NBER WORKING PAPER SERIES

DOES DIVERSITY MATTER FOR HEALTH? EXPERIMENTAL EVIDENCE FROM OAKLAND

Marcella Alsan
Owen Garrick
Grant C. Graziani

Working Paper 24787
http://www.nber.org/papers/w24787

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2018

We thank Esther Duflo, Pascaline Dupas and the J-PAL Board and Reviewers who provided important feedback that improved the design and implementation of the experiment; Ran Abramitzky, Ned Augenblick, Jeremy Bulow, Kate Casey, Arun Chandrasekhar, Raj Chetty, Stefano DellaVigna, Mark Duggan, Karen Eggleston, Erica Field, Matthew Gentzkow, Gopi Shah Goda, Susan Godlonton, Jessica Goldberg, Michael Greenstone, Guido Imbens, Seema Jayachandran, Damon Jones, Supreet Kaur, Melanie Morten, Maria Polyakova, Matthew Rabin, Al Roth, Kosali Simon, Ebonya Washington, Crystal Yang and seminar participants at UC Berkeley, Stanford, Cornell, MIT, UCLA, UCSB, Harvard Kennedy School, University of Chicago, and IFPRI for their helpful comments. Javarcia Ivory, Matin Mirramezani, Edna Idna, Anlu Xing and especially Morgan Foy provided excellent research assistance. We thank the administration at Stanford, SIEPR, and J-PAL particularly Lesley Chang, Rhonda McClinton-Brown, Dr. Mark Cullen, Dr. Douglas K. Owens, Ann Dohn, Ashima Goel, Atty. Ann James, Atty. Tina Dobleman, Nancy Lonhart, Jason Bauman, Sophie Shank, James Turitto, Florian Grosset and Luke Sonnet. A working paper version of this paper was submitted for pre-publication re-analysis to the Abdul Latif Jameel Poverty Action Lab (J-PAL), where a code replication exercise was conducted on the analysis. For more details about J-PAL's replication work visit https://osf.io/be432/. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Does Diversity Matter for Health? Experimental Evidence from Oakland
Marcella Alsan, Owen Garrick, and Grant C. Graziani
NBER Working Paper No. 24787
June 2018, Revised August 2019
JEL No. C93,I12,I14

ABSTRACT

We study the effect of physician workforce diversity on the demand for preventive care among African-American men. In an experiment in Oakland, California, we randomize black men to black or non-black male medical doctors. We use a two-stage design, measuring decisions before (pre-consultation) and after (post-consultation) meeting their assigned doctor. Subjects select a similar number of preventives in the preconsultation stage, but are much more likely to select every preventive service, particularly invasive services, once meeting with a racially concordant doctor. Our findings suggest black doctors could reduce the black-white male gap in cardiovascular mortality by 19%.

Marcella Alsan
Stanford Medical School
Center for Health Policy/PCOR
117 Encina Commons, Room 218
Stanford, CA 94304
and NBER
malsan@stanford.edu

Owen Garrick
Bridge Clinical Research
333 Hegenberger Road
Suite 208
Oakland, CA 94621
owen.garrick@bridgeclinical.com

Grant C. Graziani
University of California, Berkeley
530 Evans Hall
Berkeley, CA 94720
gcgraziani@berkeley.edu

A data appendix is available at http://www.nber.org/data-appendix/w24787
A randomized controlled trials registry entry is available at https://www.socialscienceregistry.org/trials/2497

I. Introduction

African-American men have the lowest life expectancy of any major demographic group in the United States (Arias, Heron, and Xu 2017) and live on average 4.5 fewer years than non-Hispanic white men (Murphy et al. 2017). Reasons for this disparity are multifactorial and include lack of health insurance, lower socioeconomic status, and structural racism (IOM 2003). Approximately 60% of the difference in life expectancy between black and white men is attributable to chronic diseases, which are amenable to primary or secondary prevention (Harper, Rushani, and Kaufman 2012; Silber et al. 2014). Some examples are poorly controlled hypertension (associated with stroke and myocardial infarction), diabetes (associated with end organ disease including kidney failure), and delayed diagnosis of cancers. These data suggest at least part of the mortality disparity is related to underutilized preventive healthcare services.

One frequently discussed policy prescription put forth by the Institute of Medicine (IOM) as well as the National Medical Association (NMA), the Association of American Medical Colleges (AAMC), and the American Medical Association (AMA) to address racial health disparities is to diversify the healthcare profession by increasing the number of underrepresented minorities.[1] Blacks comprise approximately 13% of the U.S. population but only 4% of physicians and less than 7% of recent medical school graduates (AAMC 2014, AAMC 2016). Evidence on whether patient and physician racial concordance improves satisfaction and health outcomes is mixed, perhaps due to methodological or contextual differences. Recent studies have found that gender- or race-match between doctors and patients in a hospital setting reduces mortality (Greenwood, Carnahan, Huang 2018; Hill, Jones, and Woodworth 2018) yet in the outpatient setting, the results are less clear. Meghani et al. (2009) perform a meta-analysis of thirty observational studies in public health and medicine concerning four racial and ethnic groups. They conclude that the evidence in favor of patient-doctor concordance in medical care is inconclusive and recommend additional research. We advance this literature by providing experimental evidence on whether and to what extent diversity in the physician workforce improves medical decisions and outcomes among minority populations.

Our study builds upon several findings in economics. First, randomized trials in development economics have demonstrated puzzlingly low demand for high return preventive healthcare services among low-income populations (for a review see Dupas 2011; Banerjee and Duflo 2011, Chapter 3). Similar patterns are found in the U.S. — compared to non-

[1] See "Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care" (IOM 2003); "Addressing Racial Disparities in Health Care: A Targeted Action Plan for Academic Medical Centers" (AAMC 2009); "Major Minority Physician Associations Come Together" (NMA 2018); and "Reducing Disparities in Health Care" (AMA 2018).

Hispanic white men, African-American men are six percentage points less likely to visit the doctor and eight percentage points less likely to report receipt of the flu shot; insurance and education do not fully explain these gaps (Blewett et al. 2018a).

Many factors likely contribute to this puzzle including lack of information, inadequate or low quality healthcare supply, and misperceptions about the etiology of disease. Given the prominent history of neglect and exploitation of disadvantaged populations by health authorities, mistrust of the medical establishment is sometimes invoked as a contributing factor. Evidence consistent with historical abuse dampening demand and increasing mistrust has been found specifically among African-American men in the immediate aftermath of the U.S. Public Health Service syphilis experiment in Tuskegee, Alabama, (Alsan and Wanamaker 2018) and persisting decades after colonial medical campaigns in Central Africa (Lowes and Montero 2018). Recent studies in public health demonstrate that African-American men continue to score higher on medical mistrust measures than other groups (Kinlock et al. 2017, Nanna et al. 2018, Hammond et al. 2010).

Second, contributions in cultural economics have highlighted how norms of behavior are influenced by social identity (Akerlof and Kranton 2000; Benjamin, Choi, and Strickland 2010). Most notably, Tabellini (2008) shows how cooperation can be sustained in a one-shot prisoner's dilemma among agents who perceive a non-economic benefit from cooperating with those closer in social distance. Third, natural experiments in labor and education have underscored how diversity, or lack thereof, may be particularly relevant in asymmetrical power relationships. For instance, Glover, Pallais, and Pariente (2017) find that minority workers exert less on the job effort in grocery stores with biased majority managers. Additional evidence on how diveristy affects hiring and job performance can be found in Stoll, Raphael, and Holzer (2004); Giuliano, Levine, and Leonard (2009); Hjort (2014); and Bertrand et al. (2018). A spate of studies has found that same race or same gender teachers are positively correlated with grades and career path, potentially through a role model effect (Ehrenberg, Goldhaber, and Brewer (1995); Dee (2004); Dee (2005); Bettinger and Long (2005); Carrell, Page, and West (2010); Fairlee, Hoffmann, and Oreopoulos (2014); and Lusher, Campbell, and Carrell (2018)).

There are several ways in which racial diversity could play a role in medicine, specifically as it relates to the patient-doctor relationship. Taste-based discrimination (Becker 1957) on the part of the patient or doctor could imply that individuals are averse to interacting with those who do not share their racial background. On the other hand, internalized racism, or negative beliefs about one's racial group, could lead to the opposite phenomenon. Third, a common racial background might facilitate communication — a critical component of clinical care as both patient and physician have potentially life-saving information to exchange.

Fourth, and not mutually exclusive, concordance may foster trust leading to cooperation (i.e. compliance with doctors' advice or willingness to engage). As noted by Arrow (1963), "...it is a commonplace that the physician-patient relation affects the quality of the medical care product."

In this study, we examine whether doctor race affects the demand for preventive care among African-American men. We induce exogenous variation by randomly assigning subjects to black and non-black doctors.[2] Our experiment was conducted in Oakland, California, where we recruited over 1,300 black men from about 20 local barbershops and two flea markets. At these recruitment sites, subjects filled out baseline questionnaires and received a coupon for a free health screening. To facilitate our experiment, we set up a clinic to provide preventive services to the subjects. The clinic was staffed with fourteen black and non-black male doctors from the Bay Area as well as a diverse team of receptionists. Doctors and staff were told the study was designed to improve the take-up of preventive care among black men in Oakland, but not specifically informed about the role of doctor race. Subjects learned of their (randomly) assigned doctor via tablet in the privacy of their own patient room.

The experiment proceeded in two stages and cross-randomized doctor race with incentives for the flu vaccine at the individual level. In the pre-consultation stage, patients were introduced to their doctor via the tablet by way of text and photo, both standardized as described in Section II. Subjects were then provided the opportunity to select which, if any, of the four advertised cardiovascular screening services they would like to receive. These services included body mass index (BMI) measurement, blood pressure measurement, diabetes screening, and cholesterol screening. The last two tests required a blood sample, and subjects were made aware of this feature. After making their selections for cardiovascular screening, subjects were informed they could also elect to receive a flu shot, administered by their assigned doctor. For subjects randomized to receive a flu incentive to encourage vaccine selection, the incentive amount was also listed. We conjectured that if subjects disliked doctors who did not share their racial background, those randomly assigned to non-black doctors would, on average, demand fewer preventives simply based on the tablet photo.

In the second stage, subjects met their assigned doctor in person. We refer to this stage throughout the paper as post-consultation (since decisions occur *after* interacting with their doctor). Subjects could revise their choice of preventives during this stage, after which the doctor administered the selected services. We therefore measure how black vs. non-black doctors affect demand between the pre- and post-consultation stages, which we refer to as

[2] Throughout the paper, we use "black" to refer to African-Americans and "non-black" to refer to Caucasian and Asian-Americans.

the *delta*, since it represents the change in selected services across the two periods. These are two choice events occurring after randomization and both represent experimental outcomes. Following the patient-doctor interaction, subjects filled out feedback forms and exited the clinic.

It is important to note that the study provided only *preventive* (i.e. care recommended during a state of relatively good health to avoid future illness – such as screenings and immunizations) as opposed to *curative* (i.e. care needed during a state of illness to restore health) interventions. Individuals often have imperfect knowledge regarding the health benefits of prevention, perhaps because they have been misinformed, never informed, or informed by someone they don't trust, which can dampen demand.[3] Hence the role of study doctors was limited to information provision on the benefits of receiving care even when not feeling sick and then providing those chosen.

Approximately half of the subjects we recruited from the community visited our clinic. Those who presented were negatively selected relative to those who completed the barbershop survey but did not come to the clinic. Subjects who redeemed the clinic coupon were 13 percentage points more likely to be unemployed (compared to 18% among non-participants) and 19 percentage points more likely to have a high school education or less (compared to 44% among non-participants). In terms of health and healthcare utilization, they had significantly lower self-reported health, were less likely to have a primary care physician, and more likely to have visited the emergency room.

Once at the clinic, subjects randomly assigned to a black doctor elected to receive the same number of preventive services as those assigned to a non-black doctor in the pre-consultation stage. In sharp contrast, we find that subjects assigned to black doctors, upon interacting with their doctor, are 18 percentage points more likely to take up preventives relative to those assigned to non-black doctors. These findings are robust to corrections for correlated error structures within doctor; the inclusion of fixed effects for clinic date, field staff, and recruitment location; as well as various permutations of the study doctors, including dropping the "best" black and "worst" non-black doctor.

Why would black male subjects randomly assigned to black male doctors elect to receive more services upon interacting with them? We provide several pieces of evidence that better communication between black subjects and black doctors explains our results, and discuss alternative mechanisms below. First, in our controlled study environment, the role of the doctor was circumscribed to informing subjects about the benefits of preventive services, and then providing those chosen. Second, for non-invasive tests (those that do not require blood or an injection), both non-black and black doctors shifted out demand in the post-

[3]According to the CDC, up to 40% of annual deaths in the U.S. are deemed preventable (CDC 2014).

consultation stage relative to the pre-consultation stage, though the effect was larger for the latter. Yet, for invasive tests, those that carry more risk and thus likely require more trust in the person providing the service, only subjects assigned to black doctors responded: increasing their take-up of diabetes and cholesterol screenings by 20 and 26 percentage points (49% and 71%), respectively. Third, subjects are more likely to talk to their assigned doctor about health issues if the assigned doctor is African-American, a result which is particularly strong among those who obtain an invasive exam.

The experimental findings highlighting improved communication for black male patients paired with black male doctors are consistent with those collected in a non-experimental manner. We surveyed 1,490 black and white adult males who matched our sample in terms of educational attainment. The respondents were asked to select a doctor of a particular race based on accessibility, quality, and communication. With respect to quality (i.e. which doctor is the most qualified) black and white respondents both selected doctors of the same race about 50% of the time. However, for questions regarding communication, in particular which doctor would understand your concerns, the proportion of respondents choosing doctors of their own racial background jumped to nearly 65% for blacks and 70% for whites.

An alternative interpretation of our results is that the estimated treatment effect is picking up an attribute correlated with the race of the doctor in our sample and which affects the outcome of interest. A prominent candidate for a hard-to-measure characteristic that may correlate with doctor race is quality.[4] The non-experimental findings cited above demonstrate black male respondents believe that non-black doctors are as qualified as black doctors. Yet, actual doctor quality within the context of our study could vary.

We address the possibility of differential quality across doctor race in the study setting in several ways. First, doctors were balanced on observables in age, experience, and medical school rank: characteristics we collected from their resumes. Moreover, all of the non-black doctors, but only 67% of black doctors, practiced internal medicine. In addition, we created a survey for the study doctors designed to assess their typical patient characteristics, their persuasiveness, and their current medical knowledge using questions typically found on medical credentialing exams. Interestingly, the non-black study doctors were more likely to state their patients comply with medical advice and that they are able to persuade both white *and* black adult male patients to take up testing they had initially refused.

If black doctors were higher quality than non-black doctors we might have expected them to be rated higher on feedback forms, yet black and non-black doctors were rated equally

[4]This could arise if, for example, black doctors are more qualified than non-black doctors in the population and we failed to draw our sample from an area of overlapping support — or if the distributions were similar, but we drew from different tails.

(highly). This compression likely reflects the design. Differences in quality that would stem from diagnostic or treatment skills were not elicited in our study, which narrowly focused on encouraging the take-up of preventives. Furthermore if black doctors were higher quality, they should perform better with *all* patients and on *all* tests. Although our recruitment efforts were focused on African-American men, 12 clients identified as from another racial or ethnic background.[5] Among this out-of-sample group, individuals were 20 percentage points *less* likely to choose invasive services in the post-consultation stage when randomized to black doctors (a finding that is more extreme than 97% of bootstrap coefficients on draws of 12 in-sample subjects). Moreover for the in-sample subjects, the differences in post-consultation preventive test take-up were much more muted for non-invasive screenings (e.g. blood pressure) than for exams that required blood (e.g. cholesterol). Thus, in order for an attribute correlated with the race of black doctors to be driving our results, it must manifest *only* when treating African-American male patients and *especially* for invasive exams.

This leads to another competing explanation, perhaps black male doctors exerted more effort with patients who shared their racial background. Since communication requires some amount of effort, this is not an interpretation to which we object (though we note if communication is more natural within concordant pairs, black doctors might be expending less effort to achieve the same or better results — i.e. communication may be more efficient). Time spent with patients has been used as a proxy for provider effort (Das et al. 2016). Equating time spent with effort is problematic in our setting because it reflects many different factors. A longer time spent could simply reflect the treatment effect (i.e. subjects elect to receive more services from black doctors), low quality (i.e. difficulty performing the services), or communication (i.e. a better patient-doctor connection facilitating credible information exchange). We find that black doctors indeed spent more time with subjects, but this finding is driven by the treatment effect — the difference in visit lengths is small and statistically insignificant after adjusting for the selected services. If we examine another potential proxy for effort, the allocation of services to the "highest need" subjects, we fail to find evidence that doctors of either race were expending effort to target interventions. Lack of targeting also reflects our instruction to the study doctors to try and encourage all patients to take up preventives.

Although years of experience in the medical field do not differ by race of doctor, it is possible that black male doctors have more familiarity with serving black male patients. This sorting would be consistent with national statistics on doctor-patient pairings as well as with the tendency for minority physicians to work in medically underserved areas with

[5]To avoid conflict, we provided services for the handful of people from other backgrounds who were consented in to the study but deleted them from the main analytical sample.

more low-income and minority patients (Komaromy et al. 1996; Walker, Moreno, Grumbach 2012). Our study doctor survey reveals that black doctors were more likely to have seen at least five black adult male patients a week, though this experience does not predict doctor fixed effects. In addition, in the context of our own experiment, non-black physicians did not "close the gap" with black doctors vis-à-vis post-consultation preventive care take-up over time.

