

INVESTIGATOR RACIAL DIVERSITY AND CLINICAL TRIAL PARTICIPATION

By Marcella Alsan, Romaine A. Campbell, Lukas Leister, Ayotomiwa Ojo*

Abstract

We investigate whether increased racial diversity of clinical trial principal investigators could increase the enrollment of Black patients, which currently lags population and disease-burden. We conducted a survey experiment in which respondents were shown a photo of a current NIH investigator in which race (Black/White) was randomized. Sex was also randomized as a relevant benchmark. Black respondents reported 0.35 standard deviation units higher interest in participating in a clinical study led by a race concordant investigator (a 12.6% increase). Sex concordance had no effect. Further analyses indicate that perceived trustworthiness and attractiveness are the most important factors explaining these results.

*Alsan: Kennedy School of Government, Harvard University, 79 John F. Kennedy St., Rubenstein Bldg R403, Cambridge, MA 02138 (email: marcella_alsan@hks.harvard.edu); Campbell: Department of Economics, Harvard University, Cambridge, MA 02138 (email: rcampbell@g.harvard.edu); Leister: Department of Economics and Business, Universitat Pompeu Fabra, Carrer de Ramon Trias Fargas, 25, 08005 Barcelona, Spain (email: lukas.leister@upf.edu); Ojo: Northwestern Memorial Hospital, Chicago, IL 60611 (email: ayotomiwa.ojo@nm.org). We note that Marcella Alsan was part of the Expert Committee for the National Academies of Science, Engineering and Medicine Report on Underrepresentation in Clinical Research referenced herein. Nikhil Shankar and Anne Fogarty provided excellent research assistance. The study is approved by IRB at Harvard University (IRB21-0864) and registered at the AEA RCT Registry (AEARCTR-0011926) and ClinicalTrials.gov (NCT04938895). We gratefully acknowledge financial support from the National Science Foundation under Grant No. DGE1745303.

I Introduction

Despite comprising 13 percent of the U.S. population and suffering disproportionately from premature morbidity and mortality, Black Americans make up only 5 percent of clinical trial participants (Alegria et al., 2021). In qualitative research, Black Americans commonly cite a lack of trust in academic and research institutions as well as investigators as the most significant attitudinal barrier to their participation in research (Scharff et al., 2010). Low enrollment of racial and ethnic minorities may compromise the generalizability of research findings and affect the opportunity for minority groups to benefit from medical innovation (Manski, Mullahy and Venkataramani, 2023). While rigorous causal evidence on the implications of low representation for patient health is sparse, simulations and immunologic evidence suggest that lower clinical trial participation among racial minorities has contributed to growing racial disparities in diabetes morbidity and cancer survival (Basu and Gujral, 2020; Awidi and Al Hadidi, 2021). In addition, a recent randomized online experiment among physicians and patients found that a lack of representativeness in clinical trials negatively affects Black patients' beliefs about stated drug efficacy and reduces the willingness of physicians who treat Black patients to prescribe new medications (Alsan et al., Forthcoming).

Strengthening diversity in clinical trials has been a policy priority for leading scientific institutions and policymakers. In 2015, the U.S. Food and Drug Administration (FDA) launched a five-year plan to increase the racial diversity and representativeness of clinical trials (FDA, 2014). However, a 2022 analysis of clinical trial data found that Black patients continue to be underrepresented, with less than 20 percent of trials reporting treatment or side effects data for Black patients (Green et al., 2022). In response, the FDA has issued draft guidance (FDA, 2022), and the U.S. Congress has introduced legislative proposals to improve diversity in clinical trial research through initiatives such as public awareness campaigns and the use of mobile technologies (H.R. 6000, H.R. 5030, H.R. 6584, H.R. 3085).

Effective strategies to inform and encourage Black patients to participate in trials are

needed to complement these policy measures. The recent *National Academies of Science Engineering and Medicine Report on “Improving Representation in Clinical Trials and Research”* outlined several potential strategies to accomplish this aim. These recommendations included increasing remuneration for study participation and fast-tracking patent applications for products tested on diverse samples. The report also highlighted the importance of ensuring a diverse and inclusive workforce, particularly in leadership roles, for all those involved in clinical trials research (National Academies of Sciences and Medicine, 2022, p. 133). Notably, both the National Institute on Minority Health and Health Disparities and the National Institute on Drug Abuse Office of Diversity and Health Disparities have established initiatives to recruit, support, and fund minority students and researchers (NIMHD, 2022; NIH, 2022a). However, the effectiveness of improved racial diversity among principal investigators for increasing clinical trial participation among Black adults has not been established empirically.

This project aims to fill the gap. We conducted a randomized survey experiment on a representative sample of Black Americans, randomizing exposure to race concordant photographs of current National Institutes of Health (NIH) investigators and assessing stated willingness to participate in a future clinical trial. Sex concordance was cross-randomized as a benchmark because it constitutes another salient dimension of similarity between subjects and investigators. We demonstrate that exposure to a race concordant investigator increases willingness to participate in medical research among Black adults by 0.35 standard deviation units or 12.6 percent. In contrast, sex concordance has no statistically significant effect.

