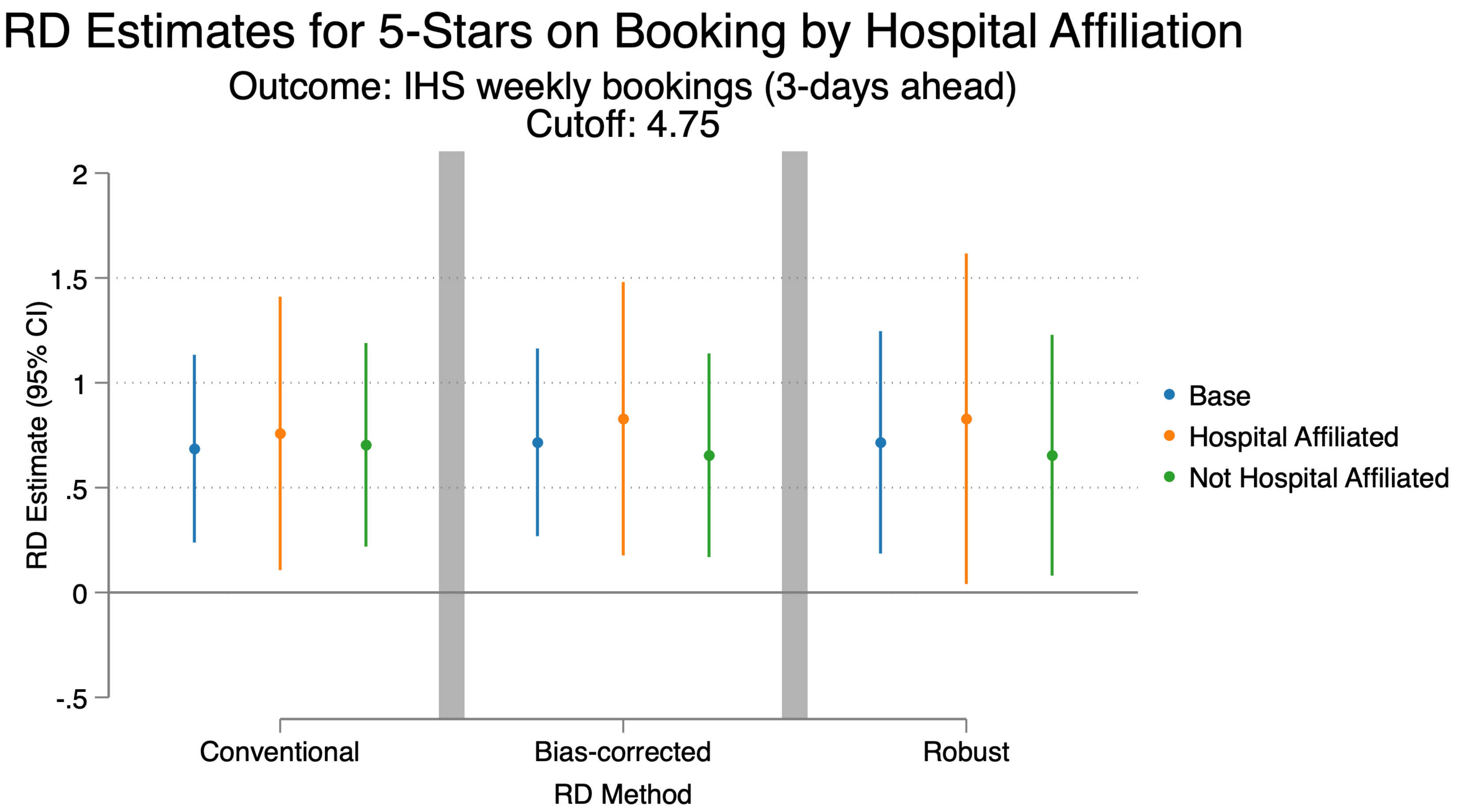
Details

Differential Impact of by Ratings: Effect Increases with No. Ratings



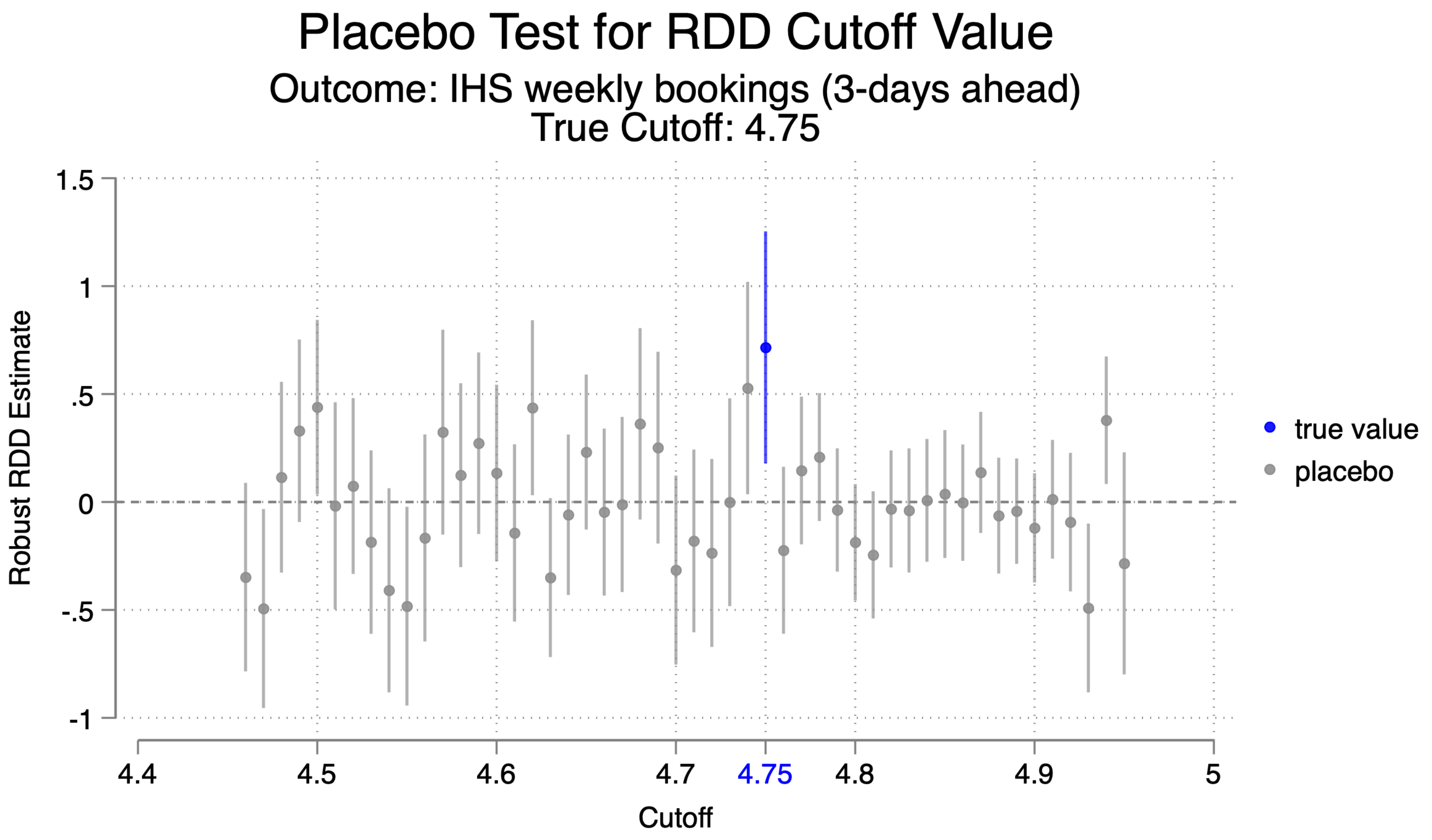
Controls: market-week, no. locations, appt length, appt type, no. reviews, hospital affiliation, offered appts
Sample: 2/24/2016-4/17/2017, primary care, with apts offered during business hours 21 days in advance,
stable ratings, excludes profiles with >4 rating removals
Specification: data-driven asymetric bandwidth, triangular kernal, NNcluster on physician