Lastly, we do not find evidence for the controversial hypothesis that subjects or doctors were discriminating against each other. First, there was no race-preference elicited in the pre-consultation (tablet) stage. Second, the comments and ratings on feedback forms were consistently positive for both sets of doctors. As for provider-level discrimination, all doctors who were involved in the study knew the goal was to improve the preventive care of black men (though were blind to the notion that their race was being randomized, thus we could not administer implicit association tests). Taste-based discrimination by doctors would again be inconsistent with non-black doctors being rated as highly as black doctors. We also failed to find evidence that doctors of different races were using distinct thresholds to test patients for disease, consistent with Chandra and Staiger (2010).

Racial concordance between subjects and doctors appears to be a particular component of social distance that is influential in affecting demand. Alternative concordance measures, such as whether subjects and assigned doctors share approximately the same age or educational attainment, do not predict healthcare demand in any meaningful way. Nor does race interact with these other concordance measures. Such findings should be interpreted with caution since these characteristics were not randomized.

Similar to prior scholarship on incentives for preventives among low-income communities, (Banerjee et al. 2010; Cohen and Dupas 2010; Thornton 2008) we find that financial incentives for the flu shot increased demand for the vaccine: by 19 percentage points for a $5 incentive and 30 percentage points for a $10 incentive in the pre-consultation stage. Yet not all those who selected an incentivized flu shot actually received it: about 18% of subjects randomized to black doctors and 26% randomized to non-black doctors declined the shot in the post-consultation stage (many cited contraindications). And regardless of incentive level, black doctors increased demand in the post-consultation stage — convincing about 26% of subjects who initially turned down an incentive and refused a flu shot to obtain it, suggesting subsidies and (meeting with) black doctors are not perfect substitutes.

In the setting of imperfect information regarding the benefit of healthcare, demand curves cease to be a sufficient statistic for welfare calculations (Pauly and Blavin 2008; Baicker, Mullainathan, and Schwartzstein 2015; Chetty, Looney, and Kroft 2009). Furthermore, we incentivized take-up for only one preventive yet demand for every preventive was affected by a black doctor treatment. Thus, to make progress on valuation, we combine published estimates on the health value of interventions offered in our clinic with results from our study. The health value estimates come from cost-effectiveness simulations in which the screen-positive population obtains and complies with guideline-recommended therapy. Using this approach, we calculate that black doctors would reduce mortality from cardiovascular disease by 16 deaths per 100,000 per year, accounting for 19% of the black-white gap in cardiovascular mortality (Kahn et al. 2010; Dehmer et al. 2017; Murphy et al. 2017; and Harper, Rushani, and Kaufman 2012). If these effects extrapolate to other leading causes of death amenable to primary or secondary prevention, such as HIV/AIDS or cancer, the gains would be even larger.

These calculations presume that there is a supply of African-American male doctors who could screen and treat black male patients. This might not be a safe assumption. Black males are especially under-represented in the physician workforce, comprising about 12% of the U.S. male population but only 3% of male doctors (AAMC (2014); Census Bureau (2013)). According to a recent report by the AAMC (2015), the number of black male medical students has been roughly constant since 1978 (when 542 matriculated into medical school compared to 515 in 2014). Returning to the non-experimental results, black male respondents were 26 percentage points less likely than white respondents to state that a doctor who matched their race and sex was available to them.

The remainder of the paper proceeds as follows. Section II describes the experimental design and the hypotheses tested. Section III describes the data, empirical approach, and the characteristics of study subjects. Section IV presents the main findings and Section V explores potential mechanisms and validity concerns. Section VI discusses health benefits and Section VII concludes.

II. Experimental Design and Hypotheses Tested

A. Design

The experiment was conducted in Oakland, California, in the fall and winter of 2017–2018 (see Figure 1 for study design and flow). We recruited men from 19 black barbershops as well as two flea markets in and around the East Bay (about 88% of all recruited were at barbershops). Field officers (FO) approached men in the barbershops to enroll in the study. After obtaining written informed consent, the subject was given a short baseline survey.[6]

[6]Protocol information and links to the pre-analysis plan as well as other study documents are provided in the Online Appendix. Field officers were mostly minority or first-generation college students planning to apply to medical school. FOs were encouraged to approach men who were black, the majority of clientele

The baseline survey included questions on socio-demographics, healthcare, and mistrust. For completing the survey, the men received a coupon (worth up to $25) for their haircut or, at the flea market, a cash incentive. After completing the baseline survey, the subjects were given a coupon to receive a free health screening for blood pressure, BMI, cholesterol, and diabetes at the clinic we operated on eleven Saturdays (see Appendix Figure 1). Subjects were encouraged to come to the clinic promptly, and subjects who did not have transport could receive a ride to the clinic courtesy of Uber. Field officers used their own smart phones to obtain the rides. Attendance at the clinic was encouraged with a $50 incentive.

Upon arrival at the clinic, subjects who had a valid coupon were escorted into a waiting room where a ticket number was dispensed. Once their ticket number was called, they were led to a private patient room by a receptionist officer (RO). ROs wore crimson polo shirts with a Stanford - Bridge Clinical logo and khaki pants. The RO would then provide the subject with a tablet, which randomized the subject to a flu vaccine incentive and to a black or non-black doctor. Fourteen doctors participated in the experiment, including eight non-black and six black. We recruited study doctors using electronic and print advertisements to Alameda-Contra Costa Medical Association as well as with announcements at various meetings throughout the Bay Area. We only refused one physician who completed the steps for the application, due to liability concerns.

SurveyCTO programmed in-form randomization using a computerized random assignment algorithm for the tablets. At least four doctors were on site every Saturday (for reference, the median number of physicians at a community health center is five (Ku et al. 2015)). Note that the tablet was the first time subjects learned about the opportunity to receive a flu vaccine, since it was not advertised. This design decision was based on our focus groups, where men expressed fear about the vaccine and we were concerned that advertising the shot would reduce attendance. The RO would collect the coupon and give the subject his $50 participation incentive, then instruct the subject on how to use the tablet. Two practice questions were answered by the subject with the RO present to make sure they could operate the tablet. Fourteen subjects were illiterate and needed to have the RO read the tablet to them. We test for robustness to dropping those observations in Section IV. The RO then exited the patient room and allowed the subject to make their medical decisions in private.

The tablet introduced the subject to their assigned doctor and emphasized the doctor

at the recruitment barbershops. However, they were also instructed they should not confront anyone who insisted on taking the survey and receiving the free haircut even if they do not appear to meet study criteria (i.e. individuals who self-identified as African-American males and who were at least 18 years of age). The net effect is that we were very successful at recruitment in the short amount of time (over 1,300 subjects in about three months) but 14 individuals who came to the clinic did not meet study criteria and were removed from the main analysis — see Figure 1. These out-of-sample subjects are used in the exploration of mechanisms discussed below.

would be providing the services:

> *Your assigned doctor for today is Dr. [Last Name]. On the next page, you will be asked to select the services you wish to receive from Dr. [Last Name]. Dr. [Last Name] will administer all the services that you choose.*

In addition, the same generic information about doctor training was provided:

> *Dr. [Last Name] is a medical doctor licensed to practice in the state of California and currently practicing in the Bay Area.*

The text was accompanied by a large headshot photo of the doctor in a white coat with a red background. Doctors were told the purpose of the photos was for identification cards and were not aware that the photos were shown to subjects on a tablet in the pre-consultation stage. Tablet screenshots can be found in Figure 2. To protect the identity of the study doctors, there are no photos in the figure. The screenshots are not shown to scale, the tablet screen was approximately 10 inches. As described further below, subjects seemed to read the text on the tablet since they responded to information about the incentive for a flu shot and whether the test required a blood sample.

The next screen listed four services in a random order (blood pressure measurement, body mass index measurement, cholesterol testing, and diabetes testing) as well as the doctor photo and queried the subjects on which services they would like to receive. The need for a finger prick of blood for diabetes and cholesterol was clearly demarcated. Selecting "none of the above" was also an option.

The following screen apprised the subject that they could also obtain the flu shot, which would "protect you and your community." Those randomized to receive an incentive were then informed they would obtain \$5 or \$10 for selecting the flu shot. The doctor's photo was shown for a third time and the subject was asked whether they would like to receive a shot from *Dr. [Last Name]*. If the subject responded affirmatively, a list of screening questions would appear for contraindications. Subjects were informed the \$5 or \$10 incentive would be given regardless of whether they had a contraindication. This was necessary to encourage reporting of any condition which could make flu vaccination potentially dangerous (e.g. allergic response). However, subjects who were reluctant to receive the shot in the first place could lie about having a problem. The RO returned to the patient room, collected the tablet, recorded the responses, and handed a clipboard to the assigned doctor. It's possible that subjects could have doubted information on the tablet, such as whether the assigned doctor would actually meet them. Yet many seemed to believe (and respond) to the flu shot incentive by choosing it and our results on this subset are similar – results available on request.

Study doctors were instructed to encourage patients to receive all preventives.[7] The doctors, subjects, and field staff were not informed that doctor race was being randomized, though they could have inferred it. They were explicitly told that the purpose of the study was to improve the take-up of preventive health screening services for African-American men (the study was officially labeled the "Oakland Men's Health Disparities Project"). Doctors were aware that subjects were randomized, so that they would only meet with subjects assigned to them. Due to the nature of the malpractice coverage we were able to provide, study doctors were instructed not to provide medical care other than the services that were covered by the study. Subjects were also informed that the doctors were only able to provide the set of preventives listed on the tablet. If subjects had alarming values on any of their tests, there was an emergency protocol in place. After the visit was completed, subjects filled out a feedback form. They were then escorted out of the clinic by an RO and the ride-sharing service was called if needed. The study was approved by the IRB committee of Stanford and by the IRB committee at NBER for the non-experimental sample. The IRB committees at Berkeley and MIT ceded authority to Stanford.

B. Conceptual Framework and Hypotheses Tested

The experimental design allows us to test two competing hypotheses, which are formalized in the Online Appendix. The model follows Pauly and Blavin (2008) and Baicker, Mullainathan, and Schwartzstein (2015), allowing patients to have false beliefs about the benefits of preventive care leading to underutilization (i.e. demanding less than what is privately optimal). This assumption mirrors what we observed in the field with many patients expressing false beliefs or present-bias. For example, one subject had been diagnosed with diabetes in the past but "refused to believe it."[8] Others thought flu shots caused sickness, or that other non-proven remedies could ward off illness instead, echoing findings in Pettey et al. (2016). Several said that they would get the shot later. One patient made a possible reference to the syphilis experiment in Tuskegee stating he did not want the flu shot out of "fear of being experimented on." We note that this subject's belief was accurate in the sense that he indeed was an enrollee in our study, and therefore part of an experiment. It also reflects many findings from the medical literature which suggest that African-Americans are wary of participating in clinical trials and part of this hesitation may be related to Tuskegee (Murthy, Krumholz, Gross 2004; Braunstein et al. 2008; Scharff et al. 2010).

[7] Similar to Coffman and Niehaus (2018) who study homophily in the context of the seller-buyer relationship, we did not provide a specific script for the doctors to use in their meetings. A script could have limited communication and made doctors appear less genuine/trustworthy in what was a real clinical encounter.

[8] "Disbelief of diabetes diagnosis" has been associated with medication non-adherence among African-American patients (Shiyanbola, Brown, and Ward 2018).

If, in addition to the discomfort associated with testing, individuals have a strong aversion for doctors who do not share their race (as in Becker 1957), then simply learning their doctor is African-American via a photo on the tablet in the pre-consultation stage is hypothesized to increase demand relative to the control group.

In the post-consultation stage, we assume a doctor provides information sufficient for the patient to correct his false belief, yet whether that information is considered credible or comprehensible depends on the social distance between the two agents (Tabellini 2008). If homophily facilitates the successful transmission of information between doctor and patient, then differences across subjects assigned to doctors of different races would only be detected after the consultation. Finally, there is the possibility that both forces are at work: aversion to a particular race of doctor in the pre-consultation stage reinforced by a lower perceived benefit, on average, from the same, in the post-consultation stage, leading to a widening of the take-up gradient across doctors.

III. Empirical Strategy and Sample Characteristics

The purpose of the study is to estimate the causal effect of doctor race on the preventive healthcare decisions of African-American men. We begin by presenting an overview of our estimation framework and the data used in the study. We then turn to describing characteristics of the study sample.

A. Estimating Equations

Using experimental data, we estimate the following equation:

$$Y_i = \alpha + \beta_1 \cdot \mathbb{1}_i^{BlackMD} + \beta_2 \cdot \mathbb{1}_i^{\$5} + \beta_3 \cdot \mathbb{1}_i^{\$10} + \Gamma' X_i + \epsilon_i \tag{1}$$

where i is an individual subject. Y_i is the demand for preventives during various stages of the experiment. $\mathbb{1}_i^{BlackMD}$, $\mathbb{1}_i^{\$5}$, and $\mathbb{1}_i^{\$10}$ are indicators for random assignment to a black doctor, a five dollar flu incentive, and a ten dollar flu incentive, respectively. X_i are characteristics of the subject and are included in some specifications to improve precision. In addition, to explore mechanisms, characteristics are interacted with randomized components. The results from our analysis of Equation 1 will show that the flu incentives only consistently affect demand for the flu.

In addition, we stack the data where each each observation is a subject-by-preventive service and we fully interact the black doctor treatment with indicator variables for each service. This allows us to test whether the treatment effect varies across services.

When estimating standard errors for the main treatment effect of interest, we approach the data as if our design involved randomizing clusters of patients to a particular doctor instead of individual assignment of subjects to doctors of a given race (Abadie et al. 2017; Bertrand, Duflo, and Mullainathan 2004). These standard errors are likely incorrect given the small number of clusters (Cameron and Miller 2015); therefore, we also report randomization inference (RI) p-values using all 3,003 $\binom{14}{6}$ combinations of doctor race.[9] When examining interactions between having a black doctor and other covariates, we generate plots of the joint distribution of RI draws (see Heß 2017) – available on request. The mode of inference decision involves a Type I and Type II error tradeoff and our focus is on minimizing the former.

To further probe mechanisms, we collected non-experimental data from a survey of 1,490 other black and white male respondents whose education profile mirrored that of our experimental sample. The sampling frame was a panel of respondents managed by Qualtrics. We designed the survey to capture whether the preference for a racially concordant provider is unique to black male respondents and whether it varies across healthcare domains. Specifically, we estimate the following equations:

$$\mathbb{1}_i^{RaceMD=k} = \alpha + \beta_1 \cdot \mathbb{1}_i^{RaceResp=k} + \Gamma' X_i + \epsilon_i \tag{2a}$$

$$\mathbb{1}_i^{RaceMD=RaceResp} = \alpha + \beta_1 \cdot \mathbb{1}_i^{BlackResp} + \Gamma' X_i + \epsilon_i \tag{2b}$$

$$\mathbb{1}_{il}^{RaceMD=RaceResp} = \alpha + \beta_1 \cdot \mathbb{1}_i^{BlackResp} + \lambda_l \cdot \mathbb{1}_l^{Domain} + \Gamma' X_i + \epsilon_{il} \tag{2c}$$

where i indicates respondent, k signifies race (black or white) and l is one of the domains cited by the World Health Organization (WHO) as features of a responsive health system: access, quality, and communication (Gostin et al. 2003). X_i in the above refers to respondent's age, education, and income. Equation 2a examines whether respondents are relatively more likely to prefer doctors who share their racial background, where $RaceMD$ and $RaceResp$ are either both black or both white. Equation 2b tests whether the preference for racial concordance differs between black and white respondents. Finally, Equation 2c investigates whether the importance of concordance differs across domains as well as by race of the respondent.

B. Sample Characteristics

We first examine characteristics of the subjects who chose to come to the clinic, then proceed to check that observable characteristics are balanced across arms before turning to our main

[9] There are 14 doctors in the experiment, six of whom comprise the "treated" group. We thus permute $\binom{14}{6}$ combinations of physician-level treatment holding the number of treated units (6) constant.

findings.

Recruitment and Participation — To examine participation in the experiment, we modify Equation 1, regressing X_i on a dummy for *Clinic Presentation*. These results are gathered in Table 1.[10] In general, those who came to the clinic were older, had lower self-reported health, visited the emergency room more in the past two years, and were less likely to have a primary care physician (PCP) compared to those that did not come. The selected men also had lower reported income; were less likely to be married; were more likely to be receiving unemployment, disability, or Supplemental Security Income; were 19 percentage points more likely to have a high school diploma or less; and were 13 percentage points more likely to be unemployed.

Recall that the visit to the clinic was incentivized and barriers associated with not having a car or a license were alleviated by providing free transport to and from the clinic. The combined reduction in transport barriers and incentive to attend likely contributed to this pattern of participation.

Balance — Treatment groups are well-balanced on observables with two exceptions (see Table 2). The cell containing subjects who were randomized to a non-black doctor and $10 incentive for flu are more likely to be uninsured and less likely to have good self-assessed health. The only significant joint F-test is on self-reported health, but including these two covariates, among others, in Equation 1 does not alter our results (see discussion below). Appendix Table 1 demonstrates that the groups are also well-balanced when examining randomization to a black doctor or a flu incentive amount separately.

