Title

Full title: Americans do not discriminate against doctors based on their race.
Short title: No racial or ethnic discrimination against doctors.

Authors

Reilly Olinger[1], Benjamin R. Matejka[1], Rohan Chakravarty[1], Margaret Johnston[1], Eliana Ornelas[1], Julia Draves[1], Nishi Jain[1], Jane H. Hentschel[1], William L. Owen[1], Yuchuan Ma[1], William W. Marx[1], Joshua P. Freitag[1], Nicholas Zhang[1], Cameron Guage[1], and Charles Crabtree[1]*.

1: Dartmouth College.
*: crabtree@dartmouth.edu
†: All authors contributed equally.

Abstract

To what extent do Americans racially discriminate against doctors? While a large literature shows that racial biases pervade the American healthcare system, there has been no systematic examination of these biases in terms of who patients select for medical treatment. We examine this question in the context of the ongoing global COVID-19 pandemic, where a wealth of qualitative evidence suggests that discrimination against some racial minorities, particularly Asians, has increased throughout the United States. Conducting a well-powered conjoint experiment with a national sample of 1,498 Americans, we find that respondents do not, on average, discriminate against Asian or other racial minority doctors. We also find no consistent evidence of treatment effect heterogeneity; Americans of all types appear not to care about the racial identity of their doctor at least in this context. This finding has broad implications for the potential limits of American prejudice.

Teaser

Racial discrimination in America is a serious problem, but Americans prioritize their medical care over their racial attitudes.

Introduction

To what extent do Americans racially discriminate against doctors? This question takes on particular importance during the ongoing COVID-19 pandemic, when more than 34.5 million people in America have been infected with the virus and more than 611,000 have died from it. With untold numbers of Americans likely to grapple with the long term, serious side effects of COVID-19 exposure (Del Rio et al 2020), public demand for healthcare across the country is expected to remain high for years. Within this context, it is crucially important to understand how racial attitudes and discrimination mighy shape healthcare interactions and outcomes. In that vein, we examine how patients select medical care providers, when they have a choice, and what role, if any, racial biases play in their decisions.

An enormous, rich, and growing literature shows that racial attitudes influence many Americans' decisions. We observe racial discrimination in the labor market, where minorities receive fewer opportunities and earn lower wages (Pager 2007), in housing, where minorities are more likely to be passed over by landlords as renters and banks as homeowners (Yinger 1995), in credit decisions, where minorities receive more loan rejections and must pay higher interest rates (Shapiro 1997), in all manner of consumer interactions, where minorities are quoted higher costs for goods and services and minority sellers or providers encounter worse treatment and lower demand (Ayers and Siegelman 1995; Doleac and Stein 2013), in politics, where minorities receive less attention, help, and representation (Butler 2014), and crucially in the domain of law, where minorities often face greater penalties for legal infractions and are regularly targeted by security forces even when they have done nothing wrong (Weaver and Prowse 2020).

More relevant to our paper, an important line of work shows vast disparities in healthcare coverage, treatment, and outcomes across racial groups (Williams, Lawrence, and Davis 2019). Prior research has documented large observational disparities in access to healthcare and in the health outcomes that result downstream (Cheon 2016; Graham 2009; Hibber and Quan 2017; Higginson and Constantini 2002; Kawachi et al 2002). Recent experimental work also indicates that these doctors might have strong biases that can arise even in the context of critical life saving situations (Crabtree, Holbein, and Monson 2021). In line with this, studies have found that Black patients were significantly more likely to perceive racial discrimination in healthcare settings and that Asian immigrants were more likely than American born Asians and whites to perceive racial discrimination in the same setting (Hausmann et, al 2012, Lauderdale et. al 2003). Taken together, the findings from this literature suggest that race plays a key role in medical treatment, healthcare processes, and patient outcomes.

Despite a growing literature on the role that discrimination plays in healthcare and in public choices about goods and services, there has been very little research on *discrimination against healthcare workers*. This is a puzzling gap given the large body of research on racial inequalities in healthcare. While the focus of this literature has been on racial discrimination against patients, there are reasons to believe that discrimination also occurs against healthcare workers as well, particularly doctors. A comprehensive literature review on biases in American healthcare finds that many racial minority physicians believe that they have been discriminated against by patients (Filut et al. 2020, 135). Building on these qualitative accounts, and a large number of anecdotal reports about discrimination against Asian doctors during the COVID-19 pandemic, we provide the first empirical investigation of the degree to which racial biases influence individual choices about which doctor they would see for medical care. Departing from earlier qualitative work, our focus here is not on how individuals treat doctors of different characteristics once visiting them, but rather the extent to which the personal attributes of doctors shape whether individuals decide to see them at all. Based on prior findings about the pervasiveness of racial discrimination in American life, we theorize that when patients have the opportunity to select from several doctors, they might rely on their individual biases against groups in their decisions about who to seek for treatment.

We think that identifying these biases is especially important considering the damaging effect they can have on doctors, who must cope with the actions and decisions of their patients, and also on discriminatory patients, who may opt for inferior care based on the doctor's identity or race. If patients discriminate, it could have concrete impacts on the job prospects, business successes, and even potentially the likelihood of dealing with malpractice suits for minority doctors. On the other hand, if patients do not discriminate, this would suggest that American racial attitudes manifest differently across contexts. The importance of our inquiry is further underscored by the fact that the American medical workforce continues to diversify - increasing the proportion of healthcare workers who might be discriminated against - and the likely reality that Americans post-COVID-19 will have greater medical needs - increasing the number of opportunities for individuals to engage in discrimination.

Data and Design

To test the extent to which Americans racially discriminate against doctors, we conducted a conjoint experiment with a national sample of 1,498 Americans recruited through Lucid Theorem with quotas for age, race, gender, educational attainment, household income, Census region, and political party. Lucid collected data from February 18 to March 4 2021. Conjoint experiments are commonly used in the social sciences to help understand how people value different attributes of possible choices (Auerbach and Thachil 2018; Bansak et al 2018; Hainmueller et al 2015; Hainmueller and Hopkins 2015; Jenke 2021). This experimental design was initially developed by market researchers in the 1970s to study what product attributes made consumers more likely to buy a given product. In recent years, conjoint experiments have exploded in popularity as a means of eliciting individual preferences (Heinmueller et al 2015).

Under a conjoint design, a researcher shows a survey or lab respondent a series of products---in our case, a fictional doctor listing---and randomizes a set of potential product attributes---in our case, the attributes of the doctors in the listings. Generally speaking, the objective of a conjoint experiment is to determine what combination(s) of a limited number of attributes is most influential in driving product choices. In a conjoint experiment, null effects are interpreted as respondents not using those attributes in their decision-making, while substantively and statistically significant effects suggest that attributes are important for respondents' choices.

We chose to conduct a conjoint design about hypothetical doctors because this is the best available research design for studying the question at hand. This is true for at least two important reasons. First, the main alternative to a conjoint experiment---looking at the actual patterns in doctor visitation across various demographic groups---would be a very inappropriate way of testing for individual biases. With this research design, we could not hold all other factors about the doctor constant. If we observed differential rates of doctor visitations across different doctor attributes, we could not know if these were caused by the attributes or other unobservable features. Ignoring this fundamental problem to making inferences about racial discrimination against doctors using observational data, there is another more fundamental concern: no comprehensive dataset like this exists. Based on these two reasons alone, we think that an experiment about fictional doctors is the most appropriate approach to examine racial discrimination in doctor selection. In addition, though, we think that conjoint experiments are a particularly powerful tool for understanding racial discrimination against doctors. This is because the experimental literature shows that conjoint experiments can be used to minimize satisficing or demand effects (Horiuchi et al 2020), even when researchers use a large number of attributes (> 10) or ask participants to complete many tasks (> 10) (Bansak et al 2018; Hainmueller et al 2015; Hainmueller and Hopkins 2015).

We implement our survey with members of the public because they are likely patients. In addition to understanding public preferences as patients, it is also important to understand if members of the public exhibit racial discrimination to doctors since both the behavior of elected officials (Healy and Maholtra 2013) and the behavior of healthcare providers may too be shaped by the preferences (Baiker et al 2014; Clinton and Sances 2018; Morgan and Cambell 2011). This means that any biases exhibited by the public might have diffuse consequences when it comes to public health priorities and even potentially medical care staffing and practices.

In our conjoint experiment, we ask respondents to choose between one of two possible doctors for medical care. It is reasonable to ask our respondents to evaluate doctors this way because many of them will likely need medical care in the next year, which contributes to the ecological validity of our experiment. We randomized several characteristics of the doctors. First, we randomized the reported Yelp review score (2.8/5, 3.9/5, 5/5) for each doctor. We selected these values by scraping Yelp for all reviews of doctor ratings and then picking the 25%, 50%, and 75% quantiles. We include this because of the central role that review ratings play in American selections of service providers, and because it allows us to test whether respondents are attentive to the doctor profiles and are responding in a way that is consistent with what we already know about public preferences. If our experiment were ecologically valid, we would expect that the effect of review ratings would be both statistically significant and substantively large. Another reason that we chose to manipulate this factor is that it allowed us to account for a potentially important form of bias in our experimental design. Specifically stating the doctors' rating allowed us to account for possible differential biases in the perceptions of ratings based on other individual characteristics. By providing this information to possible patients directly, we are potentially attenuating a mechanism that might drive discrimination. Since our (necessary) manipulation check blocks a potential channel for discrimination (i.e. one that arises from misperceptions about the overall quality that differ across race and other attributes), it allows us to estimate a vitally important quantity of interest---that is, the extent to which individuals exhibit racial discrimination independent from this form of statistical discrimination (Guryan and Charles 2013; Schwab 1986).

In addition to the Yelp review rating for each doctor, we also randomized the doctor's age (35, 41, 47, 60, 66), gender (man, woman), medical-degree granting institution (Drexel, East Carolina, Harvard, Michigan State, Tufts, UCLA), and race (White, Hispanic, Black, or Asian). We also randomized the type of clinic that the doctor practiced at (small public, large public, small private, large public) and the expected wait time (10 minutes, 15 minutes, 20 minutes). We selected these attributes based on prior literature about doctor selection and the results of two pilot studies that we describe more in the 'Materials and Methods' section. We directly stated each doctor's age, educational background, type of clinic in which they worked, and expected wait time. In line with the common practice in audit studies (Butler 2014; Costa 2017; Gaddis 2017) , we manipulated gender and race through the names we gave our fictional doctors. To help ensure that survey respondents received the correct racial manipulation, we conducted the largest known survey of American perceptions about names to date, selecting only those names that survey respondents (a) correctly perceived the intended race of at least 90% of the time, (b) thought had a college degree or higher, and (c) thought belonged to an American citizen at least 90% of the time.

We randomized this set of attributes because these are either often considered to be important when patients select doctors (e.g. review ratings, wait time, clinic type, age, and education see Salisbury 1989; Santos et al 2017), have been shown to inform decision making in other social contexts (e.g. gender and race; see Ayers and Siegelman 1995; Butler 2014, Costa 2017, Kalkan 2009, Karpowitz and Charles 2016, Paxton et al 2007), and/or have well-documented inequities in the health domain (see Cheon 2016; Graham 2009; Hibber and Quan 2017; Higginson and Constantini 2002; Kawachi et al 2002)). All of the attributes in our conjoint designs are information that individuals could know about their doctors with some consumer research.

The COVID-19 pandemic is an important context in which to conduct this experiment. The spread of COVID-19 has increased the salience of medical care and treatment. While most Americans rarely think about seeing the doctor, many must now consider this possibility daily. Another reason why it is an important context relates to patterns of racial discrimination in the United States since the early days of the pandemic, when a wealth of qualitative evidence showing that discrimination against some racial minorities has increased substantially.