Differential Impact of by Hospital Affiliation: Similar Effects



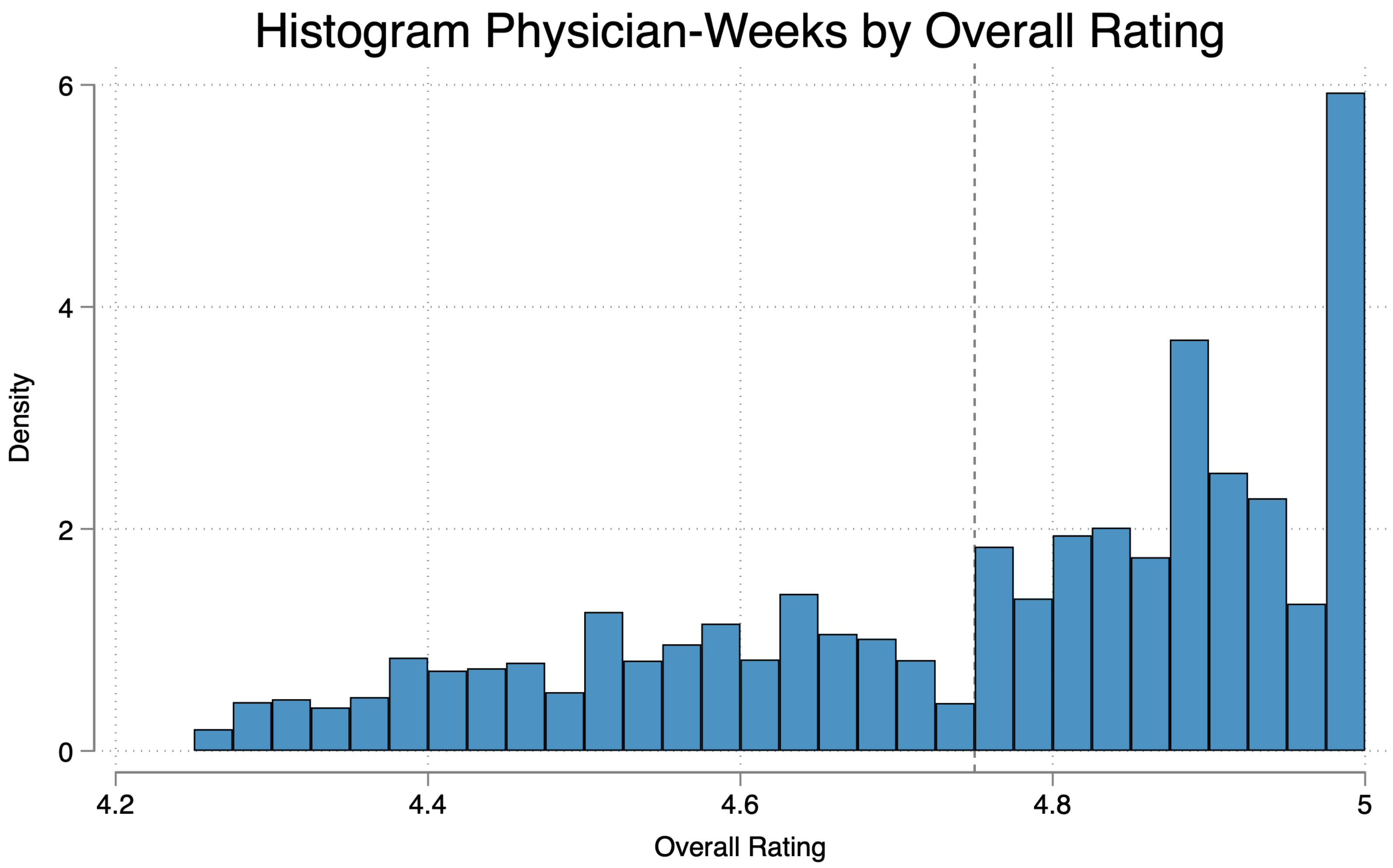
Controls: market-week, no. locations, appt length, appt type, no. reviews, hospital affiliation, offered appts
Sample: 2/24/2016-4/17/2017, primary care, with apts offered during business hours 21 days in advance,
stable ratings, excludes profiles with >4 rating removals
Specification: data-driven asymetric bandwidth, triangular kernal, NNcluster on physician