IV. Experimental Results

Main Results — We now turn to our experimental results and the principle aim of our analysis. Do black male subjects randomized to black male doctors demand more preventives? Table 3 presents the main results conditioning only on the randomized treatments: doctor race and flu incentive. Baseline results with only the black doctor treatment can be found in Section IV. In the pre-consultation stage, across every test offered, the race of the doctor in the photo did not influence demand in any significant way (see Table 3 Panel (A)). These results are also apparent when comparing the means of pre-consultation take-up among black and non-black doctors in Figure 3 (the pair of vertical bars on the left side of each figure). Such findings are inconsistent with racial aversion playing a major role in take-up decisions. Rather, they are supportive of pre-consultation case 3 of the model in the Online Appendix

[10]Our main clinic sample includes all of those who identify as African-American and are at least 18 years of age on the baseline survey as well as approximately 9% who skipped the demographic questions but were recruited in a black barbershop. In Section IV we assess sensitivity to various sample restrictions.

— in which subjects do not add doctor-related costs to their utility calculation or add it equally across doctor race types.

We find that the incentive influences pre-consultation demand for the flu shot. Approximately 20% of subjects selected the flu shot on the tablet in the absence of an incentive. A $5 incentive increased flu take-up by about 19 percentage points, and a $10 incentive increased it by 30 percentage points. With a $10 incentive, almost 50% selected the flu shot on the tablet, though, as mentioned previously, not all subjects who initially chose flu shots received it since subjects could revise their decision, usually by endorsing a contraindication. The responsiveness of pre-consultation demand to information about the incentive suggests that subjects were attending to the tablet.

In the post-consultation stage of the experiment, the effect of being randomized to a black doctor is statistically significant and, as we calculate below, medically meaningful particularly for invasive exams. Table 3 Panel (B) Column (1) shows that subjects randomized to a black doctor are 11 percentage points more likely to demand a blood pressure measurement, an increase of 15% compared to the non-black doctor mean. According to the estimates in Panel (B) Column (2), the effect of a black doctor on BMI take-up is 16 percentage points or 27%. Note that, for both of these tests, subjects assigned to non-black doctors are also demanding more exams (see Figure 3 Panels (A) and (B)); however, those assigned to black doctors do so more frequently. The RI p-values are much lower than in Panel (A) (e.g. 0.22 for BMI compared to a minimum of 0.43 across all tests in the pre-consultation panel), but are not below the conventional levels of statistical significance.

Moving to the invasive tests — those that required blood samples from the subject or involved an injection — the results demonstrate an even larger relative effect of black doctor assignment on demand for preventives among black male patients. In contrast to the non-invasive services, subjects assigned to non-black male doctors were not, on average, more likely to agree to the invasive services after meeting the doctor (See the light (gray) bars in Figure 3 Panels (C)–(F)). A subject randomly assigned to a black doctor was 20 percentage points (49%) more likely to agree to a diabetes screening and 26 percentage points (71%) more likely to accept a cholesterol screening (Table 3 Panel (B) Columns (3–4)). With respect to the flu vaccine, which was cross-randomized with an incentive, subjects randomly assigned to a black male doctor were 10 percentage points more likely (56%) to agree to the flu shot *conditional* on financial incentives offered in the pre-consultation stage to choose the flu.[11]

Columns (6) and (7) explore the effect of having a black doctor on two composite mea-

[11] See Thirumurthy, Asch, and Volpp (2019) for a discussion on the "uncertain effects" of financial incentives in health.

sures, the share of the four non-incentivized exams (all screenings except the flu shot) and the share of invasive exams (cholesterol, diabetes, and flu). Overall, the treatment effect increases take-up by 18 and 19 percentage points, respectively. These results are consistent with the conceptual framework in which all doctors relay basic information regarding the benefits of preventive care yet social distance acts to discount information from a discordant source (post-consultation case 1).

Figure 4 Panel (A) plots the post-consultation black vs. non-black doctor difference in take-up by exam. The figure reveals the percent difference between black and non-black doctors is positively correlated with the invasiveness of the test. Blood pressure is a non-invasive test and was performed in the patient room. Therefore, it is unsurprising that this low risk and low hassle test had the lowest black doctor effect relative to non-black doctors. BMI measurement required the doctor to escort the subject down the hallway to a public room where there was a scale and height machine. The doctor used both devices to measure the height and weight of the subject and then calculated the BMI. Cholesterol and diabetes tests required a finger prick of blood (usually two separate sticks). The cholesterol and diabetes tests also took longer than other tests — on average, visit lengths for subjects who selected diabetes tests were about six minutes longer; a cholesterol screening added about three minutes. For more invasive tests, the results suggest there was a greater advantage to being assigned a black doctor.

Table 3 Panel (C) presents the difference between post- and pre-consultation demand (i.e. the delta). This is similar to conditioning on the first choice, which, per above, was not statistically different across race of male doctor, and is a direct measure of how much demand changes after meeting the randomly assigned doctor. For instance, in Column (4), subjects assigned to a black doctor were 25 percentage points more likely to select a cholesterol screening after meeting their physician than those assigned to a non-black doctor.[12] To benchmark the effects shown in Panel (C), we follow DellaVigna and Kaplan (2007) and calculate persuasion rates as a measure for how much subjects changed their behavior upon exposure to a black doctor. Figure 4 Panel (B) demonstrates that the persuasion rate is high relative to studies surveyed in DellaVigna and Gentzkow (2010).

The above results suggest the treatment effect for invasive exams is stronger than for non-invasive ones. To investigate this further, we stack the data to create a subject-preventive panel, fully absorbing the black doctor coefficient by interacting it with indicators for every service (see Table 4). This specification allows us to test the joint significance of the black doctor*invasive interactions as well as the differential effects of the black doctor treatment for

[12]Appendix Figure 2 plots the histogram of delta as a share of the four non-incentivized tests (i.e. excluding the flu).

invasive exams. The results are shown in the bottom four rows. Consistent with the above, in the pre-consultation stage the F-tests for the interaction of black doctor with invasive exams and its difference with non-invasive exams are around one and not significant. The last two columns examine post-consultation and delta outcomes, respectively, and demonstrate a consistent additional marginal effect of black doctor and invasive exams which differs significantly from non-invasives, particularly in Column (3).

Robustness — In Appendix Table 2, we probe whether our results are sensitive to the inclusion of covariates thought to influence health, such as subject age (and its square), having a regular PCP, insurance, the clinic visit date, education, income, and self-assessed health. The results are very similar to those presented in Table 3 and Figure 3. We also selected covariates via double lasso, but failed to find consistent predictors of take-up other than our treatment variables (Duflo 2018; Belloni, Chernozhukov, and Hansen 2014).

As a robustness check, we include different fixed effects (RO, date, and recruitment location (Appendix Table 3 Panel (A)) and different samples (i.e. including everyone who consented regardless of their race or ethnicity, excluding those who could not read, including only those who responded to every demographic question (Appendix Table 3 Panel (B))); again the results are very similar. We also show that the results are not sensitive to dropping indicators for flu incentive levels (Appendix Table 4).[13] Finally, race appears to be a special facet of social distance — sharing the same age or educational background as doctors does not seem to positively influence take-up (see Table 5). Caution should be used in interpreting these results as neither education nor age was randomly assigned. In sum, the results presented thus far reveal that, for African-American men in our study, the opportunity to meet with a black male doctor has a consistent, large, and robust positive effect on the demand for preventives.

V. Mechanisms

In this section, we explore potential mechanisms for our results. We do so in five ways: first, by using data from the physician notes and subject feedback forms to further our understanding of the clinical encounter; second, by examining heterogeneity across subjects; third, by using non-experimental evidence from an additional survey we conducted on approximately 1,500 black and white men concerning preferences over doctors; fourth, by using publicly available, nationally representative data from a survey of health utilization; and fifth, using survey responses and background information from the study doctors. We begin by examin-

[13] In unreported results, we do not find evidence that knowing another subject at the clinic, a practice question we asked to ensure subjects could operate the tablet, affected demand.

ing the role of communication. Then we discuss other possible interpretations of our results including physician effort, quality, and discrimination.

A. Communication Between Patients and Doctors

Our primary data sources for understanding what transpired during the clinical encounter are doctors' notes on the patient and subject feedback forms about their clinical experience. As mentioned above, doctors were instructed to provide only the advertised services to subjects. In Table 6 Panel (A) Column (1) we find evidence that subjects assigned to black doctors were 10 percentage points more likely to try and talk to their doctor about issues unrelated to the preventive care. The doctors also indicated whether there was anything "notable" about the patient encounter on the patient files. Subjects were 11 percentage points more likely to have this section filled in if their assigned doctor was black (Column (3)). We analyzed the content of these notes by having three students who were blinded to the treatment hand code them as related or unrelated to the screening. Subjects assigned to black doctors were 9 percentage points more likely to discuss personal matters or health issues unrelated to the screening (Column (5)).

Although the aforementioned results lack precision, they suggest that communication was an important feature of the concordant patient-doctor interaction. Indeed, when we interact black doctor with an indicator for the subject receiving any invasive exam (even numbered columns in Panel (A) Table 6) the interaction terms are positive and large in magnitude. Thus, the relationship between black doctor and communication was strongest among subjects who obtained an invasive exam.

Qualitative evidence from the subject feedback forms and doctors' notes also support the mechanism of improved communication and the correction of false beliefs. One subject randomized to a black doctor wrote: "Dr. XXYY was excellent, he talked me into getting a flu shot and the conspiracy theories. I said 'Oh!' Great visit and putting me on track to monitor my sugar and cholesterol. Thanks!" As for the doctors' notes, a frequent phrase was "initially refused but agreed after counseling." Finally, we note that the experimental results on communication are robust to controlling for the time spent with subjects (see Appendix Table 5 Panel (A)). Thus, per minute communication was more likely to occur between racially concordant doctors.

In Table 7, we test whether subjects assigned to black doctors were more responsive to the treatment based on their baseline demographic characteristics (Panel (A)), study clinic experience (Panel (B)), or past healthcare experience (Panel (C)). We focus on invasive exams because of the evidence from Table 3 that black doctors affected demand for these services most. We fail to find strong evidence of an important interaction effect between black doctor and either low income (reported household income below $5,000), age (an indicator for younger than 40), or low education (an indicator for a high school degree or less). The absence of a statistically significant finding in this latter case is interesting, since, if black doctors were simply better at providing information to the less well-informed, the results would presumably be strongest among those with lower levels of education. Though it does comport with our conceptual framework that emphasizes how the source of the information and the connection between source and recipient, not just the information itself, matters for clinical decision-making.

In contrast, both Panels (B) and (C) reveal important interactions between the black doctor treatment and either hassle costs associated with the study clinic or limited prior healthcare experience, respectively. In particular, subjects who were randomized to a black doctor but had longer wait times (an indicator for over an hour) demanded more services than those exposed to a similarly lengthy wait time, but who were assigned to a non-black doctor. Subjects who experienced high congestion (greater than nine people in the waiting room, the 50th percentile) or those who were recruited from farther away locations (longer than 18 minutes by car, the 50th percentile) also elected to receive more services when randomized to a black doctor than a non-black doctor.[14]

African-Americans visit the emergency room more often than non-Hispanic whites, which some have linked to lack of insurance, lower socioeconomic status, and mistrust that precludes healthcare utilization until an advanced stage of illness (Arnett et al. 2016, Brown et al. 2012). Panel (C) demonstrates that those who use the emergency room more often increased their demand for services when randomized to a black doctor. This result is particularly strong for the uninsured: in unreported results, the coefficient on the interaction between black doctor and number of ER visits is roughly seven times greater if a subject reported having no insurance.[15] Similarly, those who had no recent screening had a heightened response.

Research in medicine finds that black men have higher levels of medical mistrust than their white counterparts, and this mistrust is correlated with delays in care, lower healthcare utilization, and worse health outcomes (Kinlock et al. 2017, Nanna et al. 2018, Hammond et al. 2010). As discussed above, we find that subjects increased their demand of all preventive services when assigned to a black doctor, and this effect was heightened if the screening test

[14]The wait time and congestion interactions have fewer observations due to missing data for the first two clinic days. These variables are balanced across black and non-black doctor treatment.

[15]We also asked a question about usual source of care in the baseline survey, but many subjects selected multiple options making their responses difficult to interpret. As in Zhou et al. (2017), we find that the uninsured use the ER at a similar rate to the insured, though they have fewer total hospital admissions and doctor visits. Results available on request.

was invasive. More invasive procedures, such as taking blood or providing injections, require a higher degree of trust between doctor and patient. We examine whether men who scored higher on our medical mistrust measure responded differently to the black doctor treatment than other groups. In Column (3) of Panel (C) we find that subjects were 5 percentage points more likely to obtain preventive services per a one unit increase in medical mistrust (on a collapsed scale of 1–3, see Data Appendix for details) when randomized to a black vs. non-black doctor. For the uncollapsed physician mistrust measure (scale of 1–5), 50 subjects said they would "not at all" trust doctors to make decisions on their behalf. These high-mistrust subjects were 14 percentage points (clustered s.e. 0.066) more likely to take up invasive exams after meeting with a black doctor than the least mistrustful group who "completely" trusted their doctor.

An additional source of data we use to inform mechanisms is from a survey we conducted on 1,490 African-American and white (self-identified) males. We matched the survey sample to the recruited participants in terms of education, so that approximately half of the survey respondents had a high school education or less. Given a choice between a black, white, or Asian male doctor, respondents were asked to choose which doctor ranked the highest across three WHO domains: quality, communication, and accessibility. The results are reported in Table 8.

First we examine respondent preferences for a doctor of the same race, i.e. concordance (Equation 2a). In Column (1), we find that black respondents were more likely than white respondents to select black male doctors as the most qualified. Column (2) demonstrates that white respondents selected white doctors more often than black respondents. This finding is consistent across other domains, whereby both sets of respondents were relatively more likely to choose a concordant physician rather than a discordant physician (see Columns (4) and (5) and Columns (7) and (8)).

Second, we examine whether preferences for concordance vary across race (Equation 2b). Column (3) tests whether respondents were more or less likely to rate concordant doctors as most qualified. We find that white respondents were 6 percentage points more likely to select white doctors as most qualified than black respondents select black doctors as the most qualified. Both sets of respondents view concordance as important for communication (about 69%, see Column (6)) and there is no difference between the two groups. Turning to accessibility, responses from the two groups differ significantly (Column (9)), a point we return to when discussing external validity.

Third, we estimate Equation 2c, which tests whether concordance is stronger for some domains than others. In Column (10) we find that black and white respondents were 17 percentage points more likely to select a concordant doctor when the question was about communication as opposed to when the question referred to quality. The importance of concordance for communication was similar across respondent race.

Figure 5 summarizes the results from Table 8. The figure plots the percent of respondents from a given race selecting a doctor of their own race across the three domains. We find a slight preference for concordance when it comes to quality, though both sets of respondents are very close to the (red) 50 percent line, indicating that, on average, respondents were as likely to select concordant physicians as they were to select discordant physicians. In sharp contrast, for questions related to communication, both black and white respondents shift rightwards: reflecting a clear preference for concordant doctors. Nearly 65% of black respondents and 70% of white respondents reported that a doctor of their own race would understand their concerns best.

To understand whether these patterns are also found in nationally representative data, we use the Medical Expenditure Panel Survey (MEPS), which queries individuals on characteristics of their doctor as well as utilization (Blewett et al. 2018b). Respondents were more likely to see a doctor of their own racial/ethnic group — though that varies across the race of the respondent. Specifically, 85% of white respondents and 71% of Asian respondents reported their usual medical provider was of the same race (see Appendix Table 6). Although more black respondents report their doctor is black than respondents of other backgrounds, only 26% of black respondents said they had black doctors. The pattern for Hispanics is similar. This may reflect under-representation of African-Americans and Hispanics in the physician workforce, a point we return to when discussing external validity below.

Appendix Table 7 reports correlations between patient-doctor concordance and three outcomes: whether a respondent would go to their doctor for preventive care, whether they think their doctor listens to them carefully, and whether their doctor's instructions were easy to understand. The sample is limited to adult males. The estimating equation includes indicators for patient and doctor race/ethnicity as well as concordant interactions. Results should be interpreted with caution given that patients are not randomly assigned to doctors or vice-versa. Nevertheless, the interaction between black male patient and black doctor is consistently positive and significant, indicating that said patients are more likely to seek out preventive care, feel their concerns are understood, and comprehend medical advice when paired with a black doctor. The main effect of black doctor is negative though not consistently significant — a finding that does not support differential quality of doctors across race. See also the next subsection for further evidence on (lack of) differential quality across race.

B. Threats to Internal Validity

In this section, we consider whether doctor race represents a causal effect. Race is not randomly assigned in the population. Thus, in the sample of doctors we hired, race could be correlated with a characteristic that influences the ability of doctors to encourage subjects to take up preventives (i.e. our outcome of interest). Prominent potential omitted variables include quality and effort, which are hard to measure outside of the clinic context. In addition, with a finite number of physicians, the findings might be driven by outliers in either group. Finally, there is the concern that either subjects or doctors discriminate. We discuss each of these possible interpretations in turn.

Physician Quality — Physician quality is thought to influence patient outcomes, but is acknowledged to be complex and difficult to measure, particularly in primary care (Young, Roberts, and Holden 2017; AHRQ 2016). Some measures of quality include malpractice complaints, physician report cards, and training characteristics. In this study, we use all of the aforementioned, plus an additional survey we designed to assess typical patient panel characteristics as well as persuasiveness and content knowledge of our study doctors.