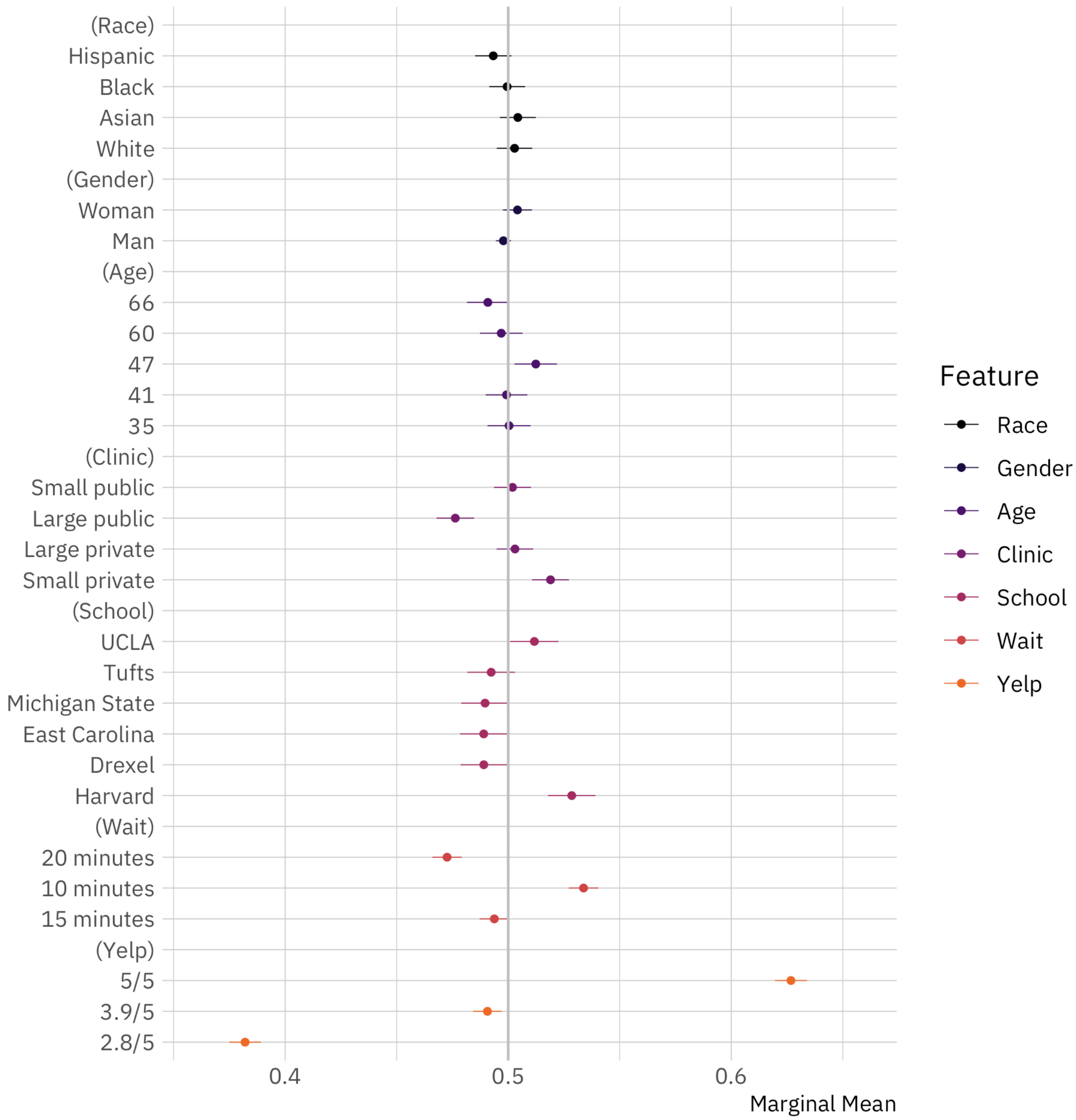
Finally, we note that in our analyses, we focus on main treatment effects (i.e. among our entire sample) and on those broken by the self-reported political party of the individuals in our study. (To be thorough, however, we examine treatment effect heterogeneity by all other available baseline characteristics in the Supporting Information.) Whether political polarization, and political attitudes more generally, moderates key personal (i.e. consumer) decisions made in the healthcare realm remains unclear. Our novel results speak to a key area with mixed findings thus far.

Results

Figure 1 shows the results from our conjoint experiment; circles denote marginal means and thin bars denote 95% confidence intervals. Conditional marginal means are grouped by experimental factor and denote the predicted probability for a doctor with a specific factor level, across all other factor levels. If the confidence intervals for a doctor attribute cross the 50% line (the grey vertical reference line), we cannot reject the null hypothesis that patient choices in regards to that factor level are random. In other words, we cannot be sure that patients discriminate either *against* or *for* doctors with that characteristic. Importantly, if individuals were selecting doctors at random, we would not expect them to exhibit any biases against or toward certain

We first examine whether survey respondents were more likely to select doctors with higher Yelp ratings, since this provides some insight into the way individuals interacted with our conjoint experiment. If respondents were paying attention to it and taking it seriously, we would expect them to select doctors with better reviews. Specifically, we would expect that the marginal means would increase with ratings and that the differences would be substantially large. In fact, this is just what we see. Survey respondents indicate that they would see doctors with 5/5 ratings about 63% [61.96%-63.39%, 95% CI] of the time, doctors with 3.9/5 ratings about 49% [48.42%-49.71%] of the time, and doctors with 2.8/5 ratings about 38% [37.49%-38.92%] of the time.

Figure 1: Public Preferences for Doctors



Note: Marginal means plot for the effect of doctor attributes on survey respondent selection. The circles represent the marginal means while the thin bars denote 95% confidence intervals. Coefficients on the left side of the grey line at 50% indicate that respondents are, all-else-equal, less likely to choose a doctor with the given characteristics on the vertical axis; those on the right are, all-else-equal, more likely to choose a doctor with the given characteristic. The unit of analysis is the respondent-choice profile. Hence, the \$N\$ reported in our models below is the number of respondents (1,498) multiplied by the number of pairwise choices (15) and individuals within those pairs (2). N = 44,940.

We next turn to whether survey respondents select doctors based on other characteristics. If our respondents were making decisions based solely on a doctor's rating, we would expect that all other characteristics manipulated in our conjoint experiment—age, clinic type, educational background, gender, and wait time—would be insignificant. Particularly, if respondents were treating racial majorities and racial minorities the same---responding to our racial conditions randomly---we would expect to see no differences across this dimension.

We find that some of the characteristics matter while others do not. Specifically, we find evidence that respondents are sensitive to the potential wait times that they would face, preferring 10 minutes (53% [52.71%-54.04%]) or 20 minutes (47% [46.60%-47.91%]). They are indifferent between doctors when the wait time is 15 minutes (49% [48.72%-50.04%]). These findings add additional credibility to our experimental design, as they indicate that respondents are reacting to the choices they face as we would generally expect them to in the real world. We also observe that educational background seems to matter. While a medical degree from Michigan State (49% [47.89%-50.03%]) or Tufts (49% [48.17%-50.31%]) does not decrease the likelihood that respondents would pick a doctor, a degree from Drexel (49% [47.87%-49.94%]) or East Carolina (49% [47.84%-49.97%]) does, though to a very small extent. On the other hand, degrees from Harvard (52% [51.78%-53.92%]) and UCLA (51% [50.09%-52.25%]) bolster a doctor's chances of being chosen, though not by much. Taken together, these educational results also fit with what we should expect about respondent behavior. Looking at the results for clinic type, we see that respondents prefer doctors who practice in small private clinics (52% [51.07%-52.72%]), would rather avoid those who practice at large public clinics (48% [46.78%-48.47%]), and are indifferent to small public and large private clinics. These findings fit broadly with past work on patient preferences.

Examining the effects of personal characteristics, we see that age appears to matter to some extent. Individuals appear not to care about how old doctors are most of the time, with the one exception being a weak preference for 47-year-olds (51% [50.28%-52.19%]). We interpret this to indicate that individuals prefer doctors who have some mix of youth and experience. More interestingly, we find that respondents do not prefer doctors of one gender to another. The marginal means for both men and women doctors are 49.8% [49.44%-50.12%] and 50.4% [49.76%-51.07%], respectively, and both estimates are statistically indistinguishable from 50% ($p > 0.2$). In other words, respondents do not seem to consider the gender of doctors, holding all else constant, when making healthcare decisions. While women face discrimination in many aspects of American life (Davis and Greenstein 2009), and face discrimination by gatekeepers within the medical profession (Kouta and Kaite 2011), it seems that prospective patients are willing to set aside their gender attitudes when selecting doctors to provide care.

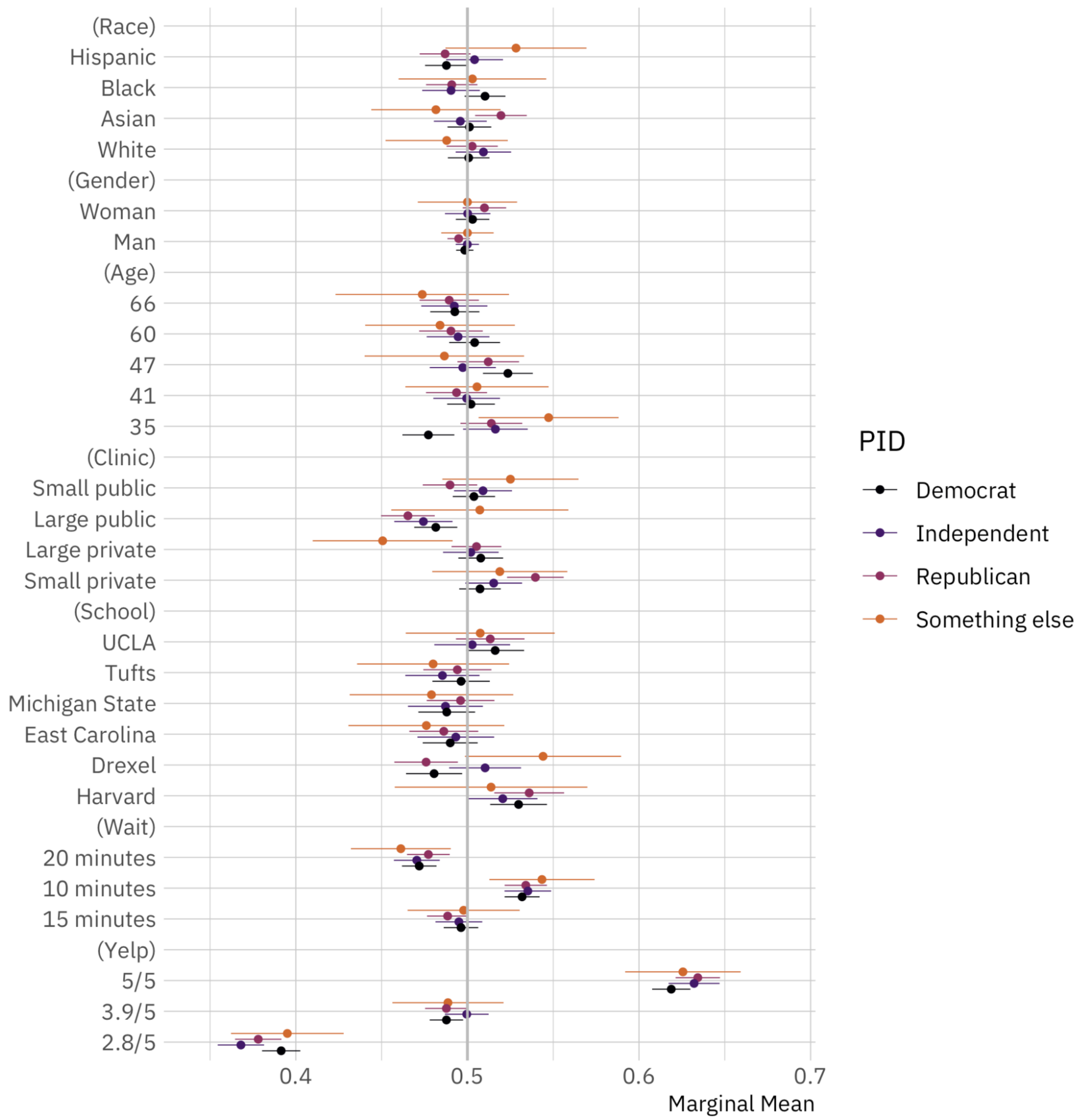
Finally turning to the effects of race, our central interest in this paper, we find that respondents *do not appear* to discriminate against racial minorities. While the marginal mean for Hispanics is slightly lower than it would be by chance (49.33% [48.52%-50.14%]), we cannot reject that this effect differs from 50% ($p > 0.1$). Comparing this effect to results from audit studies (Costa 2017; Quillian 2017; Gaddis et al 2021) or previous conjoint experiments (Auerbach and Thachil 2018; Bansak et al 2018; Hainmueller et al 2015; Hainmueller and Hopkins 2015; Jenke 2021), we also see that this effect is very small. At the least, we can say that any racial discrimination that Hispanic doctors face in this particular context appears tiny in comparison to the racial discrimination faced by them and other racial minorities in America across different contexts (Gaddis et al 2021). Importantly, we also find--albeit surprisingly--that respondents do not appear to discriminate against Asian (50.43% [49.62%-51.24%], $p > 0.29$) or Black (50% [49.14%-50.76%], $p > 0.9$) doctors. Conversely, respondents do not appear to prefer White doctors (50.28% [49.49%-51.08%], $p > 0.47$). Interestingly, this evidence concurs with our initial findings from two pilot studies. Pooling the data from these studies together, we find no evidence of racial or gender discrimination in doctor choice. For more information see Supporting Information. Taken together, our results suggest that racial biases do not influence doctor selection in America when respondents are presented with a wealth of information about their options. This potentially offers insight into the limits of American prejudice.