Robustness: Placebo Test of Main Result



Controls: market-week, no. locations, appt length, appt type, no. reviews, hospital affiliation
Sample: 2/24/2016-4/17/2017, primary care, min 8 ratings, with apts offered during business hours 21 days in advance, stable ratings, excludes profiles with >4 rating removals
Specification: data-driven asymetric bandwidth, triangular kernal, NNcluster on physician

Robustness: Visible Bunching Above Cutoff



Sample: 2/24/2016-4/17/2017, primary care, min 8 ratings, with apts offered during business hours 21 days in advance,

Conclusion

SUMMARY OF RESULTS

DISCUSSION & MECHANISMS

Summary of Findings

Descriptive Evidence

- Booking likelihood: More likely to be booked
- Booking speed: Booked further in advance

Regression Discontinuity at 5-Stars

- Patient volume via bookings: Approx. twice as many bookings
- Patient volume via vacancies: Approx. half as many vacancies

Differential Impact

- Physician gender: Effect greatest for women physicians
- Number of ratings: Effect increases with number of ratings
- Hospital affiliation: No significant difference

Robustness

- Placebo tests: Effect greatest at true cutoff
- Rating manipulation: Bunching above cutoff

Discussion: Potential Mechanisms

Gender: Effect greatest for women physicians

- But not closing a gender gap
- Correlated preferences
  - Ex: Gender, rating, and wait time
- Platform recommendation system