Turning to the quality metrics, first, all doctors were vetted by a medical liability company and Stanford attorneys as a requirement of their participation. Second, after the encounter, we asked subjects to fill out a feedback form before leaving the clinic. They rated their experience on a scale of 1 to 5 and then asked whether they would recommend their doctor to a friend. As seen in Table 6 Panel (B) Columns (3) and (5) there are no statistical differences between ratings and recommendations among those assigned to black or non-black doctors. We interpret these results with caution however, given that the satisfaction measures were all very high and without much variation. For instance, the mean experience rating was about 4.8 with 85% of subjects characterizing it as excellent (a rating of 5) and 99% saying they would recommend their doctor to a friend.[16]

Third, we gleaned details of doctors' training from their resumes and a study doctor survey querying them on their patient demographics, overall persuasiveness, and working medical knowledge using board style questions. The doctors were unblinded by the time of the survey. Characteristics such as experience, medical school rank, and board question performance were similar across groups, though black male doctors were less likely to be trained in internal medicine than their non-black counterparts (see Table 9 Panel (A)). Non-black study doctors were more likely to state their patients comply with medical advice and

[16]Patients may provide feedback based on a different maximand and satisfaction scores do not systematically covary with other quality metrics (Chandra et al. 2016). There may also be bias in patient satisfaction scores, though this would be at odds with results in the pre-consultation stage (Garcia et al. 2019; Sotto-Santiago, Slaven, Rohr-Kirchgraber 2019; Poole 2019).

that they are able to persuade patients from all backgrounds to take up testing they had initially refused (see Table 9 Panel (B)). African-American doctors were more likely to see racially concordant male patients, findings consistent with nationally representative data (Appendix Table 6).

To further analyze quality, we modify Equation 1 replacing the black doctor indicator with a fixed effect for each study doctor. We then examine what explains the correlation between doctor attributes and the fixed effect estimates (see Table 10). Physician race explains approximately 45% of the cross-sectional variation and is highly significant. In Column (2) we add experience with black male patients, although the coefficient is positive it one-sixth the size of race and is not itself significant. We also fail to find any significant effect when interacting this variable with doctor race. In Columns (3) to (5) we include doctor race with and an indicator for a top 10 medical school, years of experience, and internist, respectively. All do little to explain the variation in fixed effects. Column (6) includes all the aforementioned covariates in the same specification. Comparing Column (1) and (6), race accounts for about 86% of the R-squared.[17]

If race of doctor in the study was highly correlated with quality, then we should find black doctors perform better on subjects from all backgrounds. Twelve individuals did not identify as African-American, but were still seen at the clinic because they had been consented to participate during recruitment. These clients were randomized across eight of the fourteen study doctors, equally balanced by race, and were 20 percentage points *less* likely to choose invasive services from black doctors in the post-consultation stage. We compare this result to a placebo test where we randomly select 12 in-sample subjects and regress the share of services received on black doctor. We find that the coefficient on black doctor for the out-of-sample group is lower than 97 percent of these bootstrap coefficients (see Figure 6).To the extent that quality is a relatively stable attribute of a clinician, this finding is inconsistent with a correlation between doctor race and quality confounding the interpretation of our results. These results may understate racial discordance given that the subjects were still recruited from black barbershops. In Appendix Figure 3 Panel (A) we do not find a significant difference in the length of black study doctors' notes with this sample, however, we do find that non-criteria subjects are much *less* likely to talk to black doctors (Appendix Figure 3 Panel (B)).

[17]Additionally, the R-squared with all of the control variables except the black doctor indicator is equal to 0.183. Thus, black doctor represents 41% of the residual variation ((0.516 - 0.183) / (1 - 0.183)). We also tested whether physicians improved over time at the clinic. Specifically, we collapsed the data to the physician-visit date level and regressed the post-consultation take-up on black doctor and a count of the visit day number for each doctor. The coefficient on visit day number is -0.004 and insignificant; the black doctor coefficient is 0.187 (s.e. 0.048).

As an additional measure of within study quality, we tabulate the number of mechanical errors on the medical devices by doctor. There were very few errors in total and they did not vary across race. Finally, we find no differences in online ratings of the study doctors on the site vitals.com. On a scale of 1 to 5, the average black doctor score was 4.4 and the average non-black doctor score was 4.6, though not all of the doctors were rated on the site.

Physician Effort — Another potential explanation is that black doctors exerted more effort when working with black patients than non-black doctors. Similar to quality, physician effort is difficult to measure. Often, time spent with the patient is used as a metric, but in our study this equivalence is complicated. As mentioned in the introduction, a longer time could reflect the treatment effect (i.e. subjects elect to receive more services from black doctors), low quality (i.e. difficulty performing the test), or communication (i.e. a better patient-doctor connection facilitating credible information exchange). In Table 6 Panel (B) Column (1) we find that black doctors spent approximately four more minutes with subjects. However, this finding is mainly related to our treatment effect of black doctor on invasive exam take-up. Indeed when we condition on such take-up (Panel (B) Column (2)) neither the main effect of black doctor nor the interaction with invasive are significant, though invasive testing does lengthen the encounter by 12 minutes.[18]

We also examine whether study doctors exerted more effort by targeting services to the those at increased risk for disease (as defined by national guidelines — see Data Appendix for details). Such targeting would require clinical acumen and effort since doctors were provided no information on the subjects' medical histories prior to their brief encounter. Results in Appendix Table 8 fail to find evidence of targeting.

Outliers — A third possibility is that our results are driven by outliers. As noted above, there are no prominent differences in observables (if anything, the set of black doctors attended lower ranked medical schools and were slightly less likely to be internists, see Table 9). To test whether any particular physician is driving our results, we estimate the black doctor effect dropping one doctor at a time. The results gathered in Appendix Figure 4 demonstrate that the results are remarkably stable across the leave-one-out estimates. If we drop the "best" black doctor, we obtain a consistent coefficient of 0.120 (cluster s.e. 0.043). In the most stringent condition, we omit the "best" black and the "worst" non-black doctor and still find a treatment effect near 11 percentage points. The randomization inference procedure also addresses this concern.

Discrimination — A fourth possibility is that subjects derive disutility from non-black doctors thus decreasing demand. Our results suggest this is unlikely. First, if aversion for a

[18]Likewise, Appendix Table 5 Panel (B) Column (1) controls for each individual test and shows that visits with black doctors were about one minute longer, though not significantly so.

particular race was strong, we would have expected to observe this in the pre-consultation stage, when subjects were first introduced to the doctor by tablet photo. As previously noted, though, we find no statistical differences in the pre-consultation tablet selections (Table 3). Second, in the post-consultation stage, we find that, on average, subjects assigned to non-black doctors increased their demand relative to the pre-consultation stage (see light (gray) bars in Figure 3), just not as much of an increase as with black doctors (and not at all with invasive exams).

Discrimination by non-black doctors could manifest as higher risk thresholds to test African-American patients for disease. Chandra and Staiger (2010) note the "Beckerian intuition" of taste-based prejudice implies that "providers may consciously or unconsciously use a higher benefit threshold before providing care to minority patients (for example, recommending a treatment to non-minority patients if it prolongs their life by at least three months, but only treating minority patients if it prolongs their life by at least five months)." Yet the distributions of test results conditional on testing, presented in Figure 7, are indistinguishable across race groups, suggesting both sets of doctors were following protocol and offering preventives to all subjects. Lastly, we note that if discrimination by patients or doctors were an important part of the explanation for our results, we would have expected variation in subject feedback across doctor race and lower scores for non-black doctors. Instead we find that the average ratings were very high and there was no difference across doctor race.

We qualitatively examine the subject comments regarding the clinical experience by visualizing the occurrence of specific words using the WordStat content analysis software. The plots of the subject comments are indistinguishable across those assigned to black and non-black physicians (Appendix Figure 5). The same analysis focused on doctors' notes (Appendix Figure 6) illustrates the centrality of the word "PCP" for non-black doctors and "patient" for black doctors. We examined the nature of this difference and found that non-black doctors were often recording reasons for refusals — the most common being that the subject said that he would obtain preventives from his PCP. In fact, for patients seeing non-black doctors, 60% of the references to PCP in the doctor notes are in the context of an excuse (e.g. "cholesterol and flu shot from PCP"), whereas this occurs in only 14% of the references to PCP by black doctors.

C. Threats to External Validity

In order to benchmark our results and assess their relevance for the larger discussion on reducing health disparities in the U.S., it's important to compare our study doctors and sample to the general population, bearing in mind that extrapolation should be done with caution.

Subjects — In terms of demographic characteristics, our clinic study subjects were more likely to be uninsured (28%) and unemployed (31%), as compared to black men in the U.S. (about 17% and 7%, respectively).[19] For the entire study sample, including those who did not come to the clinic, the average uninsured rate was 26% and the unemployment rate was 24%. However, the study samples are very similar in terms of average age and education relative to the rest of the U.S. (43 years and 58% with a high school education or less in the U.S. versus 43 years and 63% with a high school education or less in the clinic sample and 43 years and 53% for all recruited).

Turning to health characteristics, the average value for systolic blood pressure was 132.7 mm Hg consistent with stage 1 hypertension (distributions of medical screening results are displayed in Figure 7). The average BMI value was 27.4 kg/m^2 consistent with an overweight categorization. The average hemoglobin A1c was 5.8%, consistent with a diagnosis of pre-diabetes. About 1.4% of the sample had a hypertensive crisis — a critically high value of blood pressure requiring urgent care, 4.4% were morbidly obese, and 3.1% of the subjects had a hemoglobin A1c value in the seriously elevated range (i.e. $>$9%).

In terms of disease prevalence, about 30% of the screened study sample had values of blood pressure, BMI, and cholesterol consistent with hypertension, obesity, and dyslipidemia, respectively; and 15% had hemoglobin A1c levels diagnostic of diabetes. Some subjects indicated that they were on medications for these conditions; we only include them in the estimate if they chose to receive a screening. Despite our sample having higher rates of unemployment and uninsurance, these figures are unfortunately very similar to the prevalence of the aforementioned conditions among black men in the U.S. more broadly, as seen in Figure 8. If anything our screened study sample was slightly healthier than the average African-American male in the U.S. Specifically, the prevalence of high blood pressure in black men in the U.S. is 41%, compared to 30% for white men, the prevalence of hypercholesterolemia is 33% for black men compared to 37% for white men, and the prevalence of diabetes is 18% for black men vs. 9% for white men (Fryar et al. 2017; Hales et al. 2017; CDC 2017b; and CDC 2017c). These comparisons suggest that our findings are not due to a sample of individuals with worse health on average (Simon et al. (2016)).

Doctors — How representative were the doctors hired for our study? All doctors who participated knew the clinic provided preventive services to black men, many of whom lacked alternative medical options. Therefore, these doctors are plausibly drawn from the *least*

[19] Calculations on the U.S. population come from 2016 1-year American Community Survey data (Ruggles et al. 2017). Our study sample also appears more disadvantaged compared to ACS summary statistics for Alameda County.

prejudiced part of the distribution. The doctors also gave up their Saturdays in exchange for a fixed hourly compensation that they received through direct deposit or check. The compensation was competitive with the market rate for moonlighting physicians in the Bay Area. Doctors of both races attended highly ranked medical schools. Across all 14 study doctors, eight graduated from schools ranked in the top 10 of the U.S. News Research Rankings, a much higher share of graduates relative to the population at large. Black doctors in the study graduated from slightly lower ranked schools, consistent with the national data (see Appendix Figure 7).

One way our study was unique, however, was that subjects had easy access to a black male doctor once randomized to them. Several studies report that minority doctors are more likely than white doctors to work in underserved areas and see patients who share their racial background (Moy and Bartman 1995; Komaromy et al. 1996; Cantor et al. 1996; Walker, Moreno, Grumbach 2012). Yet despite this allocation, there remains a difference in access. Returning to our non-experimental evidence in Figure 5, by far the largest divide between black and white male respondents is with regards to accessibility of a doctor who is of their same race and sex background (37% vs. 62%). In Table 8 Column (9), black male respondents were 26 percentage points less likely to respond that a black male doctor is available near them than white males report white male doctors are available, conditional on age, income, and education.[20]

As stated in the introduction, African-Americans comprise only 4% of practicing physicians in the U.S. Both African-American and Hispanic physicians are significantly underrepresented if comparing the ratios of the share of the recent medical school graduates to their share in the U.S. population. Non-Hispanic white physicians approach a ratio of one and Asian physicians approach a ratio of four (see Appendix Figure 8). Moreover, the pipeline of African-American medical school graduates is relatively flat — hovering around 6% for the last decade, an increasingly lower share of the African-American population (see Appendix Figure 9). This aspect of the study was also noted by one of the subjects: "Really excited about the black male doctors!!!"

VI. Health Valuation

In behavioral hazard models, individuals may underuse medical care due to misperceptions; thus the demand curve ceases to be a sufficient statistic for welfare calculations (Pauly and Blavin 2008; Baicker, Mullainathan, and Schwartzstein 2015; Chetty, Looney, and Kroft

[20]In the baseline survey, we asked how much choice individuals had in where they go for medical care — only 37% of respondents answered that they had a "great deal of choice."

2009). In addition, most of the preventives we offered were not cross-randomized with incentives. Thus, we value the effect of a black doctor in preventing cardiovascular-disease-related deaths using recently published medical studies (Kahn et al. 2010, Dehmer et al. 2017).[21] Since these models are concerned with the effect of screening on health, we combine their estimates with the coefficient on black doctor in the post-consultation stage.

We find that black doctors reduce myocardial infarctions by 1,072 per 100,000 and cardiovascular-related deaths by 622 per 100,000 (or 15.6 per year) for black men over about a 40-year time horizon. The difference in annual age-adjusted mortality rates for cardiovascular disease between non-Hispanic white (268.4 per 100,000) and non-Hispanic black males (350.3 per 100,000) in the U.S. is 81.9 per 100,000 (Murphy et al. 2017). Therefore, the treatment effect we estimate for black doctors could reduce this gap by approximately 19%. To the extent preventive services reduce ER visits, our intervention could translate into cost savings for hospitals as well (see Garthwaite, Gross, and Notowidigdo (2018)).

The difference in annual age-adjusted mortality rates for influenza and pneumonia between non-Hispanic white and non-Hispanic males in the United States is 2.7 per 100,000 (20.3 versus 17.6). Flu vaccination for adults over the age of 18 is estimated as averting 2.7 deaths per 100,000 per year (based on CDC 2016 and CDC 2017a). Multiplying the treatment effect of black doctors by the efficacy of flu vaccination to prevent flu deaths among adults, we obtain 0.27, which is roughly 10% of the gap in mortality for this cause of death.

Harper, Rushani, and Kaufman (2012) calculate that 41% of the life-expectancy gap between black and white males in 2008 was due to cardiovascular disease and diabetes. Therefore, our estimates of the black doctor treatment effect suggest the overall life-expectancy gap between black and white males exclusive of infant mortality could be reduced by approximately 8% or 5 months from cardiovascular disease and diabetes alone. If we extrapolate the screening benefit to other preventable leading causes of death and health disparities among African-American men (i.e. HIV and cancer), the life expectancy gain could be even larger since these chronic illnesses account for another 26% of the black-white male life expectancy gap. Certain types of cancer or cancer-related deaths can be prevented through care and treatment adherence (e.g. HPV vaccine, tobacco cessation, earlier stage diagnoses).

[21]Both Kahn et al. and Dehmer et al. perform a Monte-Carlo simulation on a representative U.S. population to compare screening to a no screening condition, and assume that those who screen positive receive guideline-recommended therapy. Since both studies were published relatively recently, treatment efficacy is likely to reflect the current state of care, though varying the fraction of screen-positive who obtain and follow appropriate treatment recommendations will alter the results, particularly if this fraction also interacts with doctor race. The Dehmer et al. study assumes only 90% of those offered screening take it up, thus we divide by 0.9 to make the results consistent with the Kahn et al. study. The Dehmer et al. study also provides estimates of the effects of screening subdivided by race and sex. Such stratification is not available in Kahn et al. Further details on the studies and the calculation can be found in the Online Appendix.

The assumption that all who screen positive receive appropriate care is an upper bound on the marginal impact of the screening effect of black doctors on lives saved. However, a more realistic assumption of leakage (or lack of compliance) conditional on screen positive exams could deliver even larger *differential* effects of doctor race on lives saved since compliance with recommendations conditional on having a disease might also vary with concordance (Traylor et al. 2010).

VII. Conclusion

In this study, we examine the effect of diversity of the physician workforce on the demand for preventive care among African-American men using a randomized trial. We find that, when patients and doctors had an opportunity to meet in person, patients assigned to a black doctor increased their demand for preventives, particularly those which were invasive. These findings were stronger among subjects who had limited prior experience with routine medical care. Data from the clinical encounter demonstrate that subjects brought up more issues and were more likely to seek advice from black doctors, as reflected in the doctors' notes.

These findings are consistent with a framework in which agents underestimate the benefit of preventive care, and thus have low demand. Physicians, through their counseling and rapport with patients, which varies by social distance, can help correct false beliefs and increase demand. Subsidies also increase demand, though we find financial incentives do not completely substitute for information from a trusted source. Some subjects who selected flu shots initially, encouraged by the incentive, declined to actually receive them (often citing contraindications). Moreover, black doctors continued to increase demand even among subjects who initially refused a flu shot despite a financial incentive.