Perhaps these results mask substantial heterogeneity across respondents in our sample. Put differently, individuals from some groups might exhibit more or less racial discrimination. Here we examine the extent to which respondent political leanings influence how they respond to our conjoint experiment, in general, and how they react to doctors of different races, specifically. Figure 2 shows the results of this model. As in Figure 1, circles denote marginal means and thin bars denote 95% confidence intervals. Plotted points and confidence intervals differ in color based on political identification. Marginal means are grouped by experimental factor. As a reminder, if the confidence intervals for a doctor attribute cross the 50% line (the grey vertical reference line), we cannot reject the null hypothesis that patient choices in regards to that attribute are random.

We would expect that public preferences about doctors should vary based on political identification. Contemporary American society is marked by high levels of political polarization (Fiorina and Abrams 2008; Iyengar et al 2019). In line with that, prior work has shown that political identification might shape everything from where Americans live (Bishop 2009, Shafranek 2019 hough see also Gimpel and Hui 2015, and Mummolo and Nall 2017) , who they choose to date and marry (Hersh and Gitz 2017; Nicholson 2016; Huber and Maholtra 2017), how they conduct economic transactions (Engelhardt and Utych 2018; Kam and Deichert 2017; Neilson 2010), and how they perceive salient outgroups, such as racial, ethnic, and religious minorities (Barreto and Bozonelos 2009; Enos 2016; Giles and Hertz 1994; Haner et al 2020; Kalkan 2009; Karpowitz and Charles 2016; Lajevardi 2020).

What we find, though, is the opposite. It seems that political leanings shape what Americans perceive and what they do in many areas of life but *not* what they look for in their doctors. There is weak evidence that Republicans and Democrats prefer not to visit Hispanic doctors (48.69% [47.53%-50.02%] and 48.77%[47.21%-50.18%]), but these findings are relatively small and statistically insignificant ($p > 0.10$). There is also some evidence that Republicans actually prefer Asian doctors (51.95% [50.44%-53.46%], $p < 0.05$). But, on the whole, the differences across political identification are not statistically significant or substantively important. This is a striking finding given the large, growing literature reference above that shows the central importance of political identification in driving American perceptions and choices.

Figure 2: Public Preferences for Doctors by Political Identification



Note: Conditional marginal means for the effect of doctor attributes on survey respondent selection by political identification. The circles represent the marginal means while the thin bars denote 95% confidence intervals. Coefficients on the left side of the grey line at 50% indicate that respondents are, all-else-equal, less likely to choose a doctor with the given characteristics on the vertical axis; those on the right are, all-else-equal, more likely to choose a doctor with the given characteristic. The unit of analysis is the respondent-choice profile. The N reported in our models below is the number of respondents (1,498) multiplied by the number of pairwise choices (15) and individuals within those pairs (2). N = 44,940.

In the Supporting Information, we further explore possible treatment effect heterogeneity by looking at group differences across respondent education, gender, political identification, race, and views on racial discrimination. We find no compelling, consistent differences across groups, especially after adjusting p-values for multiple comparisons. Ultimately, while there is some small variation in our treatment effect heterogeneity results, the overall trend is clear: members of the public do not provide differential treatment to doctors in our task based on race.

Discussion

Americans regularly interact with doctors, yet we know little about what racial biases, if any, drive these encounters and the extent to which they might influence practitioner choice. Our study, based on a conjoint experiment conducted with a nationally representative sample of 1,498 Americans, provides new evidence that patients in the United States do not appear to engage in racial discrimination when evaluating doctors for a potential visit. The results of our experiment stand out against the ever-growing number of studies showing that racial discrimination marks most aspects of daily life in America.

We think that there are at least three reasons why we do not observe racial discrimination, in general, or anti-Asian discrimination, in particular. One relates to public perceptions of doctors. According to a Pew study conducted in January 2019, 74% of the American public have a mostly positive view of doctors, while another 68% expressed a mostly positive view of medical research scientists. The positive approval rating of doctors has only likely increased since the COVID-19 pandemic began, especially in light of the hagiographic coverage they have received in news media. Perhaps the positive perceptions that Americans have of doctors overshadow any negative perceptions that they might have of racial minorities. A second explanation pertains to American understandings about the costs of racism. It might be that individuals recognize - at a subconscious level at least - that engaging in racial discrimination entails costs and leads to less efficient outcomes. Potentially people are willing to pay these costs in some aspects of their lives, but not when their health might be jeopardized. A third explanation relates to why we did not find anti-Asian discrimination, in particular. This might be due to widespread views of Asians as a ‘model minority’ (Chuo and Feagin 2015). This problematic view treats Asian Americans as a homogenous, successful group, glossing over achievement disparities and labor market disadvantages within the racial category (Sakamoto et al 2012). This widespread and frequently perpetuated stereotype could have lead respondents to have potentially unrealistic expectations for the care that Asian American doctors could provide and, therefore, a positive bias toward them.

We acknowledge several limitations of our work and think that they should animate additional research on the contexts in which Americans discriminate against racial minorities. First, we conducted our survey experiment in March 2021, in the middle of an ongoing global pandemic. As we discussed above, this temporal context might have caused respondents to view doctors of all races in an equally positive hue. Second, we focus here on how potential patients might choose doctors in the absence of pressing medical needs. It might be that when patients are actually experiencing physical discomfort or pain, or stress-related to those conditions, they are less likely to evaluate medical options objectively and more inclined to rely on deep-seated biases. Third, while we used quotas to ensure a nationally representative sample at the respondent recruitment stage, we introduced some demographic skew into our sample when we dropped inattentive subjects. As a result, our sample had too many white (+15.28% over the national population) and higher-income respondents (+2.77%), and too few Black (-3.105%), LatinX (-11.877%), and lower-income respondents (-3.079%). Researchers should probe the temporal validity of our work in subsequent studies, assess the extent to which individuals experiencing medical issues might be more likely to exhibit bias, and check the generalizability of our findings to other samples.

More broadly, scholars concerned about racial discrimination should continue investigating its presence and drivers in healthcare. Given the ongoing pandemic, and the societal aging trends in many developed countries, we can expect public interactions with medical personnel to increase greatly in the years and decades ahead and, consequently, the number of chances for discrimination to increase correspondingly. In attempting to understand how racial discrimination influences medical choices and outcomes, we can shed light on group dynamics in a vital, growing context.

Methods

Data and participants

In determining whether Americans discriminate against doctors based on their race or ethnicity, three studies were conducted: two pilot studies (n = 174, n = 330) and a final study (n = 1,498). All surveys were programmed in Qualtrics. Participants were recruited through Prolific for the pilot studies and through Lucid Theorem for the final study. Quotas on age, gender, race/ethnicity, party identification, and education were used in all three sample recruitments.

Survey instrument

After confirming respondents' consent to participate in our study, we conducted a two-step attention check (Aronow et al 2020). Following this attention check, we collected socio-demographic data about the respondents' political affiliation and ideology, gender identity, country of birth, American citizenship status, race/ethnicity, annual household income and education level. We then asked respondents whether they agreed that racial discrimination is a major problem in the United States.

Following these background questions, participants completed a vignette experiment where they were assigned to one of three treatments. In one, they read nothing. In the second, they read about a fictional incident of racial violence against Asian Americans that was presented as a real news excerpt. It centered on the story of an Asian American family who had their family restaurant attacked by racists. In the third, they read an automatically generated placebo condition (Porter and Velez 2021). This treatment was created using Open AI's novel large-scale unsupervised language model GPT-2 based on the seed phrase, "Jane Smith and her brother, Joe, showed up at their family's restaurant." Following Porter and Velez's recommendations, we generated a large number of these vignettes (n = 4,930) and randomly assigned one of them to recipients who received our placebo condition. After reading the placebo and fictional news vignettes, respondents were asked to write a bit about how that made them feel. The idea here was that those who read the excerpt about racial violence against Asian Americans would be forced to engage in perspective taking and thus exhibit less discrimination against Asian doctors. We find and show in the Supporting Information that none of these treatments influenced responses, probably because (as we explained above) we observe no discrimination against Asian doctors, so we do not focus on that part of the design or results in this paper.

After completing the vignette experiment, participants completed a conjoint experiment focusing on doctor choice. We asked them to evaluate two possible doctors and choose between them based on which they would be more likely to visit. We had them complete this task 15 times. We presented participants with information about the doctors in a table, varying six attributes. As described above, we presented information on (1) the doctor's name, (2) their age, (3) the medical school they attended, (4) the clinic type where the doctor practices, (5) the average wait time and (6) the doctor's rating on a five-star basis. The order of attributes in the table was fixed, but the levels of attributes were randomized.

We selected names to signal racial identities based on Crabtree et al (2021). Our age levels were based on 2018 age distribution data of U.S.-based physicians. We randomly selected medical schools across the 2020 U.S. News World and Report ranking - UCLA, Tufts, Michigan State, East Carolina, Drexel, Harvard. Clinic type was limited to “small public”, “small private”, “large public”, and “large private”, capturing broad differences in medical practice time. Physician wait time was limited to 10, 15 and 20 minutes, a range of reasonably plausible values.

To construct our Yelp ratings levels, we scraped all results for the query “Doctors” in American cities with populations larger than 100,000 (N = 382) according to the US Census. Using this procedure, we constructed a sample size of 109,551 practices; 61,021 contained a rating. We calculated averages and standard deviations for physician practices at the city level, then averaged to the national level, weighted by the cities’ populations. We then calculated the 25%, 50%, and 75% quantiles of the resulting distribution and used those values in our conjoint.

Analytic strategy

To determine the effect of each experimental factor in the conjoint, we calculated marginal means for the whole sample, and also calculated marginal means for subgroups.

Robustness checks

We find similar results when we calculate average marginal component effects (AMCEs) instead of marginal means. Our results are also surprisingly robust across subgroups, as mentioned above. In addition, we obtain similar findings if we include respondents who failed the attention check and if we drop those who used a mobile device to complete our survey.

Competing interests

There are no competing interests.

Data availability

The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

Ethics declarations

We received approval for this study from Dartmouth's Committee for the Protection of Human Subjects. All experiments were performed in accordance with relevant named guidelines and regulations. Informed consent was obtained from all participants and/or their legal guardians.

References

Aronow, P. M., Kalla, J., Orr, L., & Ternovski, J. (2020). Evidence of rising rates of inattentiveness on Lucid in 2020.

Auerbach, A. M., & Thachil, T. (2018). How clients select brokers: Competition and choice in India's slums. American Political Science Review, 112(4), 775-791.

Ayres, I., & Siegelman, P. (1995). Race and gender discrimination in bargaining for a new car. The American Economic Review, 304-321.

Bansak, K., Ferwerda, J., Hainmueller, J., Dillon, A., Hangartner, D., Lawrence, D., & Weinstein, J. (2018). Improving refugee integration through data-driven algorithmic assignment. Science, 359(6373), 325-329.

Baicker, K., Finkelstein, A., Song, J., & Taubman, S. (2014). The impact of Medicaid on labor market activity and program participation: evidence from the Oregon Health Insurance Experiment. American Economic Review, 104(5), 322-28.

Barreto, M. A., & Bozonelos, D. N. (2009). Democrat, Republican, or none of the above? The role of religiosity in Muslim American party identification. Politics and Religion, 2(2), 200-229.

Bishop, B. (2009). The big sort: Why the clustering of like-minded America is tearing us apart. Houghton Mifflin Harcourt.

Butler, D. M. (2014). Representing the advantaged: How politicians reinforce inequality. Cambridge University Press.

Cheon, S., Agarwal, A., Popovic, M., Milakovic, M., Lam, M., Fu, W., ... & Chow, E. (2016). The accuracy of clinicians' predictions of survival in advanced cancer: a review. Ann Palliat Med, 5(1), 22-29.

Chou, R. S., & Feagin, J. R. (2015). *Myth of the model minority: Asian Americans facing racism*. Routledge.

Clinton, J. D., & Sances, M. W. (2018). The politics of policy: The initial mass political effects of medicaid expansion in the states. American Political Science Review, 112(1), 167-185.

Costa, M. (2017). How responsive are political elites? A meta-analysis of experiments on public officials. Journal of Experimental Political Science, 4(3), 241-254.

Crabtree, C, Gaddis, M. S., Guage, C., Holbein, John B., Kim J. Y., and Marx, W. (2021). Validated names for experimental studies on ethnicity and race. Working paper.

Davis, S. N., & Greenstein, T. N. (2009). Gender ideology: Components, predictors, and consequences. *Annual review of Sociology*, *35*, 87-105.

Del Rio, Carlos, Lauren F. Collins, and Preeti Malani. "Long-term health consequences of COVID-19." Jama 324.17 (2020): 1723-1724.

Engelhardt, A. M., & Utych, S. M. (2020). Grand old (Tailgate) party? Partisan discrimination in apolitical settings. Political Behavior, 42(3), 769-789.