Ratings: Effect increases with number of ratings

- Consistent with Bayesian learning

Hospital Affiliation: Similar Effects

Next Steps

Extend to analysis to other cutoffs

Differential impact by apparent race and age

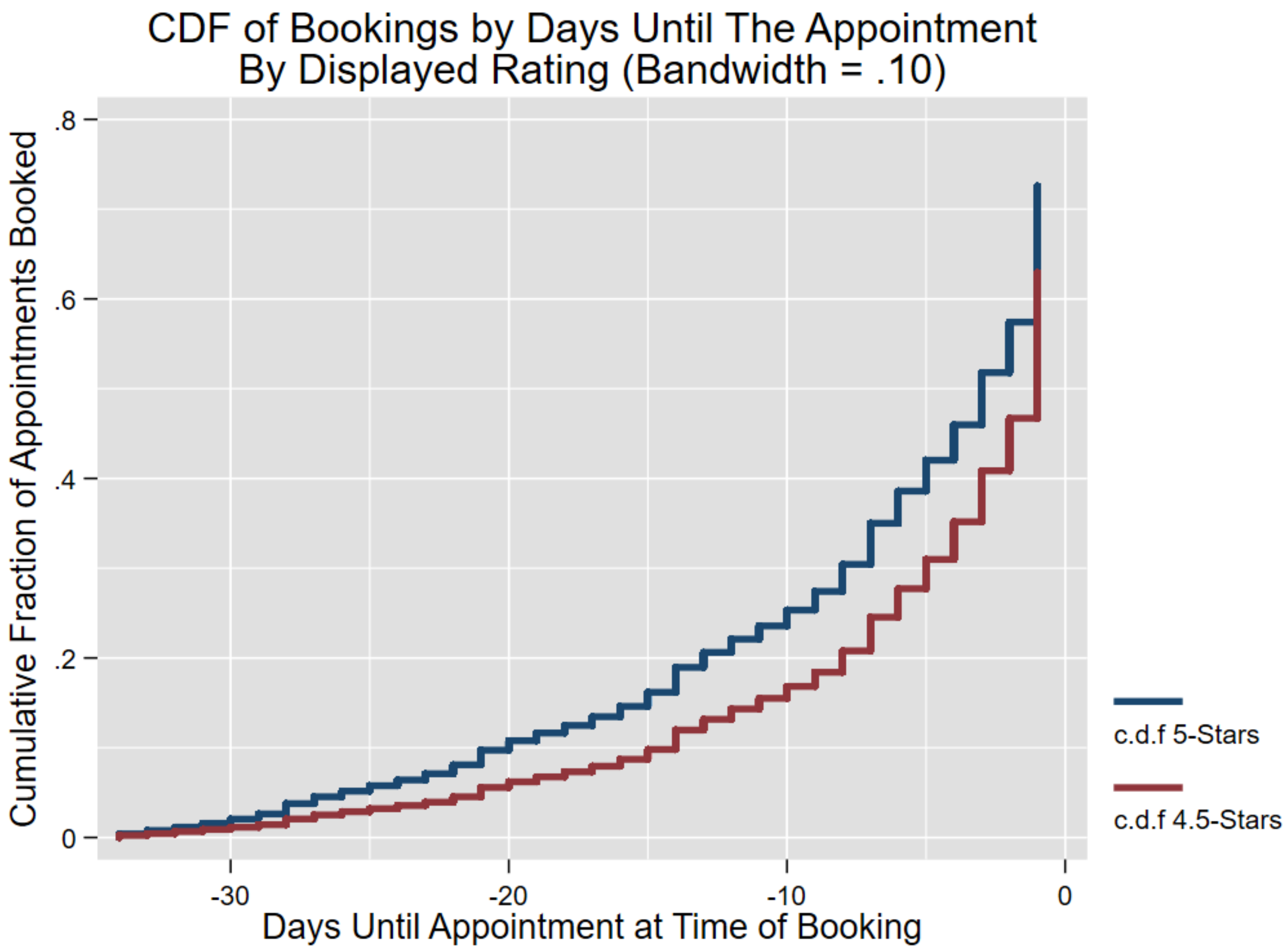
Robustness