Our back of the envelope calculations suggest the increased demand induced by black doctors could reap substantial health benefits. Specifically, we calculate that increased screening could lead to a 19% reduction in the black-white male cardiovascular mortality gap and an 8% decline in the black-white male life expectancy gap. Given the current supply of black doctors, a more diverse physician workforce might be necessary to realize these gains.

References

[1] Abadie, Alberto, Susan Athey, Guido W. Imbens, Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?" NBER Working Paper, 24003.

[2] Agency for Healthcare Research and Quality. 2016. "The Challenges of Measuring Physician Quality." Accessed June, 2018. https://www.ahrq.gov/professionals/quality-patient-safety/talkingquality/create/physician/challenges.html

[3] Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics* 115(3): 715–53.

[4] Alsan, Marcella, and Marianne Wanamaker. 2018. "Tuskegee and the Health of Black Men." *Quarterly Journal of Economics* 133(1): 407–55.

[5] American Medical Association. 2018. "Reducing Disparities in Health Care." Accessed June, 2018. https://www.ama-assn.org/delivering-care/reducing-disparities-health-care.

[6] Arias, Elizabeth, Melonie Heron, and Jiaquan Xu. 2017. "United States Life Tables, 2014." National Vital Statistics Reports 66(4): 1–63.

[7] Arnett, M. J., R.J. Thorpe, D. J. Gaskin, J. V. Bowie, and T. A. LaVeist. 2016. "Race, Medical Mistrust, and Segregation in Primary Care as Usual Source of Care: Findings from the Exploring Health Disparities in Integrated Communities Study." *Journal of Urban Health* 93(3): 456–67.

[8] Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53(5): 941–73.

[9] Association of American Medical Colleges. 2009. "Addressing Racial Disparities in Health Care: A Targeted Action Plan for Academic Medical Centers."

[10] Association of American Medical Colleges. 2014. "Diversity in the Physician Workforce: Facts and Figures 2014."

[11] Association of American Medical Colleges. 2015. "Altering the Course: Black Males in Medicine."

[12] Association of American Medical Colleges. 2016. "Diversity in Medical Education: Facts and Figures 2016."

[13] Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral Hazard in Health Insurance." *Quarterly Journal of Economics* 130(4): 1623–67.

[14] Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Dhruva Kothari. 2010. "Improving Immunisation Coverage in Rural India: Clustered Randomised Controlled Evaluation of Immunisation Campaigns with and without Incentives." *BMJ* 340: 1–9.

[15] Banerjee, Abhijit, and Esther Duflo. 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty.* 41–70. New York: PublicAffairs.

[16] Becker, Gary. 1957. *The Economics of Discrimination.* Chicago: University of Chicago Press.

[17] Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects" *Journal of Economic Perspectives*, 28(2): 29–50.

[18] Benjamin, Daniel J., James J. Choi, and A. Joshua Strickland. 2010. "Social Identity and Preferences." *American Economic Review* 100(4): 1913–28.

[19] Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119(1): 249–75.

[20] Bertrand, Marianne, Sandra E. Black, Sissel Jensen, and Adriana Lleras-Muney. 2018. "Breaking the Glass Ceiling? The Effect of Board Quotas on Female Labour Market Outcomes in Norway." *Review of Economic Studies* 0: 1–48.

[21] Bettinger, Eric P., and Bridget T. Long. 2005. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." *American Economic Review: Papers and Proceedings* 95(2): 152–7.

[22] Blewett, Lynn A., Julia A. Rivera Drew, Risa Griffin, Miriam King, and Kari C.W. Williams. 2018a. IPUMS Health Surveys: National Health Interview Survey, Version 6.3 [National Health Interview Survey, 2007–2017]. Minneapolis: University of Minnesota.

[23] Blewett, Lynn A., Julia A. Rivera Drew, Risa Griffin, Kari C.W. Williams, and Daniel Backman. 2018b. IPUMS Health Surveys: Medical Expenditure Panel Survey, Version 1.0 [Medical Expenditure Panel Survey, 2005–2015]. Minneapolis: University of Minnesota.

[24] Braunstein, Joel B., Noëlle S. Sherber, Steven P. Schulman, Eric L. Ding, and Neil R. Powe. 2008. "Race, Medical Researcher Distrust, Perceived Harm, and Willingness to Participate in Cardiovascular Prevention Trials." *Medicine* 87(1): 1–9.

[25] Brown, Lindsay E., Ryan Burton, Brian Hixon, Manasi Kakade, Parul Bhagalia, Catherine Vick, Andrew Edwards, and Mary T. Hawn. 2012. "Factors Influencing Emergency Department Preference for Access to Healthcare." *Western Journal of Emergency Medicine* 13(1): 410–5.

[26] Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50(2): 317–72.

[27] Cantor, Joel C., Erika L. Miles, Laurence C. Baker, and Dianne C. Barker. 1996. "Physician Service to the Underserved: Implications for Affirmative Action in Medical Education." *Inquiry* 33(2): 167–80.

[28] Carrell, Scott E., Marianne E. Page, and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *Quarterly Journal of Economics* 125(3): 1101–44.

[29] Centers for Disease Control and Prevention. 2014. "Up to 40 Percent of Annual Deaths from Each of Five Leading US Causes are Preventable." https://www.cdc.gov/media/releases/2014/p0501-preventable-deaths.html.

[30] Centers for Disease Control and Prevention. 2016. "Flu Vaccination Coverage, United States, 2015-16 Influenza Season." https://www.cdc.gov/flu/fluvaxview/coverage-1516estimates.htm.

[31] Centers for Disease Control and Prevention. 2017a. "Estimated Influenza Illnesses, Medical Visits, Hospitalizations, and Deaths Averted by Vaccination in the United States." https://www.cdc.gov/flu/about/disease/2015-16.htm.

[32] Centers for Disease Control and Prevention. 2017b. "National Diabetes Statistics Report, 2017: Estimates of Diabetes and Its Burden in the United States." https://www.cdc.gov/diabetes/data/statistics-report/index.html.

[33] Centers for Disease Control and Prevention. 2017c. "High Cholesterol Facts." https://www.cdc.gov/cholesterol/facts.htm.

[34] Chandra, Amitabh, and Douglas Staiger. 2010. "Identifying Provider Prejudice in Healthcare" NBER Working Paper, 16382.

[35] Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector." *American Economic Review* 106(8): 2110–44.

[36] Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99(4): 1145–77.

[37] Coffman, Lucas, and Paul Niehaus. 2018. "Pathways of Persuasion." Working Paper.

[38] Cohen, Jessica, and Pascaline Dupas. 2010. "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment." *Quarterly Journal of Economics* 125(1): 1–45.

[39] Das, Jishnu, Alaka Holla, Aakash Mohpal, and Karthik Muralidharan. 2016. "Quality and Accountability in Health Care Delivery: Audit-Study Evidence from Primary Care in India." *American Economic Review* 106(12): 3765–99.

[40] Dee, Thomas. 2004. "Teachers, Race, and Student Achievement in a Randomized Experiment." *Review of Economics and Statistics* 86(1): 195–210.

[41] Dee, Thomas. 2005. "A Teacher like Me: Does Race, Ethnicity, or Gender Matter?" *American Economic Review* 95(2): 158–65.

[42] Dehmer, Steven P., Michael V. Maciosek, Amy B. LaFrance, and Thomas J. Flottemesch. 2017. "Health Benefits and Cost-Effectiveness of Asymptomatic Screening for Hypertension and High Cholesterol and Aspirin Counseling for Primary Prevention." *Annals of Family Medicine* 15(1): 23–36.

[43] DellaVigna, Stefano, and Ethan Kaplan. 2007. "The Fox News Effect: Media Bias and Voting" *Quarterly Journal of Economics* 122(3): 1187–234.

[44] DellaVigna, Stefano, and Matthew Gentzkow. 2010. "Persuasion: Empirical Evidence" *Annual Review of Economics* 2: 643–69.

[45] Duflo, Esther. 2018. "Machinistas Meet Randomistas: Useful ML Tools for Empirical Researchers." Lecture, NBER Summer Institute, Cambridge, MA.

[46] Dupas, Pascaline. 2011. "Health Behavior in Developing Countries." *Annual Review of Economics* 3: 425–49.

[47] Ehrenberg, Ronald G., Daniel D. Goldhaber, and Dominic J. Brewer. 1995. "Do Teachers' Race, Gender, and Ethnicity Matter? Evidence From the National Education Longitudinal Study of 1988." *Industrial and Labor Relations Review* 48(3): 547–61.

[48] Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos. 2014. "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom." *American Economic Review* 104(8): 2567–91.

[49] Fryar, Cheryl D., Yechiam Ostchega, Craig M. Hales, Guangyu Zhang, and Deanna Kruszon-Moran. 2017. "Hypertension Prevalence and Control Among Adults: United States, 2015–2016." NCHS Data Brief, 289.

[50] Garcia, Luis C., Sukyung Chung, Lily Liao, Jonathan Altamirano, Magali Fassiotto, Bonnie Maldonado, Paul Heidenreich, et al. 2019. "Comparison of Outpatient Satisfaction Survey Scores for Asian Physicians and Non-Hispanic White Physicians." *JAMA Network Open* 2(2): 1–10.

[51] Garthwaite, Craig, Tal Gross, and Matthew J. Notowidigdo. 2018. "Hospitals as Insurers of Last Resort." *American Economic Journal: Applied Economics* 10(1): 1–39.

[52] Giuliano, Laura, David I. Levine, and Jonathan Leonard. 2009. "Manager Race and the Race of New Hires." *Journal of Labor Economics* 27(4): 589–631.

[53] Glover, Dylan, Amanda Pallais, and William Pariente. 2017. "Discrimination as a Self-Fulfilling Prophecy: Evidence from French Grocery Stores." *Quarterly Journal of Economics* 132(3): 1219–60.

[54] Gostin, Lawrence, James G. Hodge Jr., Nicole B. Valentine, and Helena Nygren-Krug. 2003. "The Domains of Health Responsiveness: A Human Rights Analysis." World Health Organization, EIP Discussion Paper 53.

[55] Greenwood, Brad N., Seth Carnahan, and Laura Huang. 2018. "Patient–Physician Gender Concordance and Increased Mortality among Female Heart Attack Patients." *PNAS* 115(34): 8569–74.

[56] Hales, Craig M., Margaret D. Carroll, Cheryl D. Fryar, and Cynthia L. Ogden. 2017. "Prevalence of Obesity Among Adults and Youth: United States, 2015–2016." NCHS Data Brief, 288.

[57] Hammond, Wizdom P., Derrick Matthews, Dinushika Mohottige, Amma Agyemang, and Giselle Corbie-Smith. 2010. "Masculinity, Medical Mistrust, and Preventive Health Services Delays Among Community-Dwelling African-American Men." *Journal of General Internal Medicine* 25(12): 1300–8.

[58] Harper, Sam, Dinela Rushani, and Jay S. Kaufman. 2012. "Trends in the Black-White Life Expectancy Gap, 2003–2008." *JAMA* 307(21): 2257–9.

[59] Heß, Simon. 2017. "Randomization Inference with Stata: A Guide and Software." *Stata Journal* 17(3): 630–51.

[60] Hill, Andrew, Daniel Jones, and Lindsay Woodworth. 2018. "A Doctor Like Me: Physician-Patient Race-Match and Patient Outcomes." Working Paper.

[61] Hjort, Jonas. 2014. "Ethnic Divisions and Production in Firms." *Quarterly Journal of Economics* 129(4): 1899–1946.

[62] Institute of Medicine. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: The National Academies Press.

[63] Kahn, Richard, Peter Alperin, David Eddy, Knut Borch-Johnsen, John Buse, Justin Feigelman, Edward Gregg, et al. 2010. "Age at Initiation and Frequency of Screening to Detect Type 2 Diabetes: A Cost-Effectiveness Analysis." *Lancet* 375: 1365–74.

[64] Kinlock, Ballington, Lauren J. Parker, Janice V. Bowie, Daniel L. Howard, Thomas A. LaVeist, and Roland J. Thorpe Jr. 2017. "High Levels of Medical Mistrust Are Associated With Low Quality of Life Among Black and White Men With Prostate Cancer." *Cancer Control* 24(1): 72–7.

[65] Komaromy, Miriam, Kevin Grumbach, Michael Drake, Karen Vranizan, Nicole Lurie, Dennis Keane, and Andrew B. Bindman. 1996. "The Role of Black and Hispanic Physicians in Providing Health Care for Underserved Populations." *New England Journal of Medicine* 334(20): 1305–10.

[66] Ku, Leighton, Bianca K. Frogner, Erika Steinmetz, and Patricia Pittman. 2015. "Community Health Centers Employ Diverse Staffing Patterns, Which Can Provide Productivity Lessons for Medical Practices." *Health Affairs* 34(1): 95–103.

[67] Lowes, Sara, and Eduardo Montero. 2018. "The Legacy of Colonial Medicine in Central Africa." Working Paper.

[68] Lusher, Lester, Doug Campbell, and Scott Carrell. 2018. "TAs Like Me: Racial Interactions between Graduate Teaching Assistants and Undergraduates." *Journal of Public Economics* 159: 203–24.

[69] Meghani, Salimah H., Jacqueline M. Brooks, Trina Gipson-Jones, Roberta Waite, Lisa Whitfield-Harris, and Janet A. Deatrick. 2009. "Patient–Provider Race-Concordance: Does It Matter in Improving Minority Patients' Health Outcomes?" *Ethnicity and Health* 14(1): 107–30.

[70] Moy, Ernest, and Barbara Bartman. 1995. "Physician Race and Care of Minority and Medically Indigent Patients." *JAMA* 273: 1515–20.

[71] Murphy, Sherry L., Jiaquan Xu, Kenneth D. Kochanek, Sally C. Curtin, and Elizabeth Arias. 2017. "Deaths: Final Data for 2015." National Vital Statistics Reports, 66(6).

[72] Murthy, Vivek H., Harlan M. Krumholz, and Cary P. Gross. 2004. "Participation in Cancer Clinical Trials: Race-, Sex-, and Age-Based Disparities." *JAMA* 291(22): 2720–6.

[73] Nanna, Michael, Ann Marie Navar, Pearl Zakroysky, Qun Xiang, Anne C. Goldberg, Jennifer Robinson, Veronique L. Roger, et al. 2018. "Association of Patient Perceptions of Cardiovascular Risk and Beliefs on Statin Drugs With Racial Differences in Statin Use: Insights From the Patient and Provider Assessment of Lipid Management Registry." *JAMA Cardiology* (published online).

[74] National Medical Association. 2018. "Major Minority Physician Associations Come Together." Accessed June, 2018. https://www.nmanet.org/news/398042/Major-Minority-Physician-Associations-Come-Together.htm.

[75] Pauly, Mark V., and Fredric E. Blavin. 2008. "Moral Hazard in Insurance, Value-Based Cost Sharing, and the Benefits of Blissful Ignorance." *Journal of Health Economics* 27: 1407–17.

[76] Pauly, Mark V., and Fredric E. Blavin. 2008. "Moral Hazard in Insurance, Value-Based Cost Sharing, and the Benefits of Blissful Ignorance." *Journal of Health Economics* 27: 1407–17.

[77] Pettey, Christina M., Jean C. McSweeney, Katharine E. Stewart, Mario A. Cleves, Elvin T. Price, Seongkum Heo, and Elaine Souder. 2016. "African Americans' Perceptions of Adherence to Medications and Lifestyle Changes Prescribed to Treat Hypertension." *Sage Open* 6(1).

[78] Ruggles, Steven, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. 2017. Integrated Public Use Microdata Series: Version 7.0 [American Community Survey, 2016]. Minneapolis: University of Minnesota.

[79] Scharff, Darcell P., Katherine J. Mathews, Pamela Jackson, Jonathan Hoffsuemmer, Emeobong Martin, and Dorothy Edwards. 2010. "More than Tuskegee: Understanding Mistrust about Research Participation." *Journal of Health Care for the Poor and Underserved* 21(3): 879–97.

[80] Shiyanbola, Olayinka O., Carolyn M Brown, and Earlise C Ward. 2018. " 'I Did Not Want to Take That Medicine': African-Americans' Reasons for Diabetes Medication Nonadherence and Perceived Solutions for Enhancing Adherence." *Patient Preference and Adherence* 12: 409–21.

[81] Silber, Jeffrey H., Paul R. Rosenbaum, Richard N. Ross, Bijan A. Niknam, Justin M. Ludwig, Wei Wang, Amy S. Clark, et al. 2014. "Racial Disparities in Colon Cancer Survival: A Matched Cohort Study." *Annals of Internal Medicine* 161(12): 845–54.

[82] Simon, Kosali K., Angshuman Gooptu, Seth Freedman, and Victoria Perez. 2016. "Securing the Right to Health Care and Well Being: Ten Years Later." In *The Covenant with Black America: Ten Years Later*. Edited by Tavis Smiley. Smiley Books.

[83] Sotto-Santiago, Sylk, James E. Slaven, and Theresa Rohr-Kirchgraber. 2019. "(Dis)Incentivizing Patient Satisfaction Metrics: The Unintended Consequences of Institutional Bias." *Health Inequity* 3(1): 13–8.