Enos, R. D. (2016). What the demolition of public housing teaches us about the impact of racial threat on political behavior. American Journal of Political Science, 60(1), 123-142.

Fiorina, M. P., & Abrams, S. J. (2008). Political polarization in the American public. Annu. Rev. Polit. Sci., 11, 563-588.

Filut, A., Alvarez, M., & Carnes, M. (2020). Discrimination toward physicians of color: a systematic review. Journal of the National Medical Association, 112(2), 117-140.

Gaddis, S. M. (2017). How black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies. Sociological Science, 4, 469-489.

Gaddis, S.M., Larsen, E.N., Crabtree, C., & Holbein, J.H. (2021) Meta-Analysis of Correspondence Audits: Racial Discrimination is More Common in Hiring than Other Contexts. Unpublished manuscript.

Giles, M. W., & Hertz, K. (1994). Racial threat and partisan identification. American Political Science Review, 88(2), 317-326.

Gimpel, J. G., & Hui, I. S. (2015). Seeking politically compatible neighbors? The role of neighborhood partisan composition in residential sorting. Political Geography, 48, 130-142.

Graham, H. (2009). Understanding health inequalities. McGraw-hill education (UK).

Guryan, J., & Charles, K. K. (2013). Taste‐based or statistical discrimination: the economics of discrimination returns to its roots. The Economic Journal, 123(572), F417-F432.

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. Proceedings of the National Academy of Sciences, 112(8), 2395-2400.

Hainmueller, J., & Hopkins, D. J. (2015). The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants. American Journal of Political Science, 59(3), 529-548.

Haner, M., Sloan, M. M., Cullen, F. T., Graham, A., Lero Jonson, C., Kulig, T. C., & Aydın, Ö. (2020). Making America safe again: Public support for policies to reduce terrorism. Deviant Behavior, 1-19.

Hausmann, L. R., Hannon, M. J., Kresevic, D. M., Hanusa, B. H., Kwoh, C. K., & Ibrahim, S. A. (2011). Impact of perceived discrimination in health care on patient-provider communication. *Medical care*, *49*(7), 626.

Healy, A., & Malhotra, N. (2013). Retrospective voting reconsidered. Annual Review of Political Science, 16, 285-306.

Hersh, E., & Ghitza, Y. (2018). Mixed partisan households and electoral participation in the United States. PloS one, 13(10), e0203997.

Hibberd, C. S., & Quan, G. M. (2017). Accuracy of preoperative scoring systems for the prognostication and treatment of patients with spinal metastases. International scholarly research notices, 2017.

Higginson, I. J., & Costantini, M. (2002). Accuracy of prognosis estimates by four palliative care teams: a prospective cohort study. BMC palliative care, 1(1), 1-5.

Horiuchi, Y., Markovich, Z. D., & Yamamoto, T. (2020). Does conjoint analysis mitigate social desirability bias?.

Huber, G. A., & Malhotra, N. (2017). Political homophily in social relationships: Evidence from online dating behavior. The Journal of Politics, 79(1), 269-283.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. Annual Review of Political Science, 22, 129-146.

Jenke, L., Bansak, K., Hainmueller, J., & Hangartner, D. (2021). Using eye-tracking to understand decision-making in conjoint experiments. Political Analysis, 29(1), 75-101.

Kalkan, K. O., Layman, G. C., & Uslaner, E. M. (2009). “Bands of others”? Attitudes toward Muslims in contemporary American society. The Journal of Politics, 71(3), 847-862.

Kam, C. D., & Deichert, M. (2020). Boycotting, buycotting, and the psychology of political consumerism. The Journal of Politics, 82(1), 72-88.

Knox, D., Lowe, W., & Mummolo, J. (2020). Administrative records mask racially biased policing. American Political Science Review, 114(3), 619-637.

Knox, D., & Mummolo, J. (2020). Making inferences about racial disparities in police violence. Proceedings of the National Academy of Sciences, 117(3), 1261-1262.

Kouta, C., & Kaite, C. P. (2011). Gender discrimination and nursing: α literature review. *Journal of Professional Nursing*, *27*(1), 59-63.

Lajevardi, N. (2020). Outsiders at home: The politics of American Islamophobia. Cambridge University Press.

Kawachi, I., Subramanian, S. V., & Almeida-Filho, N. (2002). A glossary for health inequalities. Journal of Epidemiology & Community Health, 56(9), 647-652.

Lauderdale, D. S., Wen, M., Jacobs, E. A., & Kandula, N. R. (2006). Immigrant perceptions of discrimination in health care: the California Health Interview Survey 2003. *Medical care*, 914-920.

Morgan, K. J., & Campbell, A. L. (2011). Delegated Governance in the Affordable Care Act. Journal of health politics, policy and law, 36(3), 387-391.

Mummolo, J., & Nall, C. (2017). Why partisans do not sort: The constraints on political segregation. The Journal of Politics, 79(1), 45-59.

Neilson, L. A. (2010). Boycott or buycott? Understanding political consumerism. Journal of Consumer Behaviour, 9(3), 214-227.

Nicholson, S. P., Coe, C. M., Emory, J., & Song, A. V. (2016). The politics of beauty: The effects of partisan bias on physical attractiveness. Political Behavior, 38(4), 883-898.

Pager, Devah, and Hana Shepherd. "The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets." Annu. Rev. Sociol 34 (2008): 181-209.

Paxton, P., Kunovich, S., & Hughes, M. M. (2007). Gender in politics. Annu. Rev. Sociol., 33, 263-284.

Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. Proceedings of the National Academy of Sciences, 114(41), 10870-10875.

Ritchie, Hannah, Esteban Ortiz-Ospina, Diana Beltekian, Edouard Mathieu, Joe Hasell, Bobbie Macdonald, Charlie Giattino, Cameron Appel, Lucas Rodés-Guirao and Max Roser

(2020) - "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/coronavirus' [Online Resource]

Salisbury, C. J. (1989). How do people choose their doctor?. British Medical Journal, 299(6699), 608-610.

Sakamoto, A., Takei, I., & Woo, H. (2012). The myth of the model minority myth. *Sociological Spectrum*, *32*(4), 309-321.

Santos, R., Gravelle, H., & Propper, C. (2017). Does quality affect patients' choice of doctor? Evidence from England. The Economic Journal, 127(600), 445-494.

Schwab, S. (1986). Is statistical discrimination efficient?. The American Economic Review, 76(1), 228-234.

Shafranek, R. M. (2021). Political considerations in nonpolitical decisions: a conjoint analysis of roommate choice. Political Behavior, 43(1), 271-300.

Weaver, V. M., & Prowse, G. (2020). Racial authoritarianism in US democracy. Science, 369(6508), 1176-1178.

Acknowledgments
The authors would like to thank Dan Hopkins, Matt Golder, Sona Golder Hakeem Jefferson, D.G. Kim, Yusaku Horiuchi, Natalie Masuoka, Tanika Raychaudhuri and participants at the Asian Politics Online Seminar Series and the Politics of Race, Immigration, and Ethnicity Consortium meeting held at the University of California, Santa Barbara for their extremely helpful comments.

Author contributions:
Conceptualization: RO, BRM, RC, MJ, EO, JD, NJ, JHH, WLO, YM, WM, JPF, NZ, CG, CC
Methodology: RO, BRM, RC, MJ, EO, JD, NJ, JHH, WLO, YM, WM, JPF, NZ, CG, CC
Investigation: RO, BRM, RC, MJ, EO, JD, NJ, JHH, WLO, YM, WM, JPF, NZ, CG, CC
Visualization: RO, BRM, RC, MJ, EO, JD, NJ, JHH, WLO, YM, WM, JPF, NZ, CG, CC
Supervision: CC
Writing—original draft: RO, BRM, RC, MJ, EO, JD, NJ, JHH, WLO, YM, WM, JPF, NZ, CG, CC
Writing—review & editing: RO, BRM, RC, MJ, EO, JD, NJ, JHH, WLO, YM, WM, JPF, NZ, CG, CC

Competing interests:
All authors declare they have no competing interests.

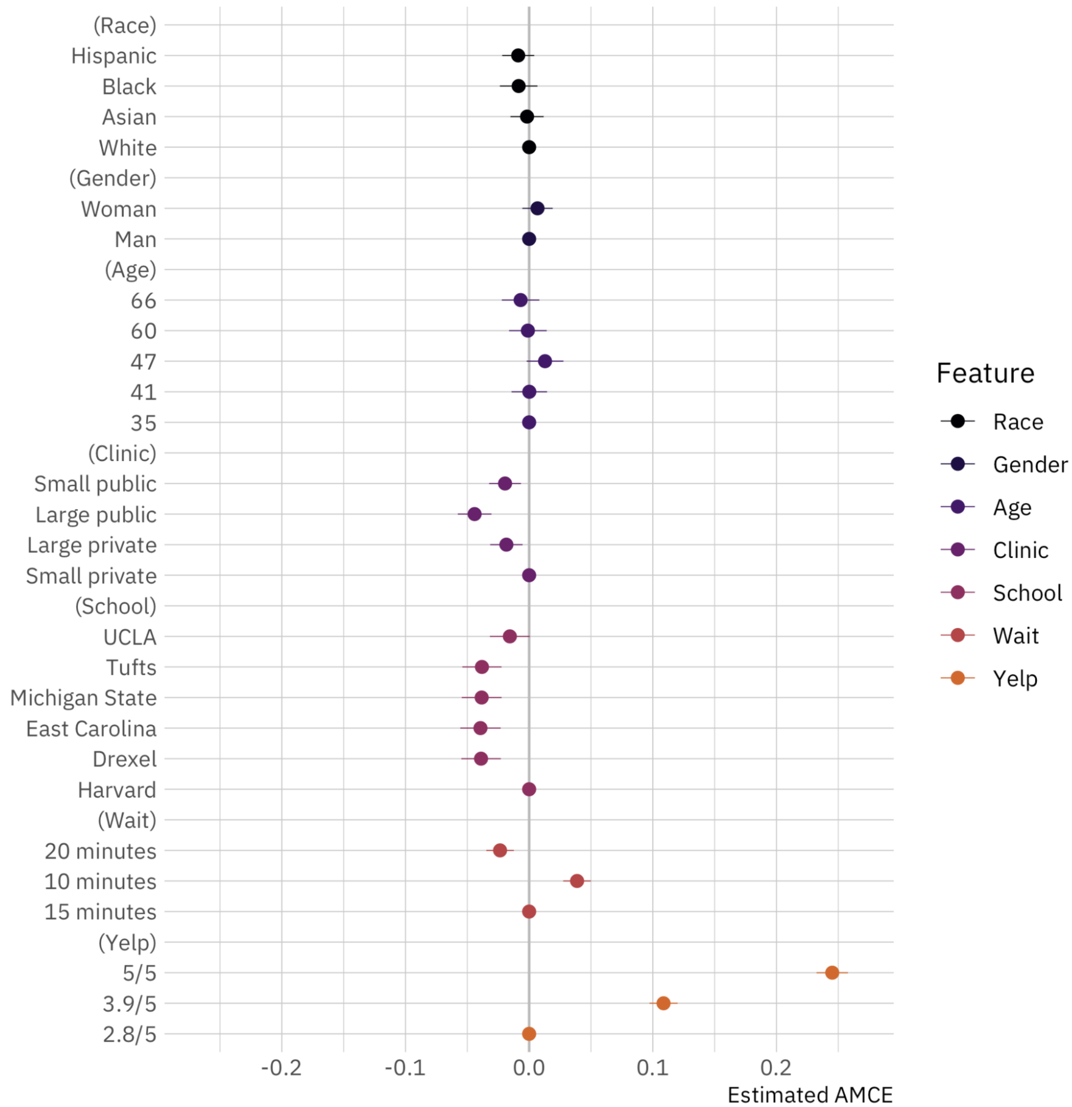
Data and materials availability:
All data, code, and materials used in the analyses will be deposited at the Harvard Dataverse upon publication.

Supplementary Information for "Americans do not discriminate against doctors based on their race"

SI 1: Study AMCEs

In this SI, we report the average marginal component effects from our study. They are substantively similar to the marginal means that we report in the main text.