- Mass at cutoff
  - “Donut” regression discontinuity
- Covariate balance

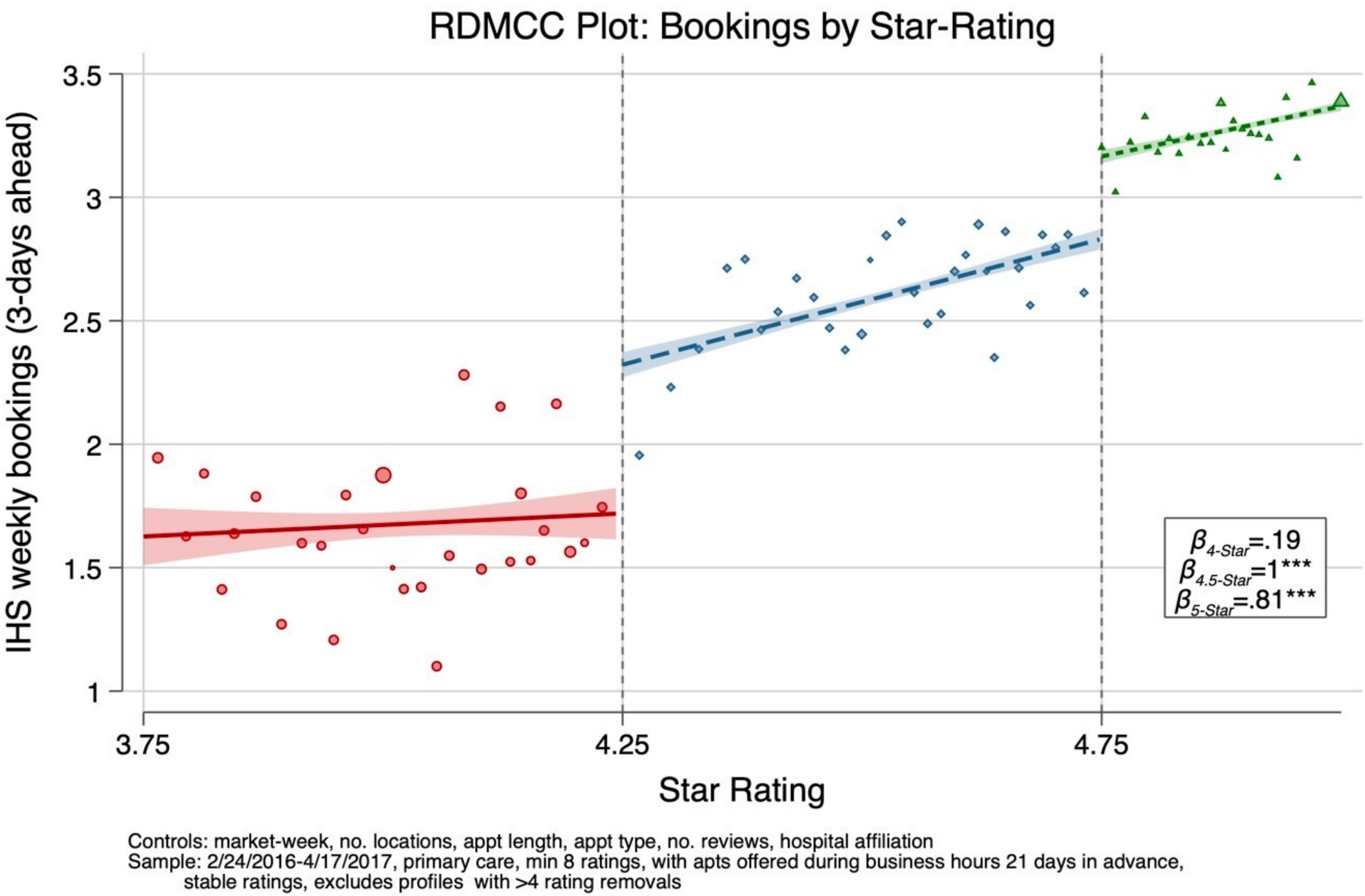
Appendix

Booking by Time (CDF Comparison)

Use a pilot bandwidth of .1 to compare these cdfs of physicians just above and just below the 4.75 threshold to have five stars.