[84] Stoll, Michael A., Stephen Raphael, and Harry J. Holzer. 2004. "Black Job Applicants and the Hiring Officer's Race." *Industrial and Labor Relations Review* 57(2): 267–287.

[85] Tabellini, Guido. 2008. "The Scope of Cooperation: Values and Incentives." *Quarterly Journal of Economics* 123(3): 905–50.

[86] Thirumurthy, Harsha, David A. Asch, and Kevin G.Volpp. 2019. "The Uncertain Effect of Financial Incentives to Improve Health Behaviors" *JAMA* 321(15): 1451–2.

[87] Thornton, Rebecca L. 2008. "The Demand for, and Impact of, Learning HIV Status." *American Economic Review* 98(5): 1829–63.

[88] Traylor, Ana H., Julie A. Schmittdiel, Connie S. Uratsu, Carol M. Mangione, and Usha Subramanian. 2010. "Adherence to Cardiovascular Disease Medications: Does Patient-Provider Race/Ethnicity and Language Concordance Matter?" *Journal of General Internal Medicine* 25(11): 1172–7.

[89] Walker, Kara O., Gerardo Moreno, and Kevin Grumbach. 2012. "The Association Among Specialty, Race, Ethnicity, and Practice Location Among California Physicians in Diverse Specialties." *Journal of the National Medical Association* 104(0): 46–52.

[90] Young, Richard A., Richard G. Roberts, and Richard J. Holden. 2017. "The Challenges of Measuring, Improving, and Reporting Quality in Primary Care." *Annals of Family Medicine* 15(2): 175–82.

[91] Zhou, Ruohua A., Katherine Baicker, Sarah Taubman, and Amy Finkelstein. 2017. "The Uninsured Do Not Use the Emergency Department More — They Use Other Care Less." *Health Affairs* 36(12): 2115–22.

Table 1: Participation in Experiment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	PANEL A: Health and Medical Care						
	Self-Reported Health	*Any Health Problem*	*Hospital Visits*	*ER Visits*	*Nights Hospital*	*Medical Mistrust*	*Has Primary MD*
Clinic Presentation	-0.126	0.033	0.244	0.513	-0.332	-0.011	-0.072
	(0.025)	(0.028)	(0.469)	(0.183)	(0.746)	(0.042)	(0.029)
Mean	0.81	0.57	4.74	1.24	1.93	1.64	0.69
Observations	1,148	1,241	935	1,031	1,041	1,232	1,096
	PANEL B: Socio-demographics						
	Uninsured	*Age*	*Married*	*Unemployed*	*≤ High School Education*	*Low Income*	*SSI/DI/UI*
Clinic Presentation	0.038	3.411	-0.053	0.129	0.190	0.198	0.113
	(0.027)	(0.811)	(0.022)	(0.025)	(0.029)	(0.027)	(0.024)
Mean	0.24	41.06	0.20	0.18	0.44	0.25	0.18
Observations	1,074	1,241	1,201	1,176	1,141	1,171	1,198

Note: Table reports results from a regression of various baseline characteristics on clinic presentation. Observation count varies due to missing responses in the baseline survey. Reported mean is among subjects that did not present to the clinic. See Data Appendix for other variable definitions. Robust standard errors in parentheses.

Table 2: Balance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean (S.D.)	*Non-Black MD - $5*	*Non-Black MD - $10*	*Black MD - $0*	*Black MD - $5*	*Black MD - $10*	*p-value*	*N*
Self-Reported Health	0.72	-0.033	-0.181	0.007	-0.016	0.004	0.067	563
	[0.45]	(0.066)	(0.067)	(0.065)	(0.064)	(0.063)		
Any Health Problem	0.62	-0.026	0.036	-0.015	-0.025	-0.021	0.940	614
	[0.49]	(0.068)	(0.065)	(0.069)	(0.067)	(0.066)		
ER Visits	1.69	-0.149	0.867	-0.212	0.145	-0.391	0.247	511
	[3.54]	(0.434)	(0.609)	(0.443)	(0.558)	(0.419)		
Nights Hospital	1.20	-0.392	0.839	1.956	-0.214	0.230	0.249	511
	[3.52]	(0.415)	(0.734)	(1.490)	(0.466)	(0.663)		
Medical Mistrust	1.61	0.162	-0.046	0.032	0.016	-0.034	0.430	611
	[0.74]	(0.105)	(0.100)	(0.105)	(0.105)	(0.100)		
Has Primary MD	0.63	-0.042	0.033	-0.059	0.008	-0.019	0.838	537
	[0.49]	(0.074)	(0.070)	(0.073)	(0.070)	(0.071)		
Uninsured	0.22	0.042	0.146	0.112	0.057	0.010	0.223	517
	[0.42]	(0.066)	(0.067)	(0.070)	(0.064)	(0.062)		
Age	44.96	-1.051	-0.100	-0.261	-1.109	-0.495	0.990	620
	[14.76]	(1.973)	(2.001)	(1.982)	(2.048)	(1.944)		
Married	0.14	0.043	-0.037	0.069	-0.015	0.024	0.348	586
	[0.35]	(0.052)	(0.045)	(0.055)	(0.047)	(0.050)		
Unemployed	0.32	-0.045	-0.008	-0.051	0.008	0.025	0.853	570
	[0.47]	(0.066)	(0.066)	(0.065)	(0.065)	(0.065)		
≤ High School Education	0.62	0.006	-0.006	-0.029	0.055	0.034	0.886	556
	[0.49]	(0.070)	(0.070)	(0.072)	(0.068)	(0.068)		
Low Income	0.47	-0.026	-0.033	-0.043	0.022	-0.042	0.936	571
	[0.50]	(0.072)	(0.071)	(0.072)	(0.070)	(0.069)		
Attrition	0.03	0.022	0.045	0.031	0.015	-0.029	0.129	684
	[0.18]	(0.033)	(0.034)	(0.034)	(0.031)	(0.025)		

Note: Column (1) reports the mean and standard deviation. Columns (2)–(6) report regression coefficients and standard errors for each randomization group relative to the omitted group (Column (1), the non-black doctor and no incentive group). Column (7) shows the p-value associated with the F-statistic testing whether the treatment arms are jointly equal to zero. Observation count varies due to missing responses in the baseline survey. Attrition is an indicator for the 47 subjects that did not complete the study because they left before the clinic encounter (3 of the 50 subjects who attrited self-identified as a race/ethnicity other than African-American or as a female and are therefore excluded). See Data Appendix for other variable definitions. Robust standard errors in parentheses. Standard deviations in brackets.

Table 3: Pre-Consultation, Post-Consultation, and Delta Demand for Preventives

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Blood Pressure	*BMI*	*Diabetes*	*Cholesterol*	*Flu Vaccination*	*Share Four*	*Share of Invasives*
PANEL A: Pre-Consultation							
Black Doctor	0.025	0.023	0.050	0.010	-0.009	0.027	0.017
	{0.045}	{0.043}	{0.048}	{0.052}	{0.039}	{0.040}	{0.039}
$5 Incentive	0.028	-0.059	0.085	0.067	0.192	0.030	0.115
	{0.037}	{0.041}	{0.045}	{0.030}	{0.047}	{0.022}	{0.028}
$10 Incentive	-0.023	-0.009	0.028	-0.014	0.299	-0.004	0.104
	{0.037}	{0.035}	{0.028}	{0.039}	{0.031}	{0.020}	{0.027}
$Prob(\lvert\beta^{RI: Black Dr}\rvert > \lvert\beta^{Study Est.}\rvert)$	0.635	0.645	0.431	0.875	0.850	0.637	0.746
Control Mean	0.56	0.50	0.37	0.35	0.20	0.44	0.36
PANEL B: Post-Consultation							
Black Doctor	0.107	0.161	0.204	0.256	0.100	0.182	0.186
	{0.074}	{0.099}	{0.062}	{0.071}	{0.037}	{0.064}	{0.046}
$5 Flu Incentive	0.044	0.019	0.110	0.065	0.221	0.059	0.132
	{0.036}	{0.056}	{0.047}	{0.036}	{0.039}	{0.032}	{0.030}
$10 Flu Incentive	-0.026	-0.010	0.054	-0.004	0.219	0.003	0.090
	{0.038}	{0.028}	{0.045}	{0.040}	{0.026}	{0.028}	{0.028}
$Prob(\lvert\beta^{RI: Black Dr}\rvert > \lvert\beta^{Study Est.}\rvert)$	0.251	0.220	0.039	0.023	0.047	0.036	0.017
Control Mean	0.72	0.60	0.42	0.36	0.18	0.53	0.37
PANEL C: Delta (Post - Pre)							
Black Doctor	0.082	0.138	0.154	0.246	0.108	0.155	0.169
	{0.100}	{0.101}	{0.059}	{0.072}	{0.050}	{0.077}	{0.051}
$5 Flu Incentive	0.017	0.078	0.024	-0.002	0.029	0.029	0.017
	{0.052}	{0.043}	{0.022}	{0.029}	{0.035}	{0.026}	{0.016}
$10 Flu Incentive	-0.003	-0.001	0.026	0.010	-0.080	0.008	-0.015
	{0.028}	{0.028}	{0.047}	{0.055}	{0.030}	{0.028}	{0.039}
$Prob(\lvert\beta^{RI: Black Dr}\rvert > \lvert\beta^{Study Est.}\rvert)$	0.468	0.277	0.051	0.005	0.098	0.116	0.016
Control Mean	0.16	0.11	0.05	0.01	-0.02	0.08	0.01
Observations	637	637	637	637	637	637	637

Note: Table reports OLS estimates of Equation 1. The outcome varies by column heading. Control mean refers to subjects randomized to a non-black doctor for the non-flu screenings and to subjects randomized to a non-black doctor and no incentive for the flu vaccination. Robust standard errors clustered at the doctor level in curly brackets. *Prob* refers to the randomization inference p-value from permuting doctor race for each of the 3,003 $\binom{14}{6}$ combinations.

Table 4: Heterogeneity by Invasiveness of Preventive Service

	(1)	(2)	(3)
	Pre	*Post*	*Delta*
Body Mass Index	-0.062	-0.114	-0.052
	{0.032}	{0.025}	{0.026}
Cholesterol	-0.210	-0.358	-0.148
	{0.019}	{0.077}	{0.076}
Diabetes	-0.185	-0.293	-0.108
	{0.027}	{0.064}	{0.061}
Flu	-0.213	-0.395	-0.182
	{0.030}	{0.053}	{0.057}
Black MD * BP	0.022	0.103	0.081
	{0.044}	{0.075}	{0.099}
Black MD * BMI	0.017	0.157	0.140
	{0.043}	{0.097}	{0.102}
Black MD * Cho	0.008	0.254	0.245
	{0.052}	{0.071}	{0.072}
Black MD * Dia	0.051	0.205	0.154
	{0.048}	{0.061}	{0.058}
Black MD * Flu	0.002	0.109	0.107
	{0.040}	{0.039}	{0.051}
*Black Doctor * Invasive Test Interactions = 0*	1.257	7.434	4.364
Prob($F^{RI} > F^{Study}$)	0.482	0.058	0.151
Sum of Invasives = Sum of Non-Invasives	0.054	4.366	9.921
Prob($F^{RI} > F^{Study}$)	0.829	0.088	0.008
Observations	3,185	3,185	3,185

Note: Table reports OLS estimates. The outcome variable is individual test take-up by stage. The omitted category is blood pressure. Robust standard errors clustered at the doctor level in curly brackets. Indicators for incentive levels are included but not reported. The two F-tests test the joint significance of the black doctor*invasive interactions *(Black Doctor * Invasive Test Interactions = 0)* as well as the differential effects of the black doctor treatment for invasive exams *(Sum of Invasives = Sum of Non-Invasives)*. The cholesterol, diabetes, and flu tests are considered invasive as they require a blood draw or an injection. *Prob* indicates randomization inference p-value based on F-tests using the 3,003 $\binom{14}{6}$ combinations of doctors.

Table 5: Demand for Preventives with Alternative Concordance Measures

	(1)	(2)	(3)	(4)	(5)	(6)
X =	*Age, 5 Years*		*Age, 10 Years*		*Education*	
X	0.008	-0.005	0.015	-0.019	0.002	-0.024
	{0.030}	{0.039}	{0.026}	{0.034}	{0.052}	{0.098}
X * Black Doctor		0.008		0.037		-0.018
		{0.053}		{0.048}		{0.118}
Black Doctor		0.165		0.153		0.157
		{0.051}		{0.059}		{0.057}
Observations	620	620	620	620	556	556

Note: Table reports OLS estimates of Equation 1. The outcome is the delta share of the invasive screenings. Columns (1) and (2) explore age concordance (i.e. doctor and subject born within five years of each other), Columns (3) and (4) examine concordance within a wider age window (i.e. doctor and subject born within 10 years of each other), and Columns (5) and (6) explore concordance across educational attainment (i.e. subject has at least a bachelor of arts degree). Indicators for incentive levels are included but not reported. Robust standard errors clustered at the doctor level in curly brackets.

Table 6: Communication, Time Spent, and Satisfaction with Doctor

	(1)	(2)	(3)	(4)	(5)	(6)
	PANEL A: Communication					
Outcome =	*Subject Talk to MD*		*Doctor Notes about Subject*		*Non-Preventive Notes*	
Black Doctor * Invasive		0.115 {0.088}		0.167 {0.072}		0.100 {0.057}
Black Doctor	0.100 {0.150}	0.006 {0.132}	0.111 {0.130}	0.004 {0.159}	0.089 {0.076}	0.024 {0.105}
Invasive Test		0.039 {0.053}		-0.149 {0.070}		-0.079 {0.042}
Control Mean	0.35	0.35	0.32	0.32	0.08	0.08
Observations	637	637	637	637	637	637
	PANEL B: Time Spent and Satisfaction					
Outcome =	*Length Visit, Minutes*		*Subject Rating of Experience*		*Subject Recommend MD*	
Black Doctor * Invasive		0.399 {1.363}		-0.110 {0.140}		-0.015 {0.026}
Black Doctor	4.384 {1.730}	2.253 {1.307}	-0.019 {0.053}	0.049 {0.134}	-0.0005 {0.009}	0.006 {0.026}
Invasive Test		11.996 {1.234}		0.116 {0.125}		0.036 {0.022}
Control Mean	20.53	20.53	4.80	4.80	0.99	0.99
Observations	498	498	574	574	597	597

Note: Table reports OLS estimates from a modified version of Equation 1. Even columns include an interaction between black doctor and an indicator for whether the subject chose any invasive preventive service (cholesterol, diabetes, or flu). Indicators for incentive levels are included but not reported. See Data Appendix and text for variable definitions. Robust standard errors clustered at the doctor level in curly brackets.

Table 7: Heterogeneity by Demographics, Hassle Costs, and Medical Care Experience

	(1)	(2)	(3)
	Outcome = Delta Share Invasives		
	PANEL A: Demographics		
X =	*Low Income*	*≤ High School Education*	*Younger than 40*
Black Doctor * *X*	0.061 {0.091}	-0.043 {0.071}	0.020 {0.039}
X	0.031 {0.031}	0.044 {0.057}	0.014 {0.013}
Black Doctor	0.130 {0.037}	0.180 {0.037}	0.160 {0.051}
Observations	571	556	620
	PANEL B: Hassle Costs		
X =	*Long Wait Time*	*High Congestion*	*Long Commute*
Black Doctor * *X*	0.157 {0.050}	0.150 {0.034}	0.093 {0.077}
X	-0.033 {0.019}	-0.042 {0.012}	-0.020 {0.026}
Black Doctor	0.135 {0.049}	0.115 {0.058}	0.126 {0.055}
Observations	451	451	618
	PANEL C: Medical Care Experience		
X =	*No Recent Screening*	*ER Visits*	*Medical Mistrust*
Black Doctor * *X*	0.113 {0.044}	0.010 {0.006}	0.047 {0.032}
X	-0.003 {0.034}	-0.005 {0.004}	-0.009 {0.017}
Black Doctor	0.144 {0.047}	0.151 {0.050}	0.092 {0.061}
Observations	604	511	611

Note: Table reports OLS estimates from a modified version of Equation 1 including interactions between black doctor and certain baseline characteristics. The outcome variable for every specification is the delta in demand for the share of invasive preventives. Observation count varies due to missing responses in the baseline survey. Indicators for incentive levels are included but not reported. See Data Appendix and text for variable definitions. Robust standard errors clustered at the doctor level in curly brackets.

Table 8: Perceptions of Doctors among Black and White Male Respondents

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Quality			*Communication*			*Access*			*Communication vs. Quality*
	Which MD most qualified?			Which MD understands me?			Which MD available near me?			
	Black MD	White MD	Concordance	Black MD	White MD	Concordance	Black MD	White MD	Concordance	Concordance
Black Respondent	0.350		-0.055	0.531		-0.001	0.241		-0.255	-0.028
	(0.025)		(0.030)	(0.024)		(0.029)	(0.024)		(0.029)	{0.025}
White Respondent		0.273			0.479			0.175		
		(0.029)			(0.027)			(0.030)		
Communication										0.171
										{0.014}
Mean	0.11	0.27	0.54	0.12	0.19	0.69	0.11	0.43	0.62	0.54
R-squared	0.12	0.08	0.03	0.23	0.24	0.04	0.09	0.04	0.07	0.06
Observations	1,490	1,490	1,490	1,490	1,490	1,490	1,490	1,490	1,490	2,980

Note: Columns (1), (2), (4), (5), (7), and (8) report OLS estimates of Equation 2a, testing whether respondents have a preference for doctors of the same race with respect to three domains of healthcare: quality, communication, and access, respectively. Columns (3), (6), and (9) report OLS estimates of Equation 2b testing whether preference for own race varies across black and white respondents. Column (10) reports OLS estimates of Equation 2c comparing preference across domain and race. The comparison group mean is the average white respondents who prefer black doctors in Columns (1), (4), and (7); the average black respondents who prefer white doctors in Columns (2), (5), and (8); the average white respondents who prefer white doctors in Columns (3), (6), and (9); and the average white respondents who select concordance in regards to quality in Column (10). See Data Appendix and text for variable definitions. All specifications include categorical controls for age, education, and household income levels. Robust standard errors in parentheses for Columns (1)–(9). Robust standard errors clustered at the respondent level in curly brackets for panel analysis (Column (10)).