Figure SI 1a: Public Preferences for Doctors

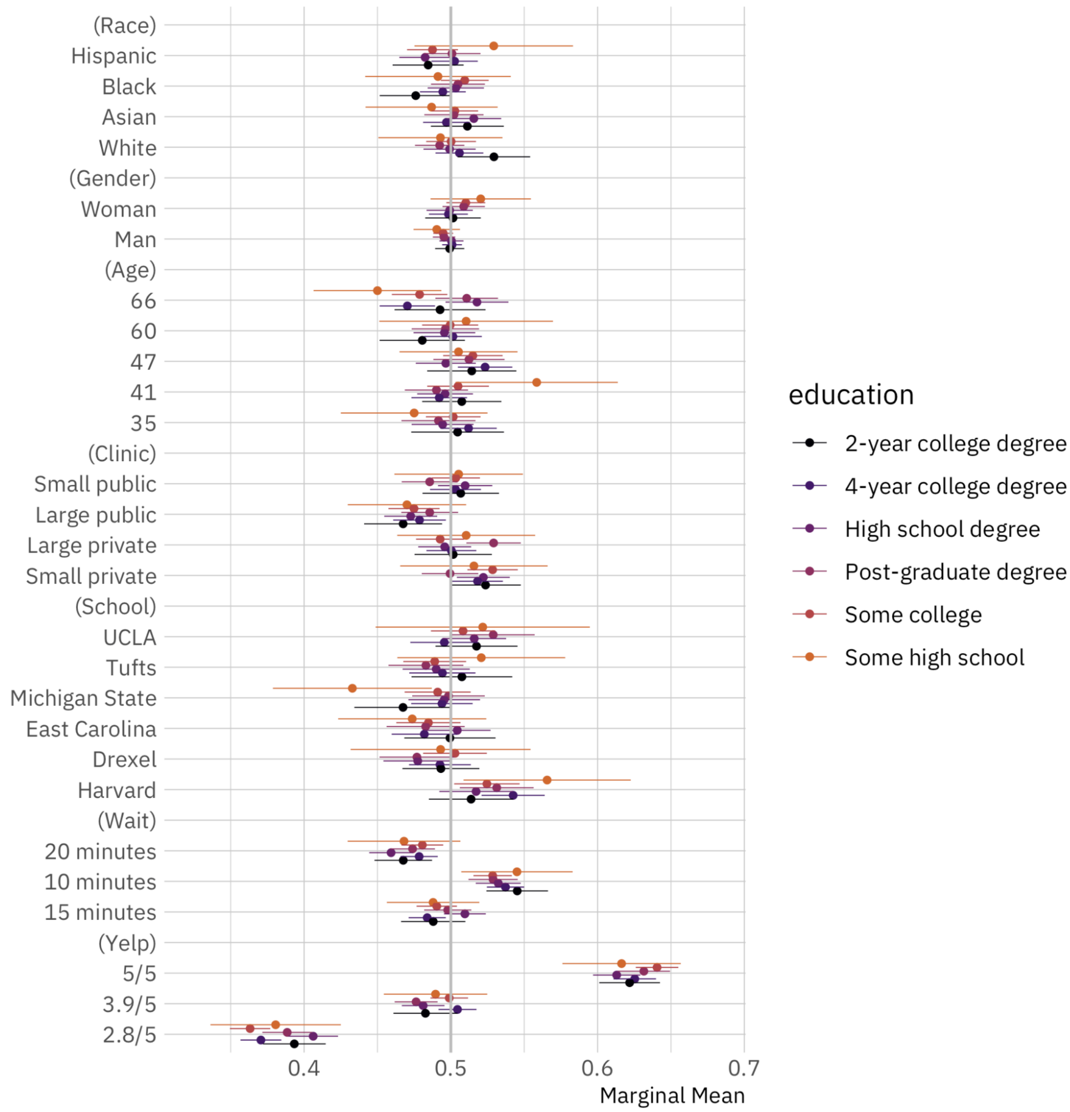


Note: AMCE plot for the effect of doctor attributes on survey respondent selection. The circles represent the marginal means while the thin bars denote 95% confidence intervals. Coefficients on the left side of the grey line at 50% indicate that respondents are, all-else-equal, less likely to choose a doctor with the given characteristics on the vertical axis; those on the right are, all-else-equal, more likely to choose a doctor with the given characteristic. The unit of analysis is the respondent-choice profile. Hence, the N reported in our models below is the number of respondents (1,498) multiplied by the number of pairwise choices (15) and individuals within those pairs (2). N = 44,940.

SI 2: Study treatment effect heterogeneity

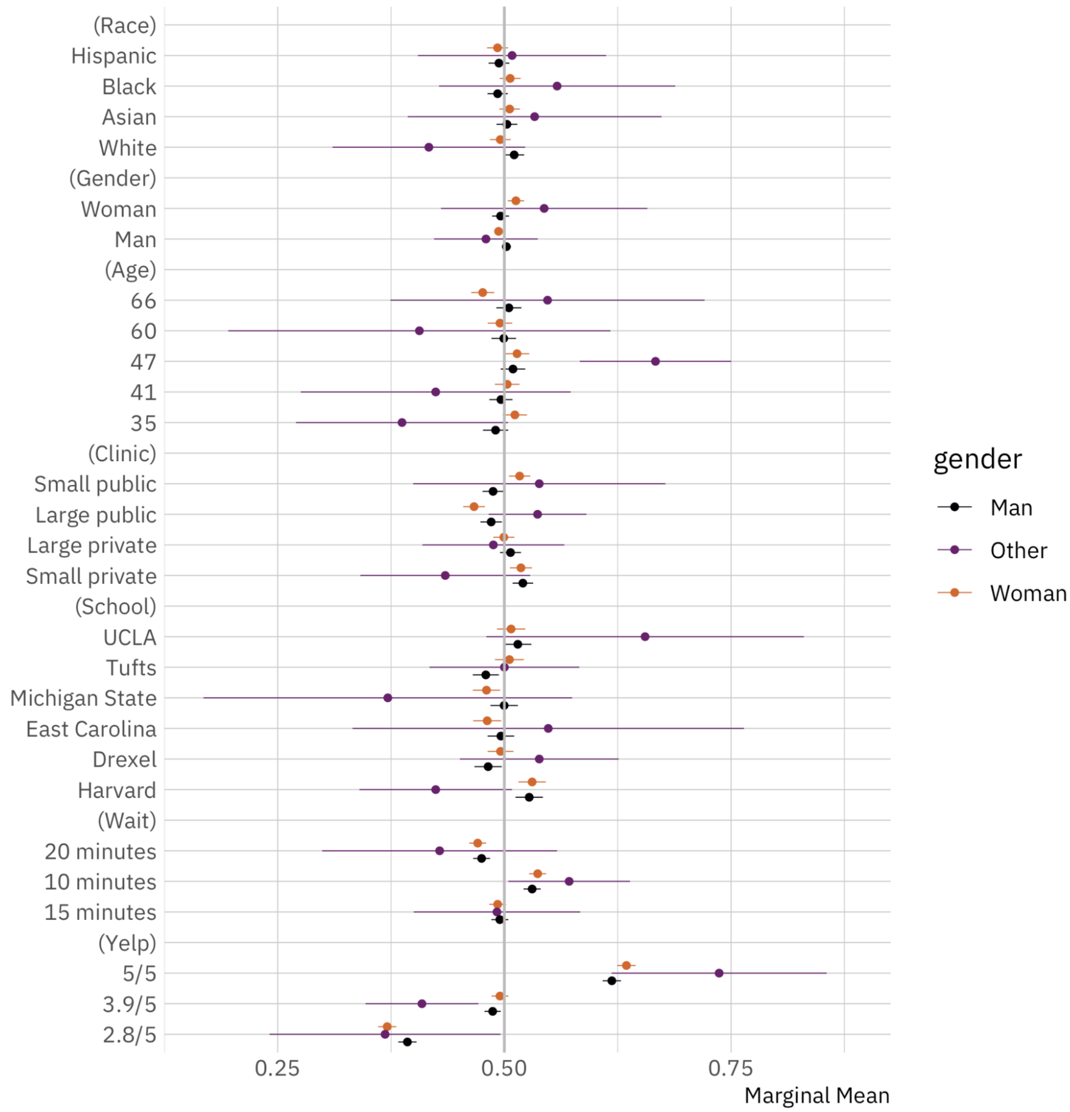
In this SI, we report the results from several additional treatment effect heterogeneity tests. As described in the main text, we explore possible treatment effect heterogeneity by looking at group differences across respondent education, gender, political identification, race, and views on racial discrimination. We find no compelling, consistent differences across groups, especially after adjusting p-values for multiple comparisons. Ultimately, while there is some small variation in our treatment effect heterogeneity results, the overall trend is clear: members of the public do not provide differential treatment to doctors in our task based on race.

Figure SI 2a: Public Preferences for Doctors by Education



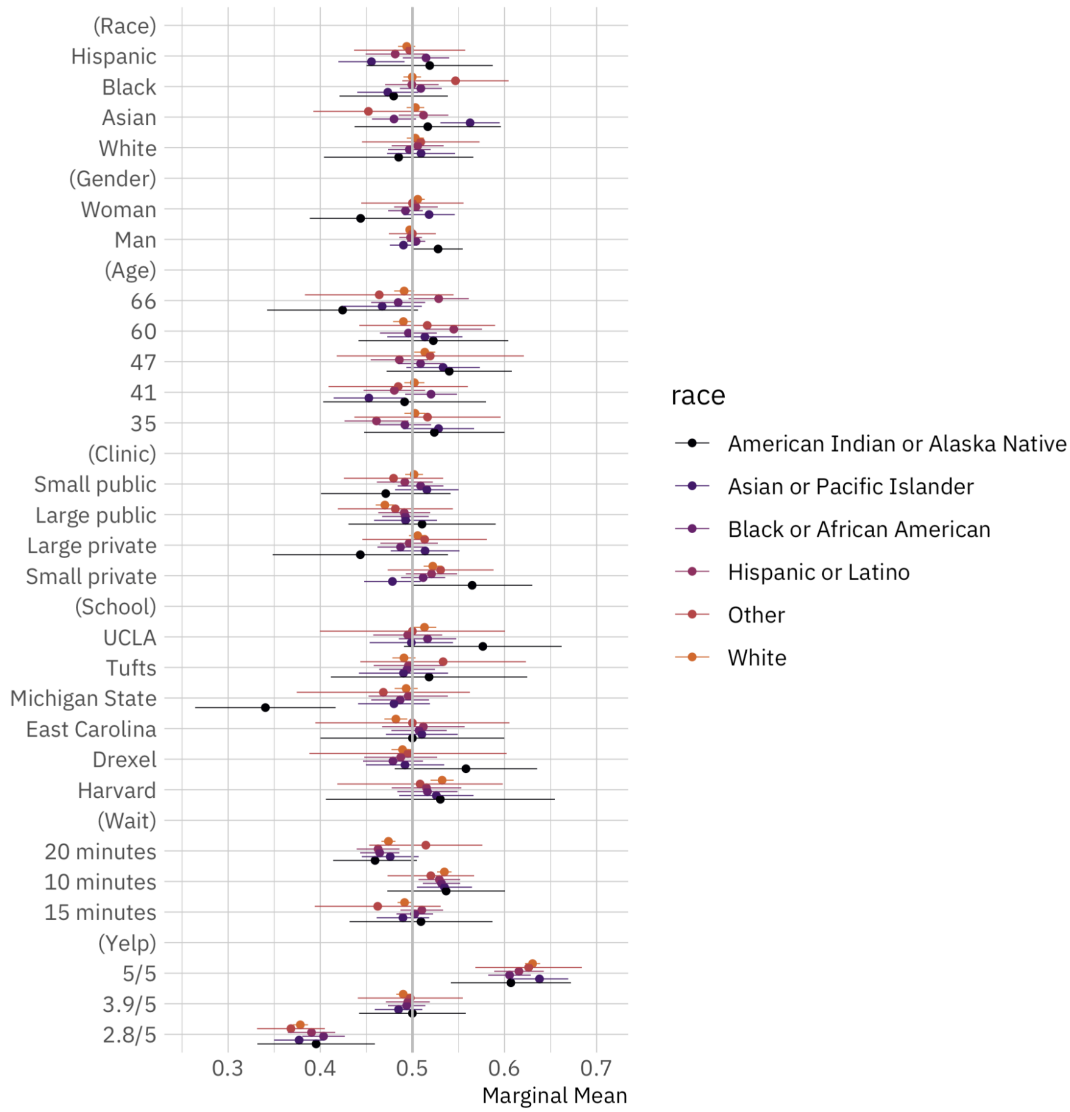
Note: Conditional marginal means for the effect of doctor attributes on survey respondent selection by respondent education level. The circles represent the marginal means while the thin bars denote 95% confidence intervals. Coefficients on the left side of the grey line at 50% indicate that respondents are, all-else-equal, less likely to choose a doctor with the given characteristics on the vertical axis; those on the right are, all-else-equal, more likely to choose a doctor with the given characteristic. The unit of analysis is the respondent-choice profile. The N reported in our models below is the number of respondents (1,498) multiplied by the number of pairwise choices (15) and individuals within those pairs (2). N = 44,940.