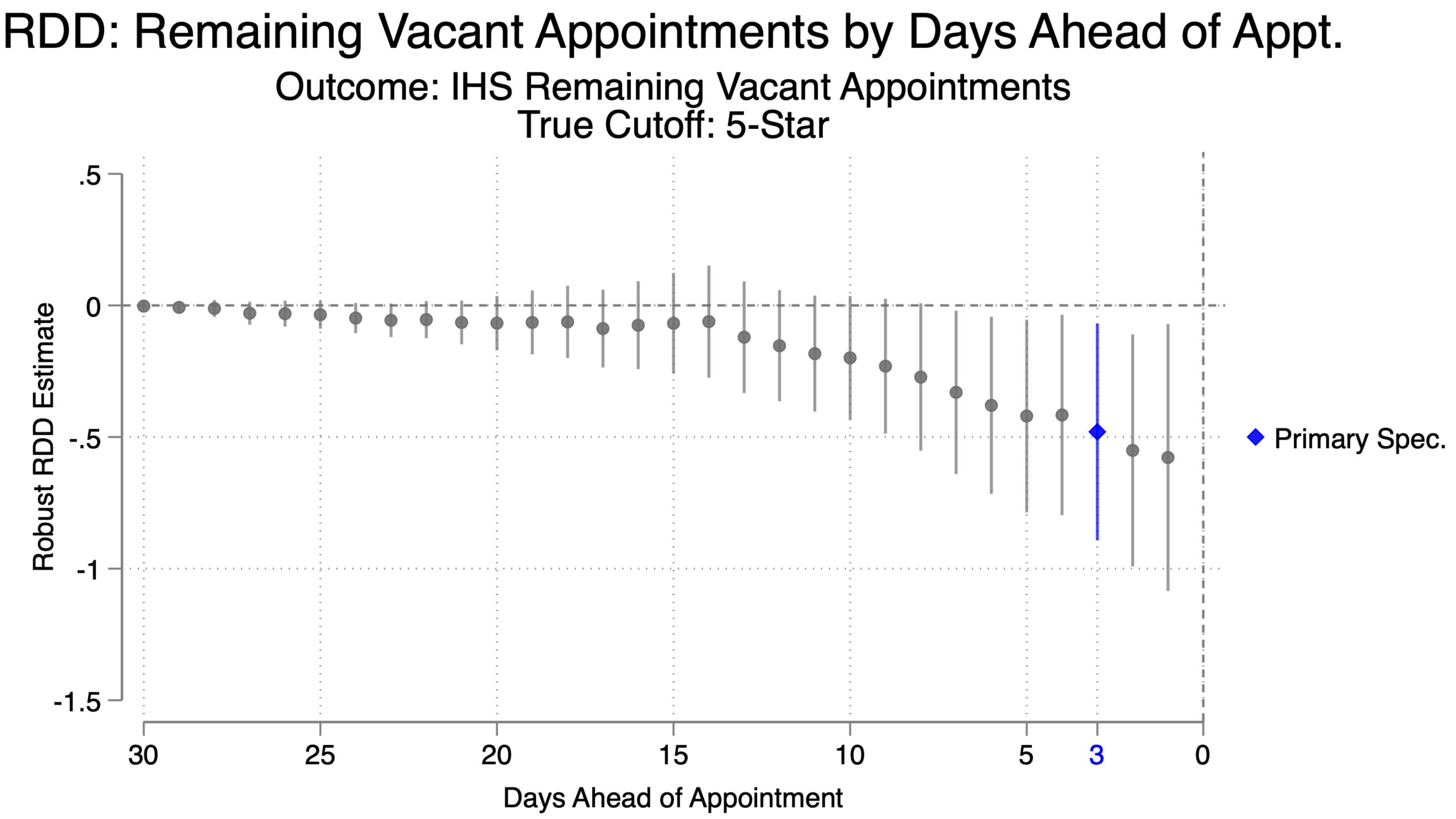


Patient Volume by Star Rating: 4, 4.5, and 5-Stars



5-star Results

Impact of Ratings on Patient Volume: Fewer Vacancies



Controls: market-week, no. locations, appt length, appt type, no. reviews, hospital affiliation
Sample: 2/24/2016-4/17/2017, primary care, min 8 ratings, with apts offered during business hours 21 days in advance, stable ratings, excludes profiles with >4 rating removals
Specification: data-driven asymetric bandwidth, triangular kernal, NNcluster on physician

Booking Results

Booked Appointments by Page Rank Proxy

We are also interested in platform mechanics. Here, we take advantage of the fact that page rank is a function of availability.