Table 9: Doctor Characteristics and Quality

	(1)	(2)	(3)	(4)
	PANEL A: Occupational Characteristics			
	Years of Experience	*Medical School Rank*	*Internist*	*Board Question Performance*
Black Mean	15.17	24.00	0.67	0.78
Non-Black Mean	12.25	11.00	1.00	0.83
p-value	0.74	0.85	0.09	0.66
	PANEL B: Persuasiveness and Patient Panel Characteristics			
	Persuade Black Men	*Persuade White Men*	*Most Comply*	*>5 Black Patients / Week*
Black Mean	0.50	0.33	0.50	0.67
Non-Black Mean	0.75	0.75	1.00	0.38
p-value	0.35	0.13	0.03	0.30
Observations	14	14	14	14

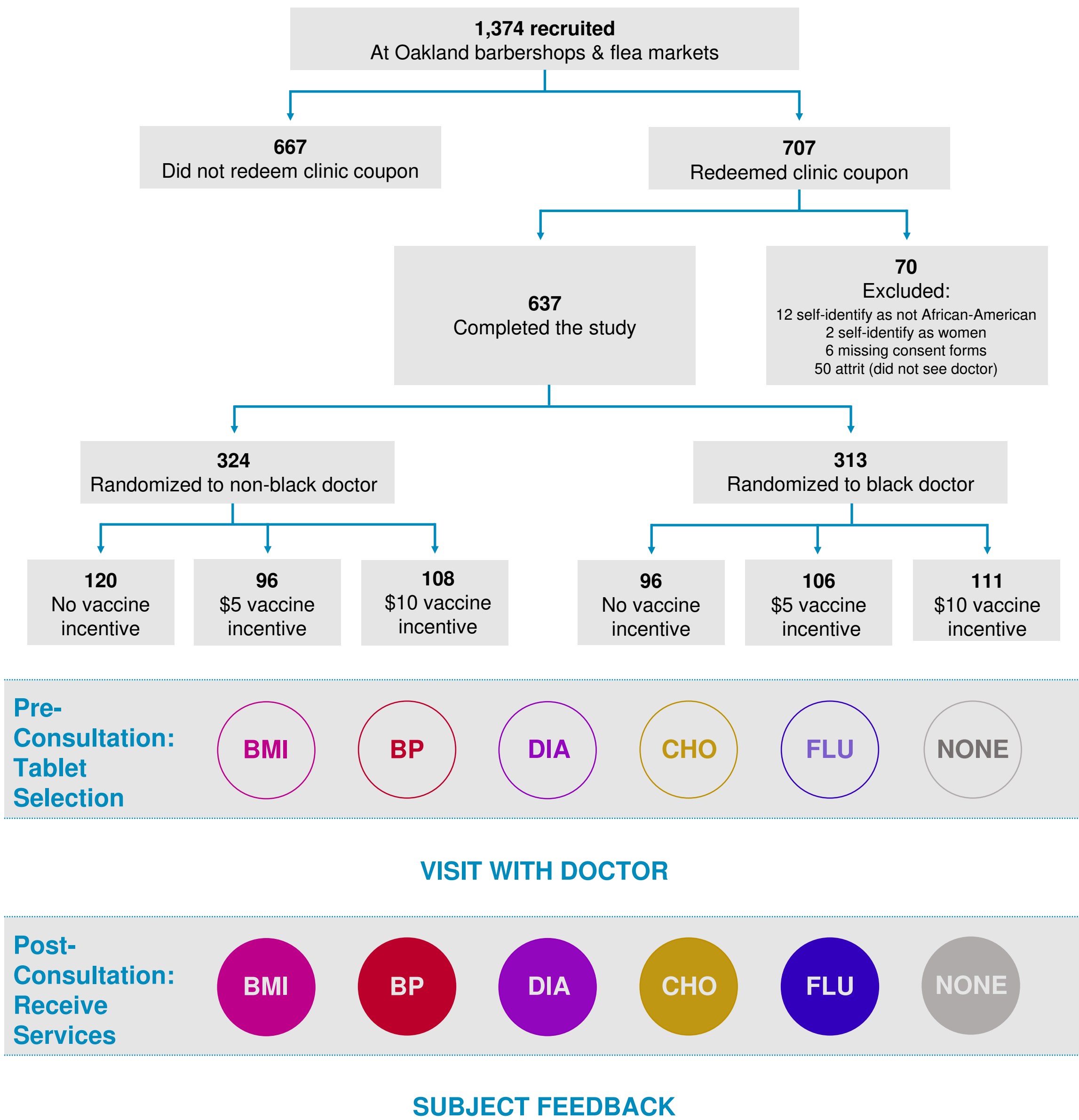
Note: Table reports mean doctor characteristics by race. See Data Appendix and text for variable definitions. Wilcoxon rank-sum test p-values are reported in row 3 and 6.

Table 10: Examining Doctor Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)
				Doctor Fixed Effects		
Black Doctor	0.142	0.135	0.143	0.137	0.151	0.177
	(0.049)	(0.043)	(0.054)	(0.050)	(0.069)	(0.068)
>5 Black Patients / Week		0.023				-0.037
		(0.039)				(0.073)
Top 10 Ranked Medical School			-0.005			-0.008
			(0.052)			(0.065)
Years of Experience				0.002		0.004
				(0.002)		(0.004)
Internist					0.028	0.102
					(0.071)	(0.079)
$Prob(\mid \beta^{RI:\ Black\ Dr} \mid > \mid \beta^{Study\ Est.} \mid)$	0.010	0.024	0.021	0.025	0.008	0.019
R-squared	0.445	0.455	0.445	0.472	0.451	0.516
Observations	14	14	14	14	14	14

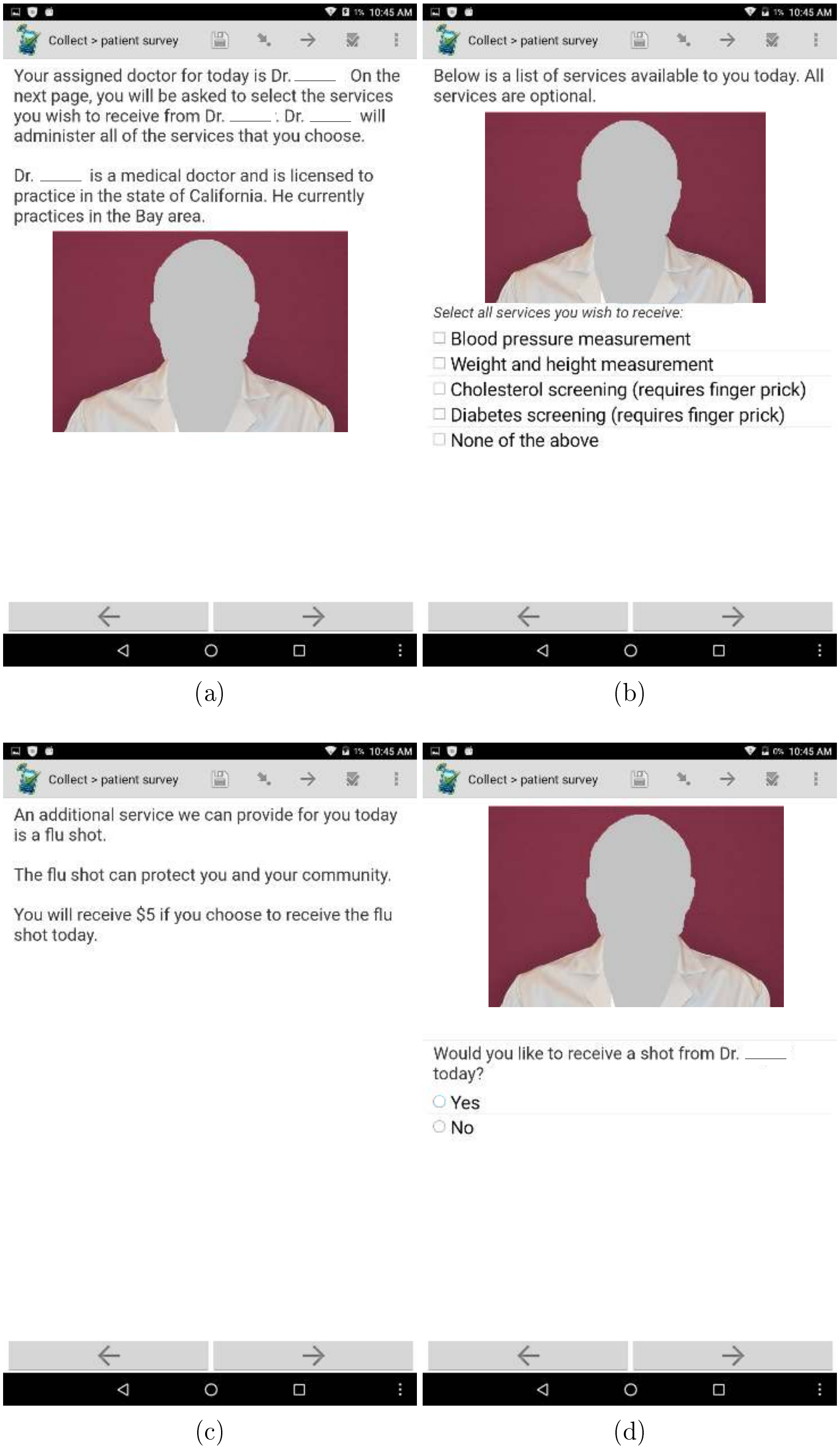
Note: Table reports OLS estimates. The outcome variable is the individual doctor fixed effects for the delta share invasives. Robust standard errors in parentheses. *Prob* indicates randomization inference p-value based on 3,003 $\binom{14}{6}$ combinations of doctors.

Figure 1: Study Design and Flow



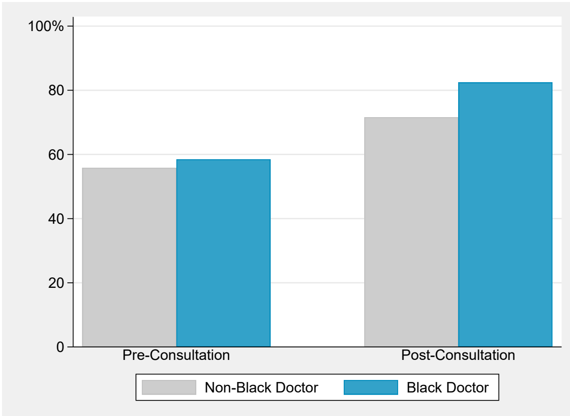
Note: Two-stage randomization design and flow of subjects from recruitment through clinic exit. Note that 70 subjects were randomized but are not included in the analysis study either because they did not meet criteria (i.e. they self-identified as a different race/ethnicity or as a female, were underage, or did not consent) or they left before the clinic encounter (i.e. attrited).

Figure 2: Tablet Photos

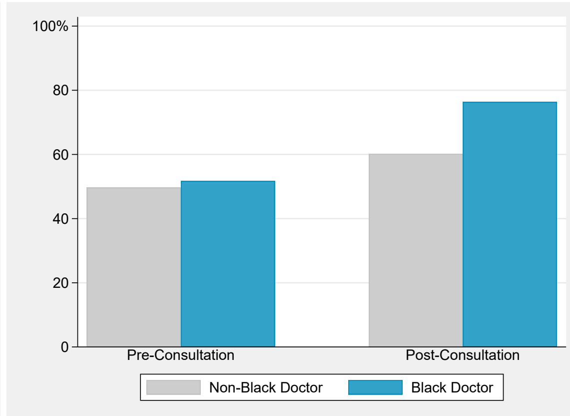


Note: Screenshots of clinic survey tablet: Panel (A) introduces subject's doctor; Panel (B) presents the non-incentivized screenings available (the order was randomized); Panel (C) informs the subject about the flu shot and associated incentive (if applicable); Panel (D) asks the subject whether he would like to receive a flu vaccination. Screenshots not shown to scale; tablet screen was approximately 10 inches.

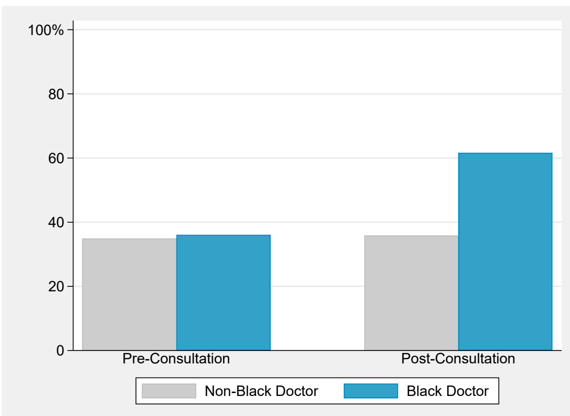
Figure 3: Demand for Preventives



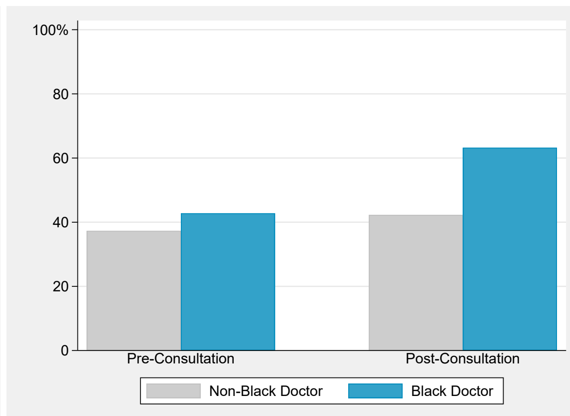
(a) Blood Pressure



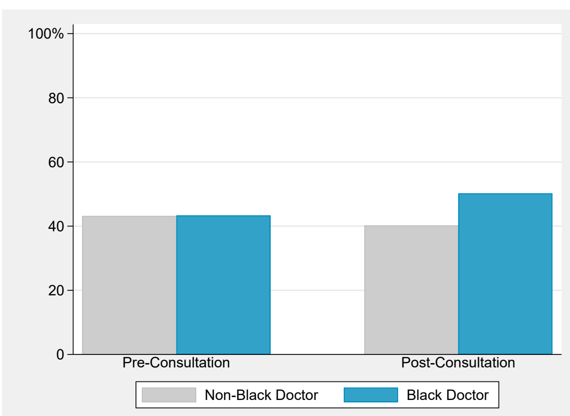
(b) BMI



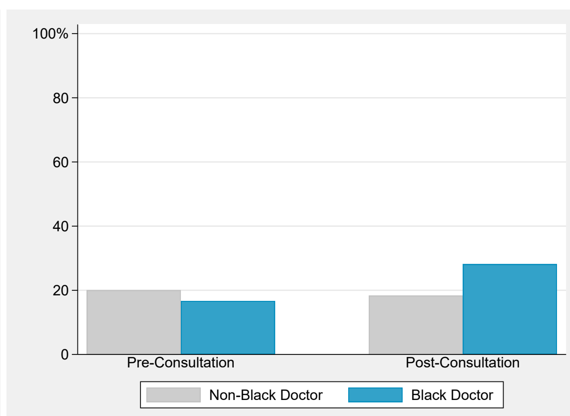
(c) Cholesterol



(d) Diabetes



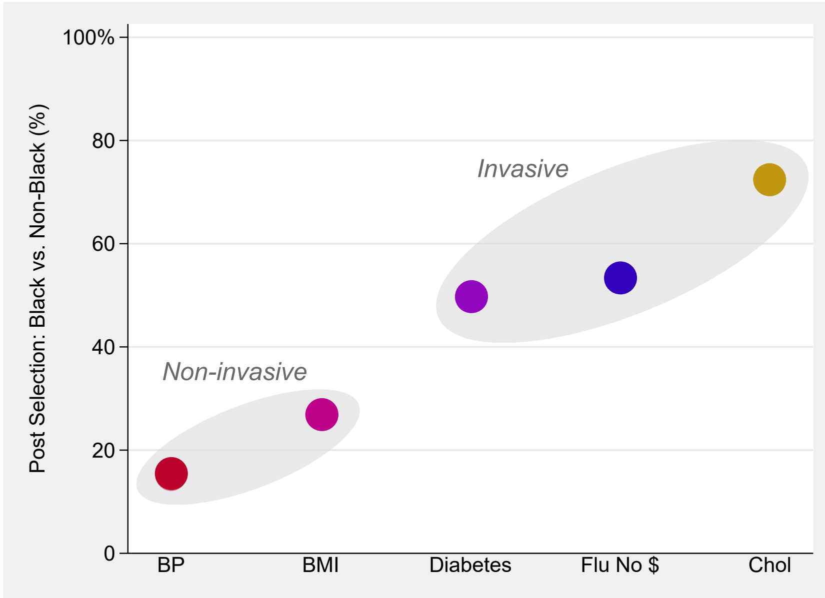
(e) Flu Shot: With Incentive



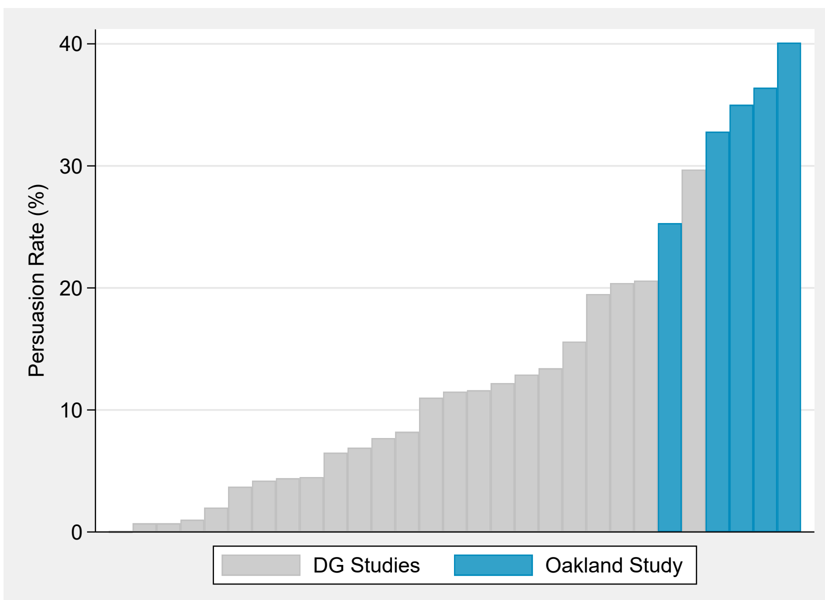
(f) Flu Shot: Without Incentive

Note: Pre- and post-consultation selection for preventives by randomized doctor race.