Figure SI 2b: Public Preferences for Doctors by Gender



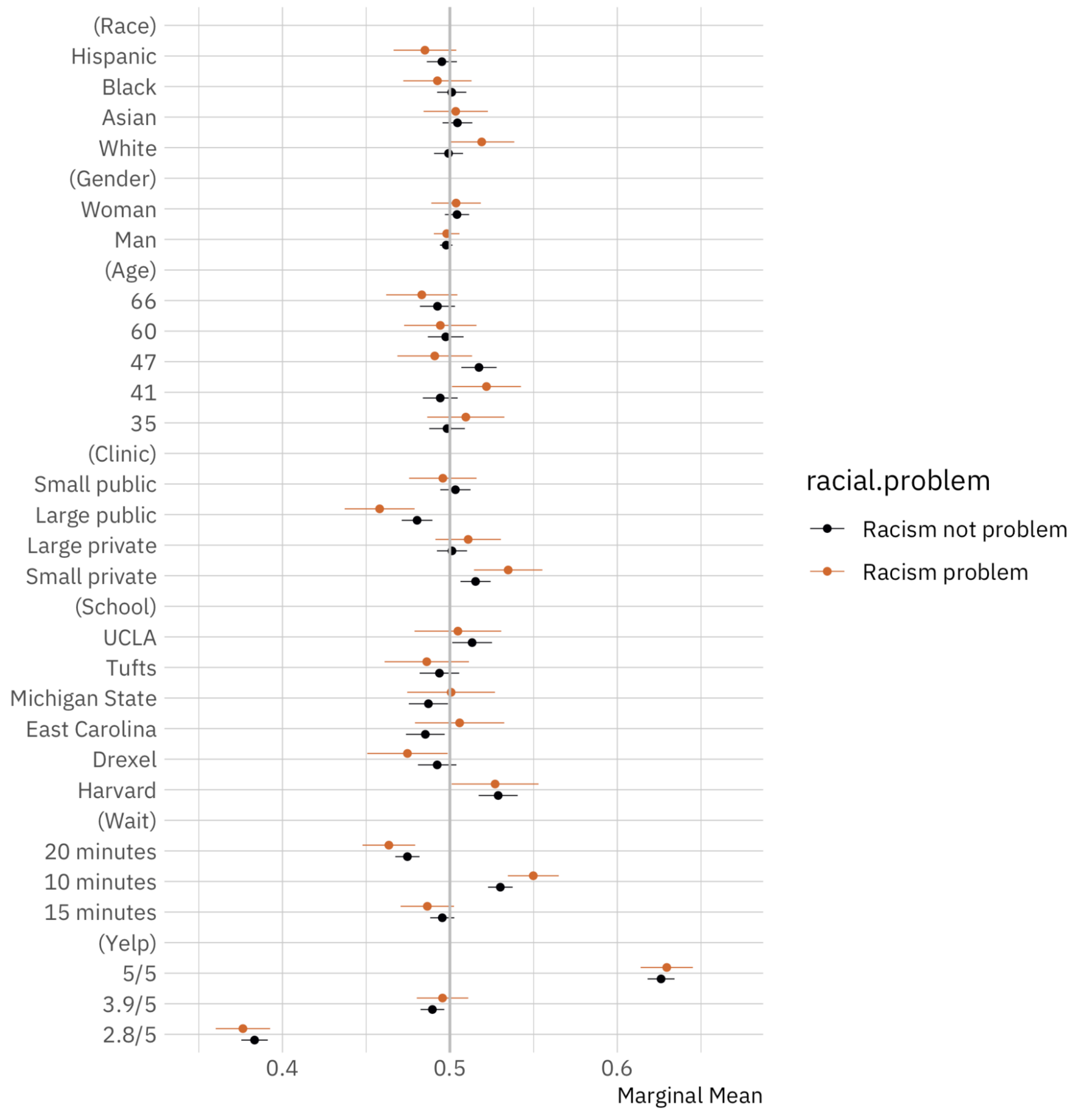
Note: Conditional marginal means for the effect of doctor attributes on survey respondent selection by respondent gender. The circles represent the marginal means while the thin bars denote 95% confidence intervals. Coefficients on the left side of the grey line at 50% indicate that respondents are, all-else-equal, less likely to choose a doctor with the given characteristics on the vertical axis; those on the right are, all-else-equal, more likely to choose a doctor with the given characteristic. The unit of analysis is the respondent-choice profile. The N reported in our models below is the number of respondents (1,498) multiplied by the number of pairwise choices (15) and individuals within those pairs (2). N = 44,940.

Figure SI 2c: Public Preferences for Doctors by Race



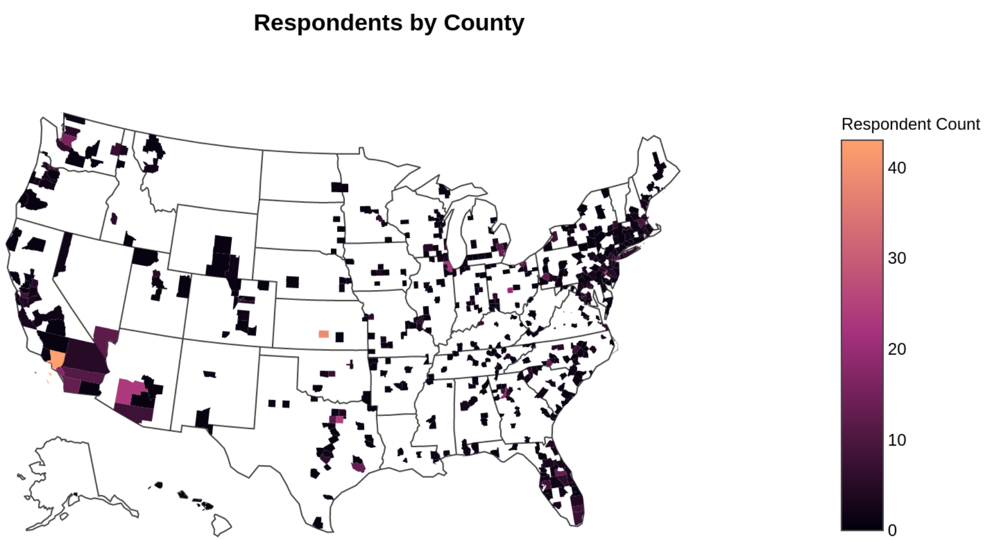
Note: Conditional marginal means for the effect of doctor attributes on survey respondent selection by respondent race. The circles represent the marginal means while the thin bars denote 95% confidence intervals. Coefficients on the left side of the grey line at 50% indicate that respondents are, all-else-equal, less likely to choose a doctor with the given characteristics on the vertical axis; those on the right are, all-else-equal, more likely to choose a doctor with the given characteristic. The unit of analysis is the respondent-choice profile. The N reported in our models below is the number of respondents (1,498) multiplied by the number of pairwise choices (15) and individuals within those pairs (2). N = 44,940.

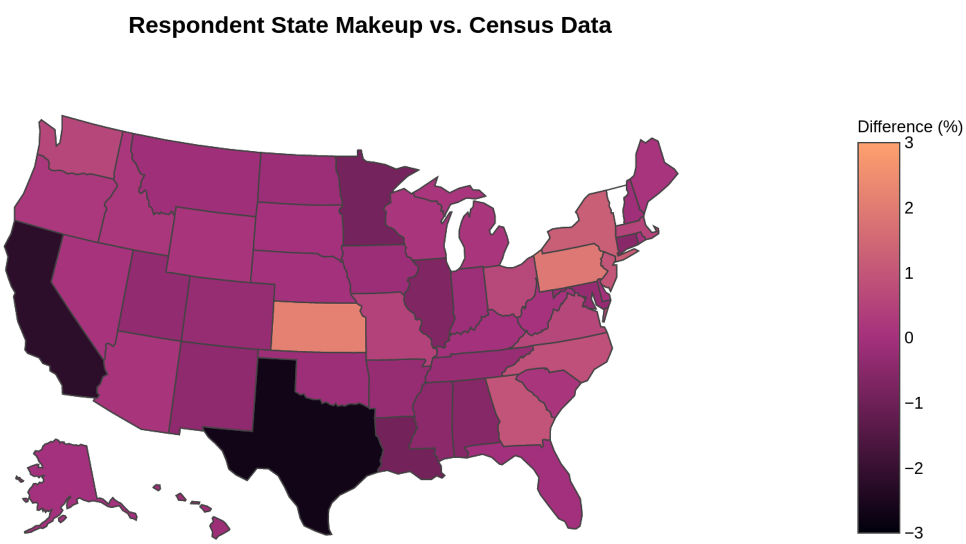
Figure SI 2d: Public Preferences for Doctors by Views on Racism



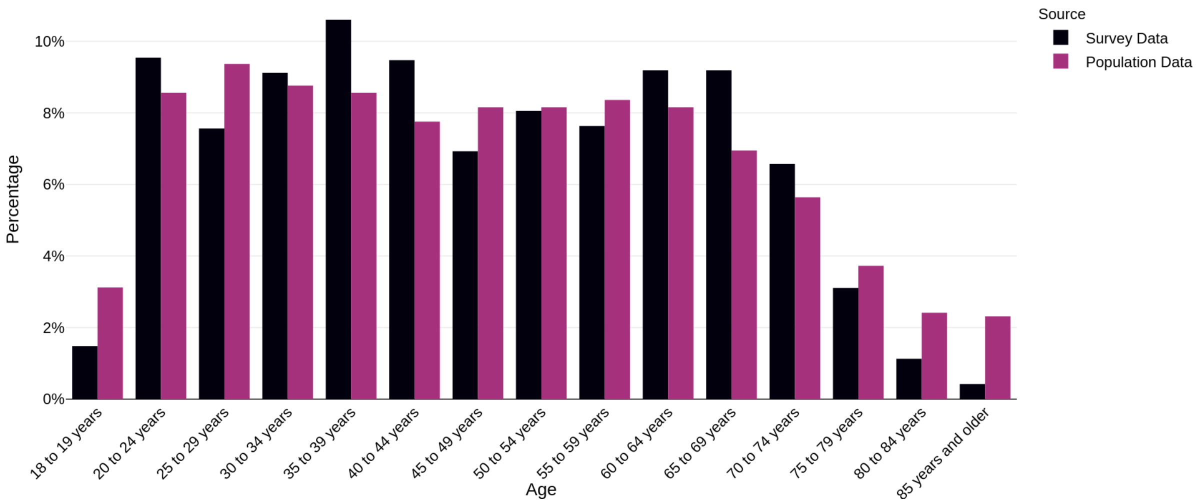
Note: Conditional marginal means for the effect of doctor attributes on survey respondent selection by respondent views on racism. The circles represent the marginal means while the thin bars denote 95% confidence intervals. Coefficients on the left side of the grey line at 50% indicate that respondents are, all-else-equal, less likely to choose a doctor with the given characteristics on the vertical axis; those on the right are, all-else-equal, more likely to choose a doctor with the given characteristic. The unit of analysis is the respondent-choice profile. The N reported in our models below is the number of respondents (1,498) multiplied by the number of pairwise choices (15) and individuals within those pairs (2). N = 44,940.