Observation Level: Physician-day

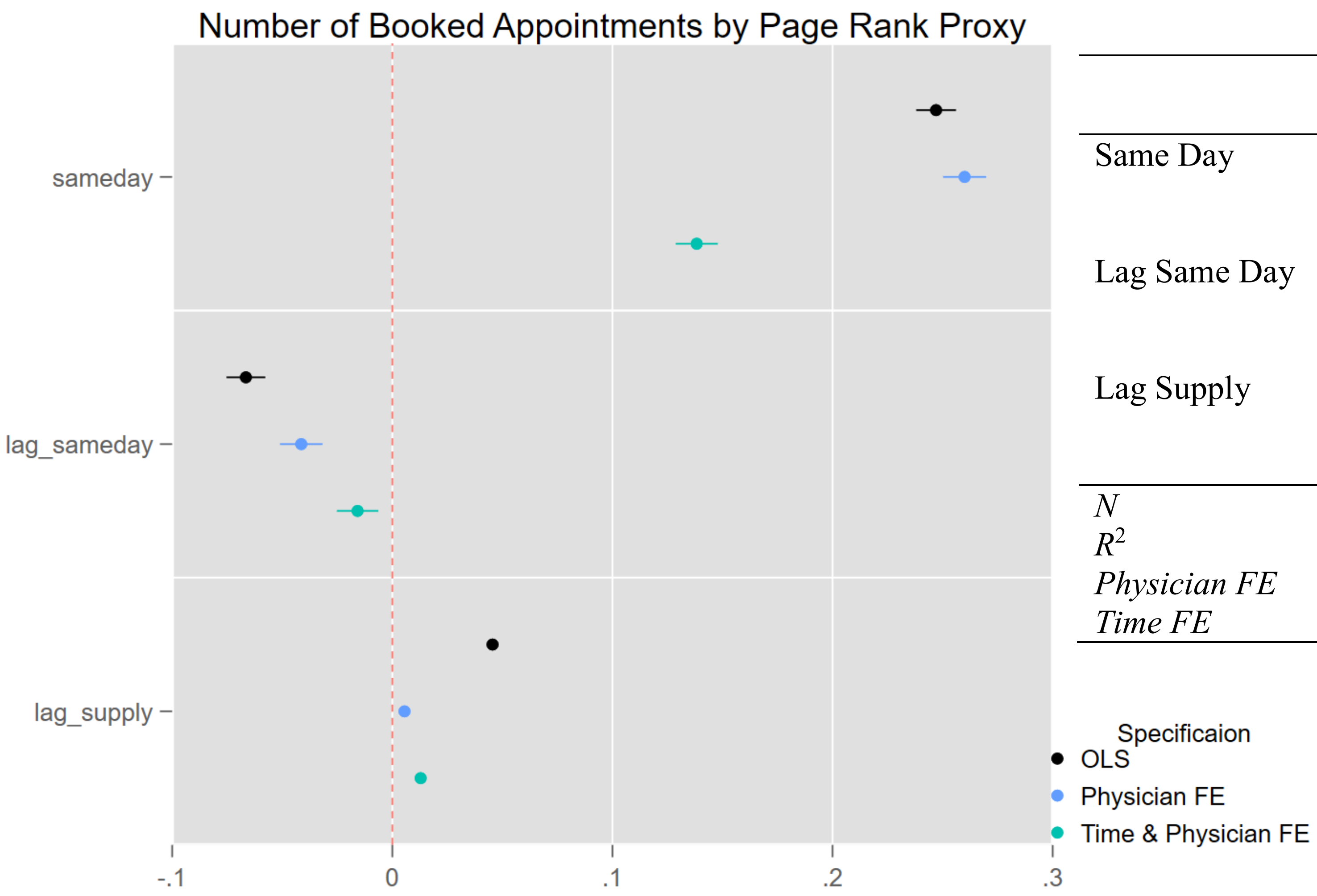
Dependent Variable: Count of appointments booked that day

Variables of Interest: Proxy for page rank with the number of same day appointments available, and the lag of same day appointments available.

Intuition: If page rank has no effect, we might expect these coefficients to be negative. A positive coefficient suggest page rank is indeed important.

Controls: Number of available appointments, physician FE, and time FE.

Booked Appointments by Page Rank Proxy



	(1) booked	(2) booked	(3) booked
Same Day	0.247***	0.260***	0.138***
	(53.49)	(52.12)	(28.31)
Lag Same Day	-0.0665***	-0.0413***	-0.0158**
	(-14.78)	(-8.38)	(-3.28)
Lag Supply	0.0455***	0.00554***	0.0129***
	(45.60)	(5.56)	(11.76)
N	487408	487408	487408
R^2	0.012	0.006	0.186
Physician FE		X	X
Time FE			X