Figure 4: Post-Consultation Take-Up and Persuasion Rates by Preventive



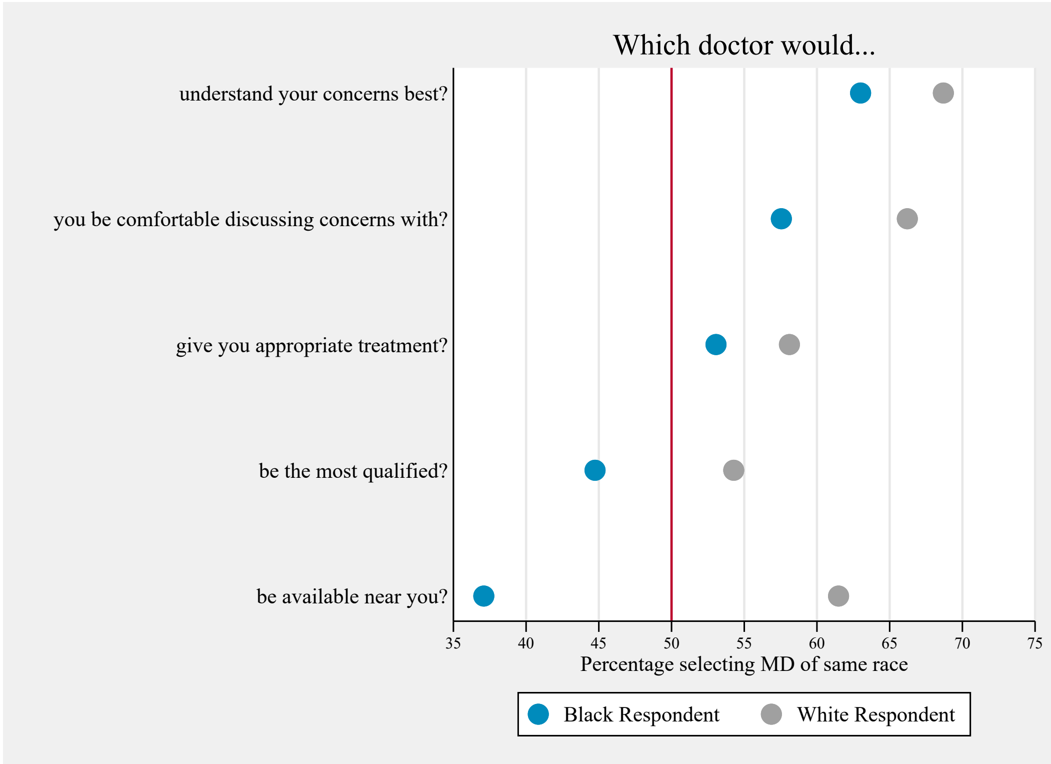
(a) Post % Differences by Preventives



(b) Persuasion Rates

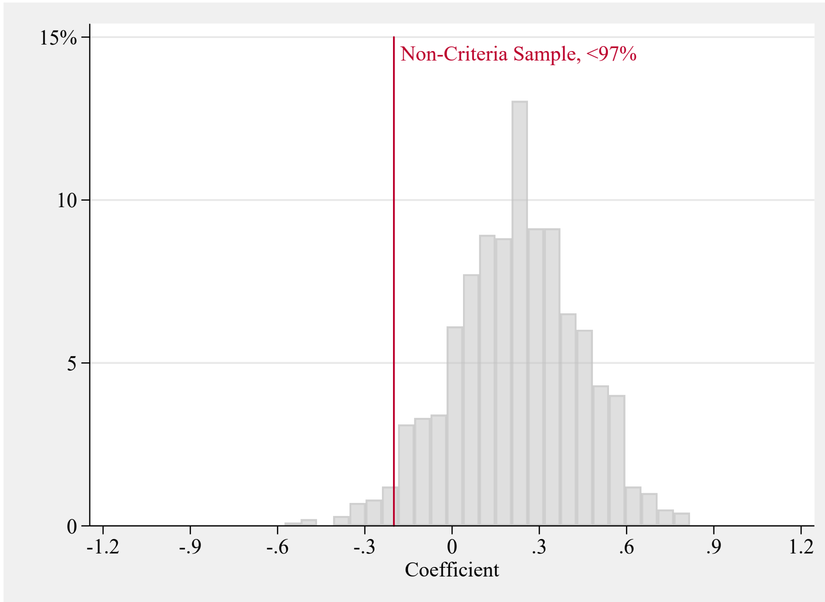
Note: Panel (A) plots the percent difference between black doctors vs. non-black doctors in post-consultation demand by preventive. Note that the percent difference in demand for the flu with an incentive (not shown) is equal to about 25%. The preventives are ordered by their y-axis value. Panel (B) plots persuasion rates (see text for more details). Each blue bar represents the persuasion rate for one of the five non-incentivized clinic screenings: from left to right, blood pressure, body mass index, diabetes, flu without an incentive, and cholesterol. Gray bars represent persuasion rates of studies from Table 1 of DellaVigna and Gentzkow (2010).

Figure 5: Non-Experimental Preference for Concordance

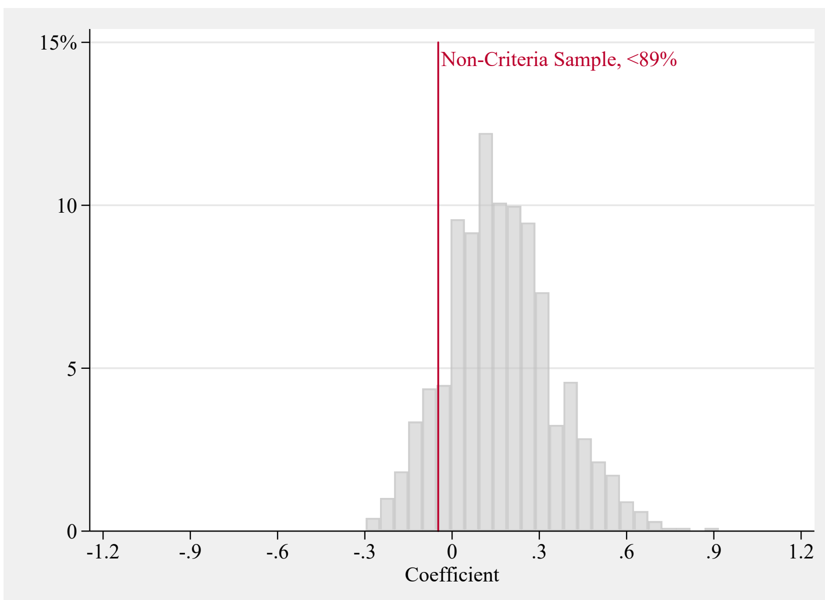


Note: Figure plots the percent of black and white survey respondents who select a doctor of the same race in response to various questions. Choice set included black, white, or Asian male doctors.

Figure 6: Permutation Test of Black Doctor Effect on Non-Criteria Sample



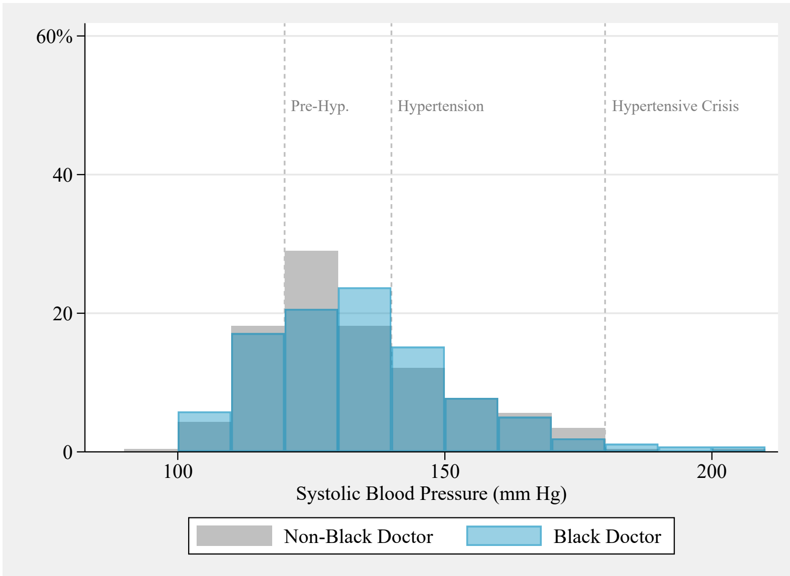
(a) Post



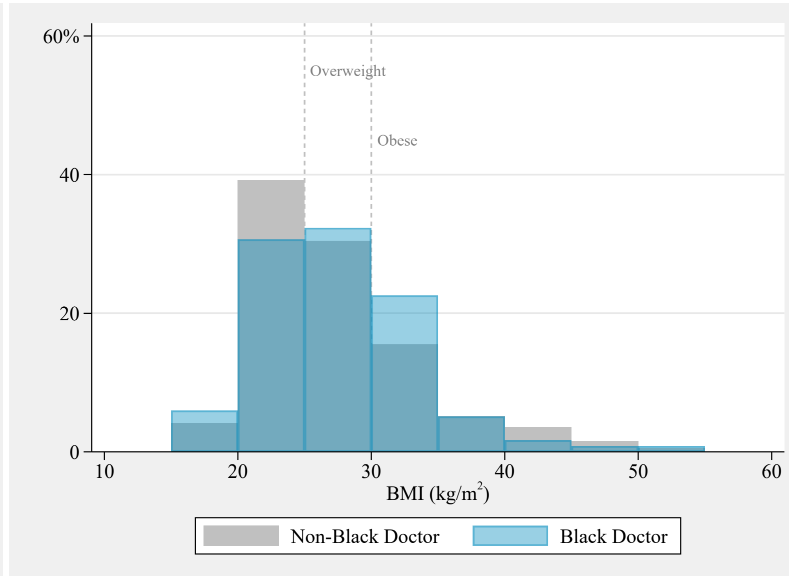
(b) Delta

Note: Figure plots the black doctor coefficient on a random selection of N subjects with replacement, where $N = 12$. We limit the random selection to subjects who were assigned to the eight doctors who saw the 12 out-of-sample subjects. Permutation test runs the main regression (Equation 1) 1,000 times. Vertical (red) line signifies the coefficient from the subjects who did not meet study criteria.

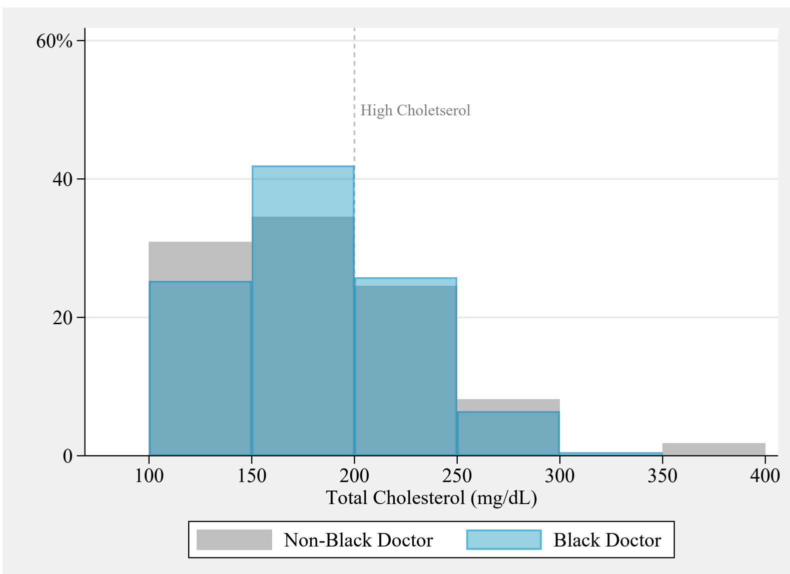
Figure 7: Distribution of Medical Screening Results



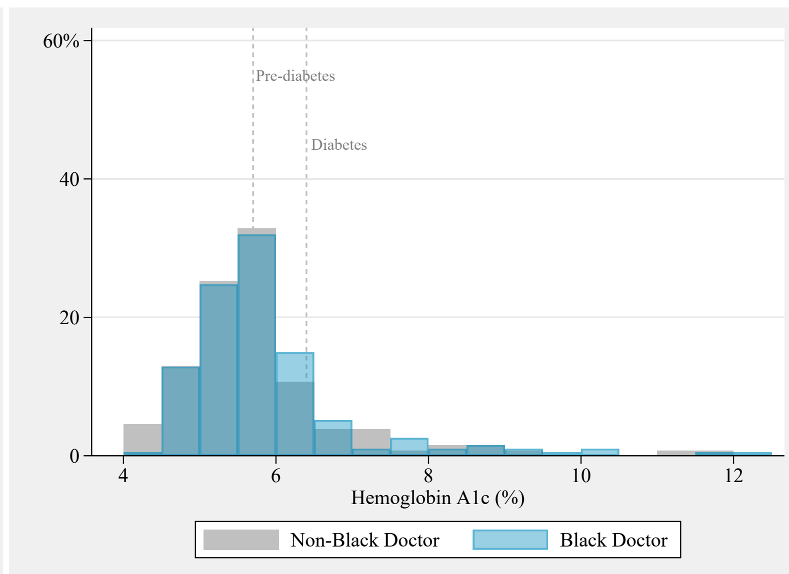
(a) Blood Pressure: Systolic



(b) Body Mass Index



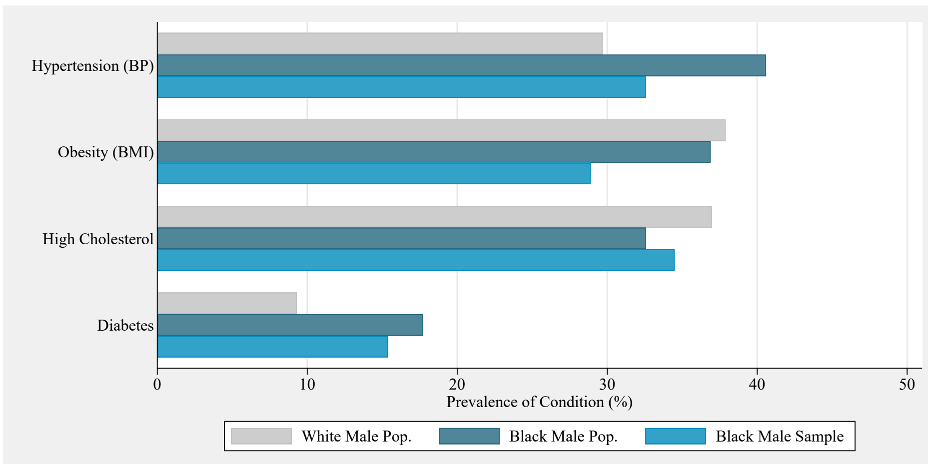
(c) Cholesterol



(d) Diabetes

Note: Distribution of medical screening results for subjects who elected to receive preventives by race of doctor.

Figure 8: Health of Study Sample vs U.S. Population



Note: Figure plots the percentage of each demographic group diagnosed with the listed conditions. Hypertension is defined as a systolic blood pressure value greater or equal to 140 mm Hg, obesity as a BMI greater or equal to 30 kg/m^2, high cholesterol as a cholesterol value greater or equal to 200 mg/dL, and diabetes as an A1c value greater or equal to 6.5%. Study sample values are for subjects who opted to receive a screening. Values for the U.S. population are from Fryar et al. (2017), Hales et al. (2017), and CDC (2017b, 2017c).