SI 3: Study demographics

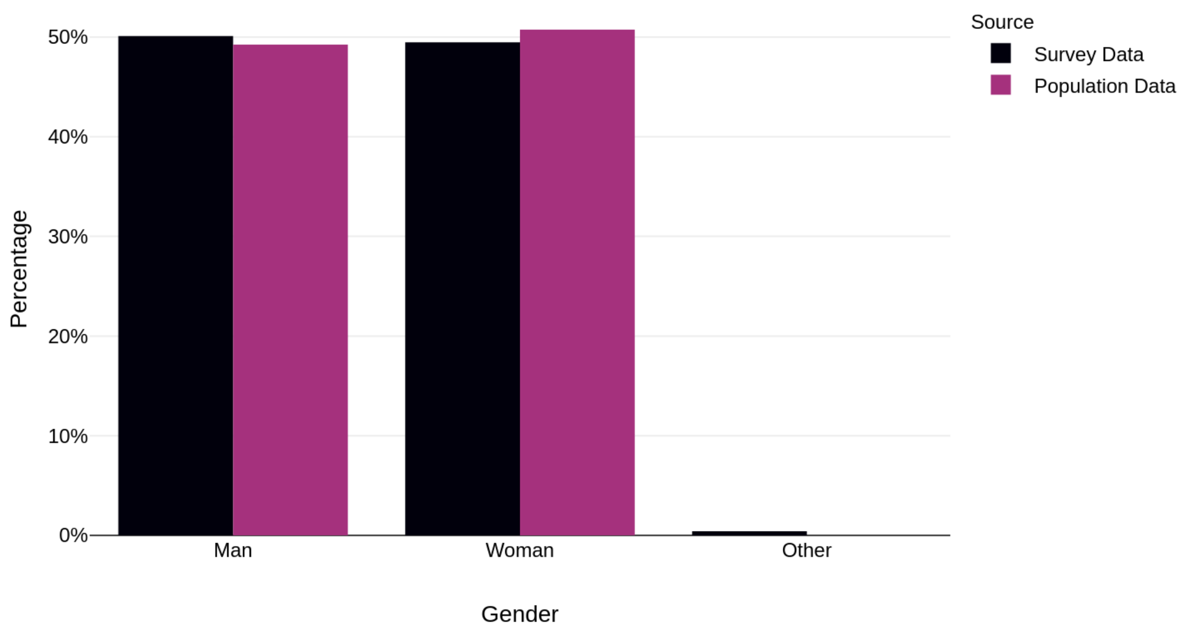




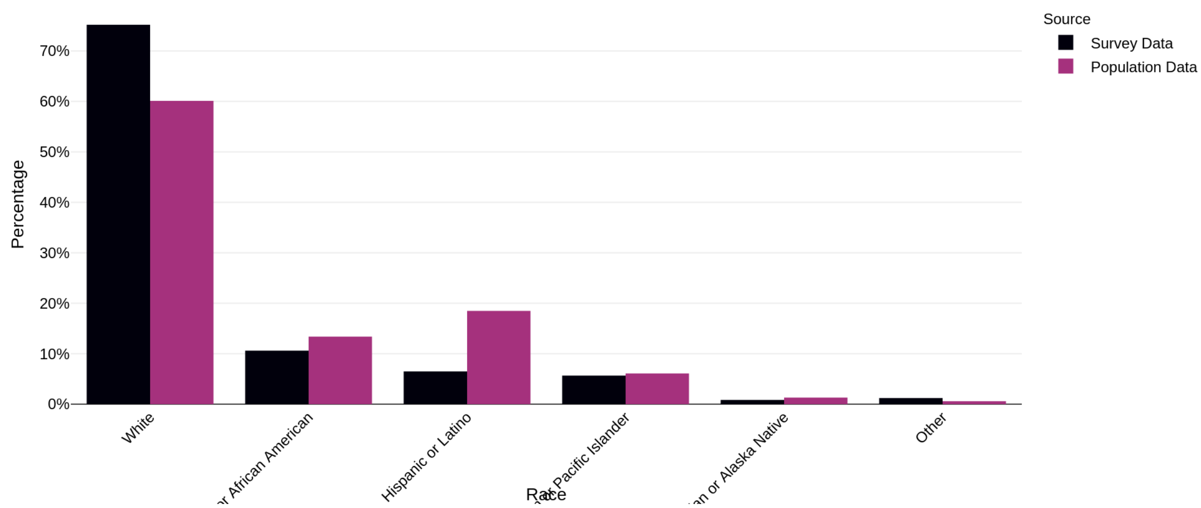
Estimated Respondent Age by Birth Year



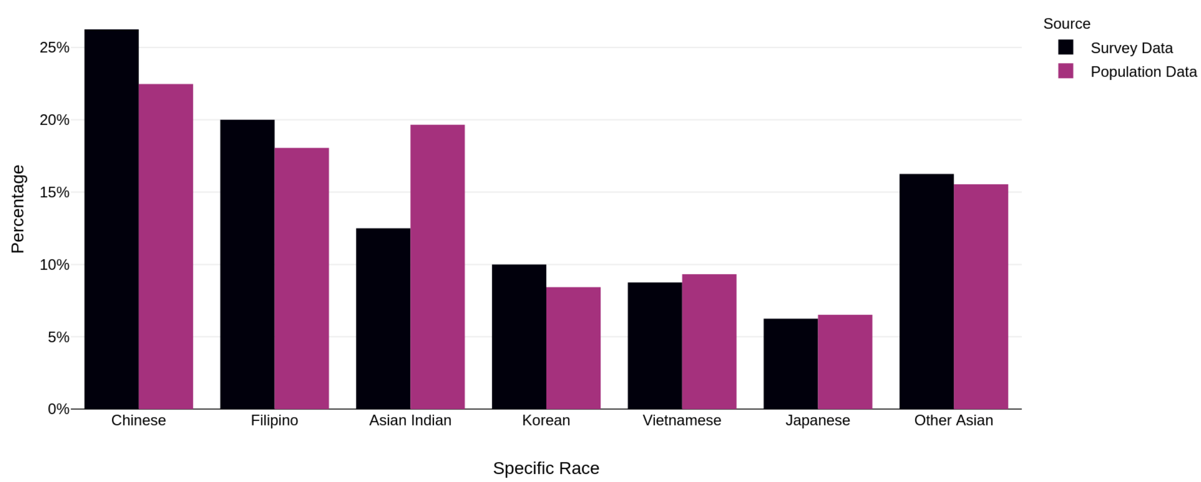
Gender



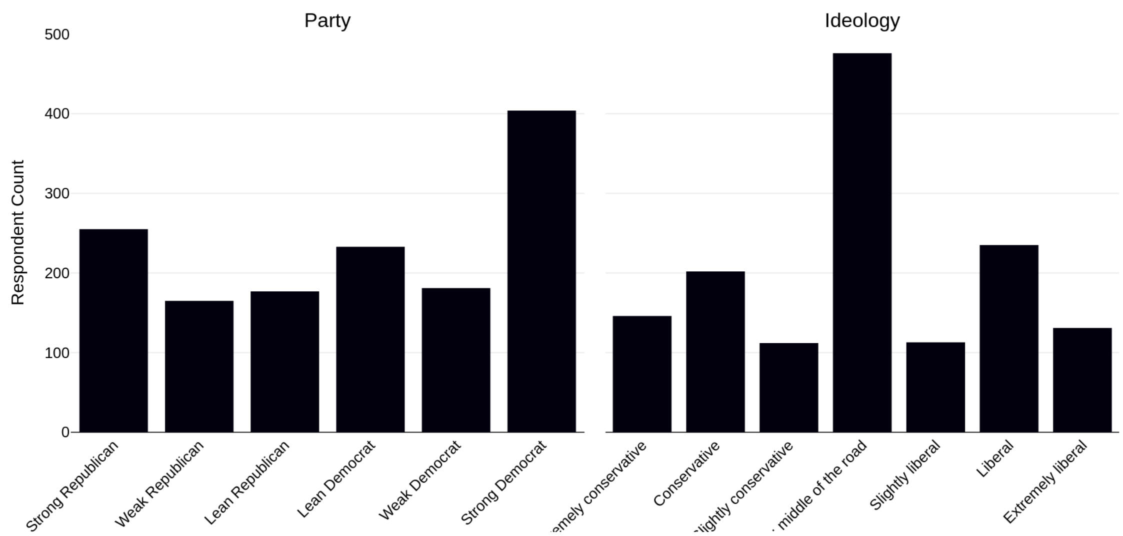
Race



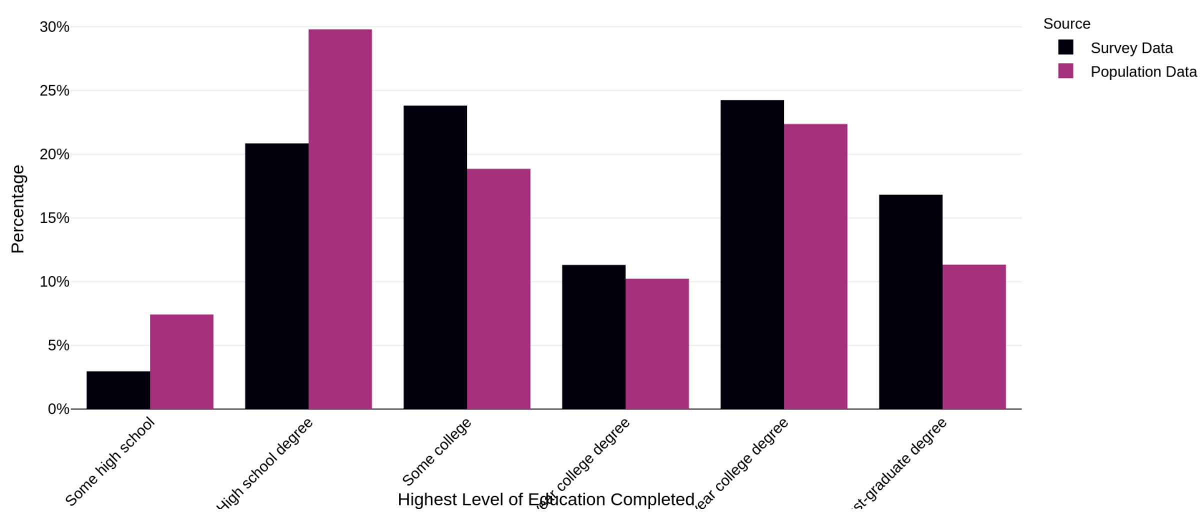
Specific Race



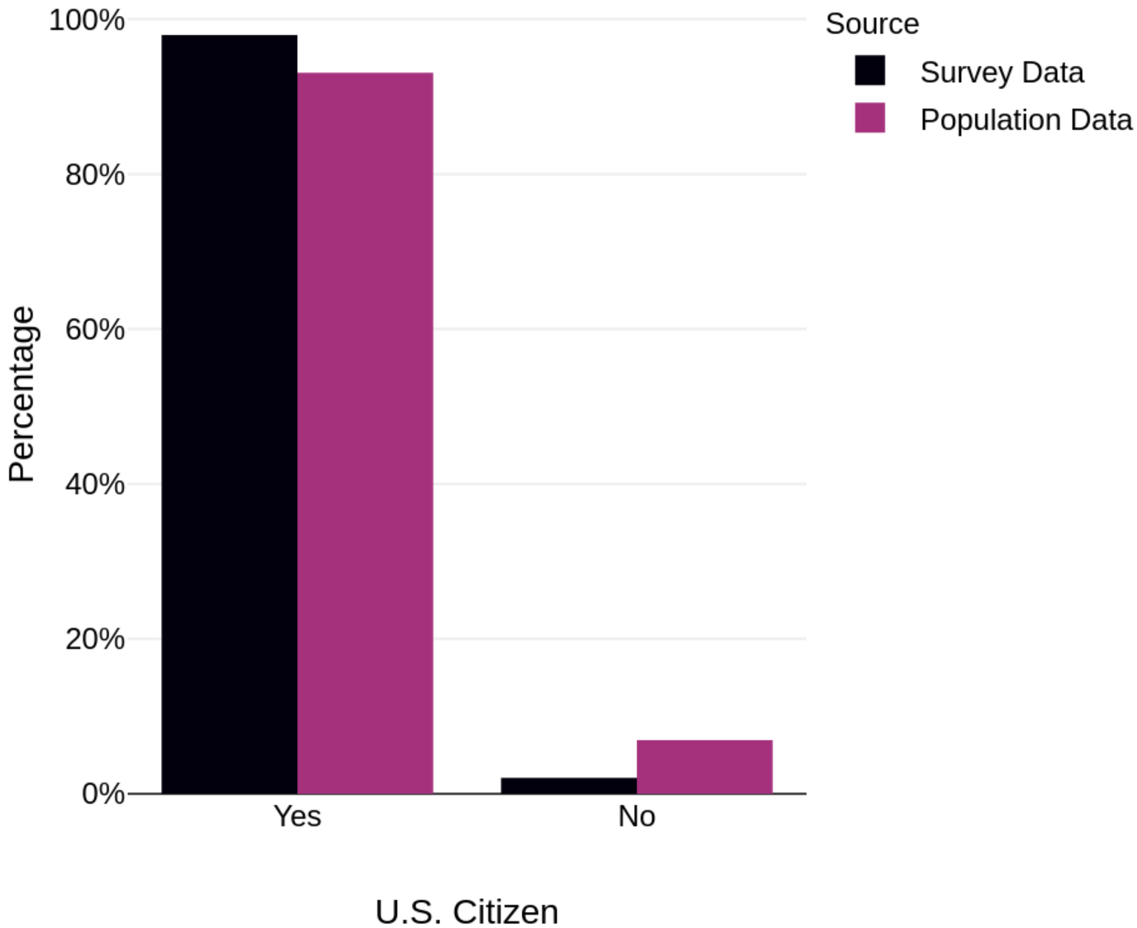
Political Stance



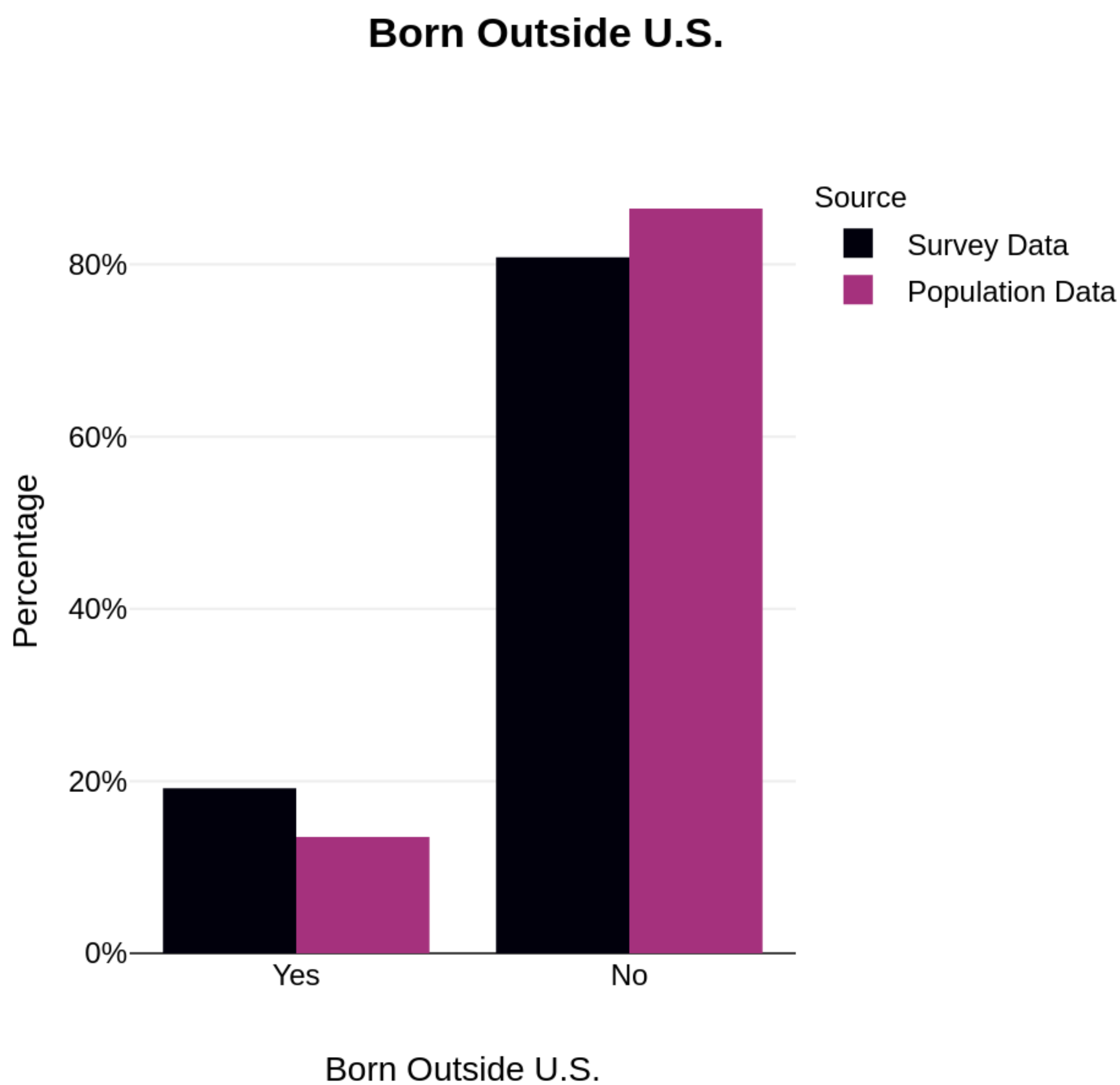
Highest Completed Level of Education



U.S. Citizen



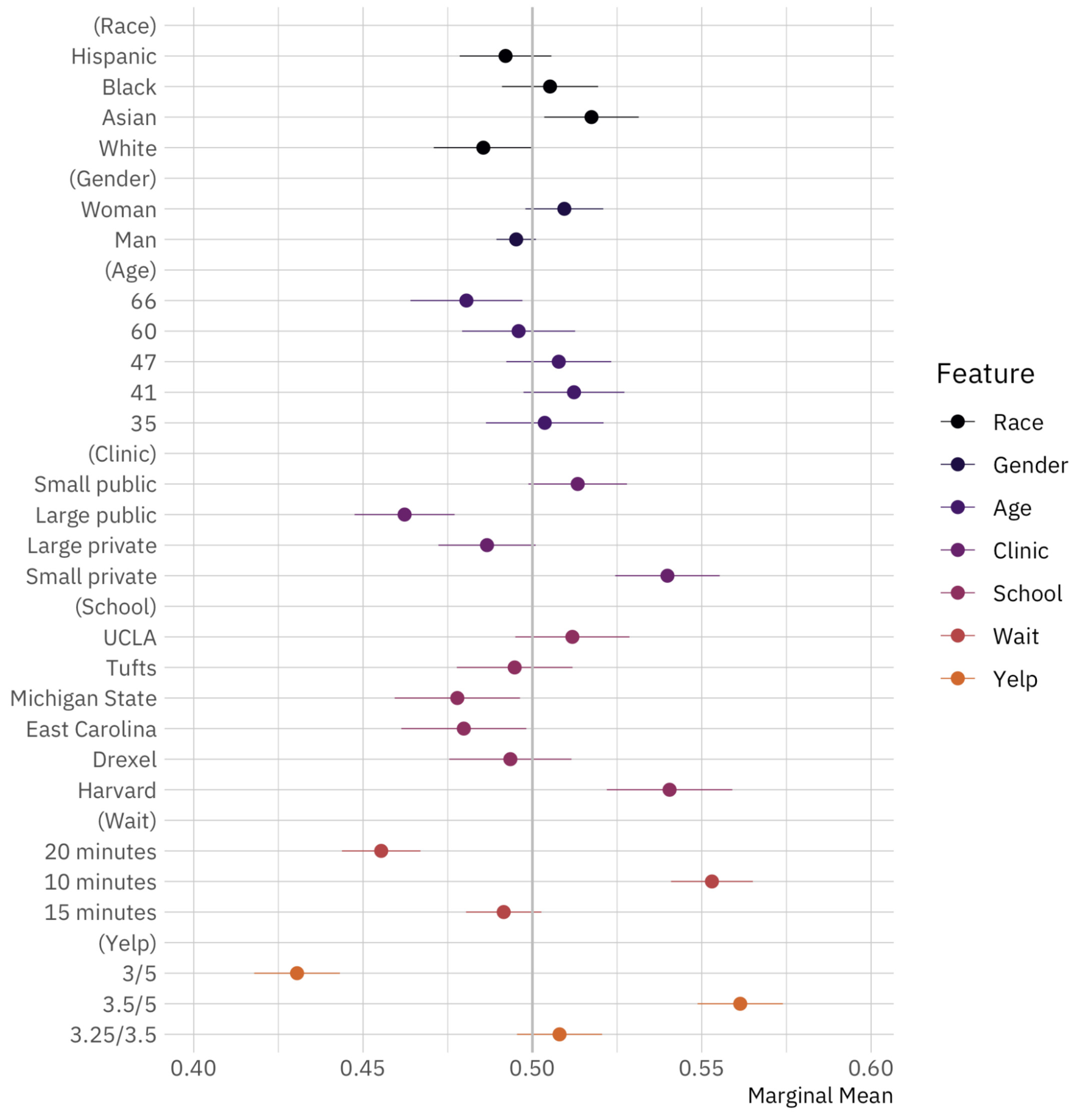
100%
80%
60%
40%
20%
0%
Percentage
Yes
No
U.S. Citizen
Source
Survey Data
Population Data


Born Outside U.S.
Source
Survey Data
Population Data
Percentage
80%
60%
40%
20%
0%
Yes
No
Born Outside U.S.

SI 4: Pilot marginal means

In this SI, we report the results from our two pilot studies (n = 174, n = 330). In these analyses, we pooled the data from the pilots. Our estimand is the marginal mean. As described in the main text, we find similar results in our pilots to the main study. That is there is no evidence that respondents discriminated against doctors based on their race in our conjoint task.

Figure SI 4a: Public Preferences for Doctors

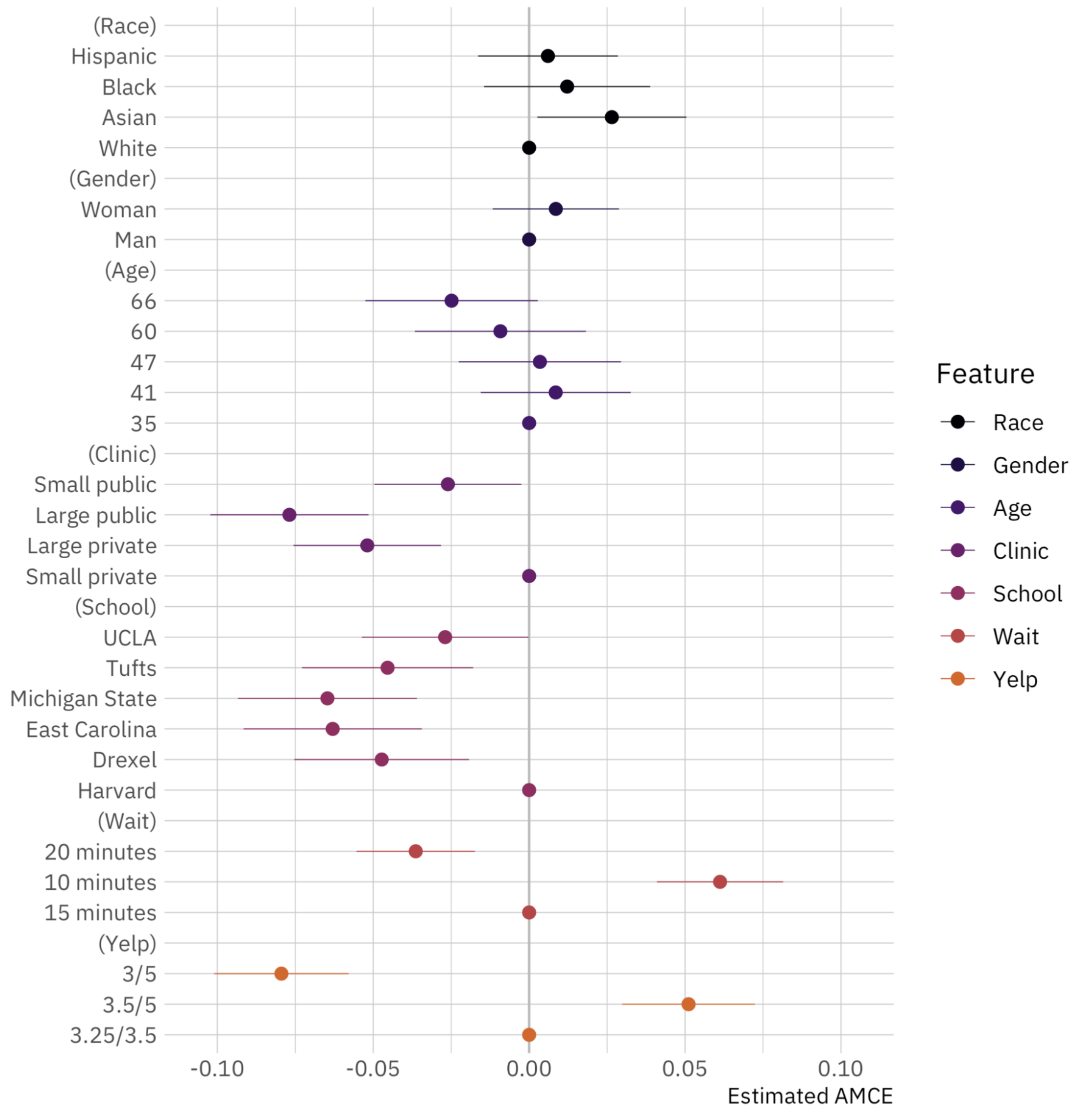


Note: Marginal means plot for the effect of doctor attributes on survey respondent selection. The circles represent the marginal means while the thin bars denote 95% confidence intervals. Coefficients on the left side of the grey line at 50% indicate that respondents are, all-else-equal, less likely to choose a doctor with the given characteristics on the vertical axis; those on the right are, all-else-equal, more likely to choose a doctor with the given characteristic. The unit of analysis is the respondent-choice profile. The N reported in our models below is the number of respondents (504) multiplied by the number of pairwise choices (15) and individuals within those pairs (2). N = 15,120.

SI 5: Pilot AMCEs

In this SI, we report the results from our two pilot studies (n = 174, n = 330). In these analyses, we pooled the data from the pilots. Our estimand is the average marginal component effect. As described in the main text, we find similar results in our pilots to the main study. That is there is no evidence that respondents discriminated against doctors based on their race in our conjoint task.

Figure SI 5a: Public Preferences for Doctors



Note: AMCE plot for the effect of doctor attributes on survey respondent selection. The circles represent the marginal means while the thin bars denote 95% confidence intervals. Coefficients on the left side of the grey line at 50% indicate that respondents are, all-else-equal, less likely to choose a doctor with the given characteristics on the vertical axis; those on the right are, all-else-equal, more likely to choose a doctor with the given characteristic. The unit of analysis is the respondent-choice profile. The N reported in our models below is the number of respondents (504) multiplied by the number of pairwise choices (15) and individuals within those pairs (2). N = 15,120.