Our results are robust to the inclusion of respondent demographics and machine learning (ML)- generated ratings of image characteristics, such as image quality and the emotions expressed by the investigator (Serengil and Ozpinar, 2021; Ocampo, 2022). We conduct three additional checks to validate our results. A common concern with survey experiments is that findings may be driven by demand effects. To assess this, we exclude the

6 percent of participants who mentioned race- or sex-related terms in an open-text question about the study's purpose from our analysis, and our results persist. We also add image fixed effects to our specification and demonstrate that the estimated fixed effects within treatment category are statistically indistinguishable from one another, suggesting that the findings are not driven by "outlier" images. Finally, we show that the racial concordance effect is robust to the inclusion of risk aversion, generic trust in others, and altruism — preferences that could be potential confounders in our setting.

So, what *is* driving our effect? To explore mechanisms, we query subjects on the perceived attributes of the investigator on several margins after introducing the experimental variation. These include perceived age, attractiveness, educational attainment, relative quality, and trustworthiness. We chose these attributes based on prior literature supporting their importance in persuasion. Bertrand et al. (2010) randomized features of a direct mailer for loans in South Africa and found that an attractive photo was equivalent to a 25 percent lower interest rate on the loan. Educational attainment and quality can signal professionalism or legitimacy and have experimentally been shown to garner a return in the medical labor market (Deming et al., 2016). Age might be perceived as greater experience or outmoded information and thus may be a positive or negative signal. Lastly, trustworthiness is important given the history of exploitation and abuse particularly targeting Black Americans (Brandt, 1987; Roberts, 1997; Alsan and Wanamaker, 2018; Washington, 2006). We decompose our overall effect by iteratively including each investigator feature in our main specification individually. We find that the perceived trustworthiness of the researcher has a substantial explanatory effect, almost doubling the fit of the model (*i.e.*, R^2 squared from 0.116 to 0.218) and decreasing the race concordant treatment effect by about one-third. Perceived attractiveness displays a similar pattern, although the two are only modestly correlated in the data. Perceived quality matters much less than either trustworthiness or attractiveness, and no other attribute is shown to matter. A formal decomposition method following Gelbach (2016) confirms the importance of trustworthiness and

attractiveness.

Our paper is related to several strands of literature, including the importance of trust in the delivery of medical care (Alsan and Wanamaker, 2018; Banerjee et al., 2020) and the role of racial concordance in improving health outcomes (Alsan, Garrick and Graziani, 2019; Alsan and Eichmeyer, Forthcoming; Anderson et al., 2020; Frakes and Gruber, 2022; Hill, Jones and Woodworth, 2020; Greenwood et al., 2020). We add to this literature by rigorously examining the decision to participate in medical research, which is important for innovation and confidence in research results (Alsan et al., Forthcoming). Methodologically, we borrow from Ludwig and Mullainathan (2023) by combining real-world images with ML-rated attributes and using the output in a choice experiment for a policy-relevant topic.

The paper proceeds as follows. In Section II, we provide more information on racial gaps in investigator funding and leadership as well as previous studies on concordance. In Section III, we describe our sample and intervention, including investigator image selection and outcomes. We then present results in Section IV. Finally, we conclude with a discussion of the potential implications of our results.

II Background

Several studies have highlighted the gap between the demographics of clinical trial samples and those in the population who either have the relevant condition or would be primary targets for prevention. In a recent review of cancer trials, Nazha et al. (2019) find that only 5 percent of the samples were Black across trials for novel checkpoint inhibitors. While COVID-19 vaccine trial samples were more representative at nearly 9 percent Black (Artiga et al., 2022), these trials were unique in terms of the considerable scrutiny from the press, pressure by the U.S. government, and major advance purchase agreements. Recent trials for Alzheimer’s, a disease that has a larger impact on Black communities, continue

to display the pre-COVID pattern for trial representativeness: Eli Lilly's 1,736-person donanemab trial included only 19 Black participants who received the drug (Reardon, 2023).

Against this backdrop, diversity among investigators has been championed as a potential path to reduce the racial recruitment gap. Diverse investigators might ask different research questions, develop more successful outreach, enrollment, and retention strategies for minorities in trials, and increase trust both in the specific studies they are conducting and in the research process more broadly. Yet, there is no causal evidence to establish the effectiveness of this approach. Previous research has documented that physician-patient race, sex, or ethnic concordance can improve patient satisfaction (Takeshita et al., 2020; Cooper et al., 2003), patient trust in health information (Loeb et al., 2023), utilization of preventative services (Alsan, Garrick and Graziani, 2019), mortality among those with manageable illnesses (Frakes and Gruber, 2022), neonatal outcomes (Greenwood et al., 2020), and survival of heart attacks (Greenwood, Carnahan and Huang, 2018). Religious concordance between medical experts and patients increased intent to take up the COVID-19 vaccine among American Christians (Chu, Pink and Willer, 2021). Because the investigator-trialist relationship is distinct from the doctor-patient relationship, evidence specific to this setting is needed.

Yet the topic is difficult to study for two main reasons. First, investigators are not randomly assigned to studies. Second, systemic discrimination and structural inequality have led to persistent racial disparities in access to education, promotion, and funding opportunities (Laurencin et al., 2023; National Academies of Sciences and Medicine, 2023). Consequently, Black investigators comprise less than 4 percent of NIH-funded researchers (NIH, 2022*b*). The funding gap was first documented by Ginther et al. (2011) who found that African American applicants were 13 percentage points less likely to receive NIH investigator-initiated funding than White investigators. While recent updates suggest that racial representation and funding rate gaps have narrowed since 2011, non-White NIH principal investigators remain significantly underrepresented and underfunded (NIH,

2023). Therefore, we devised an online trial in which we randomized profiles of real Black and White NIH investigators to subjects.

III Experimental Design

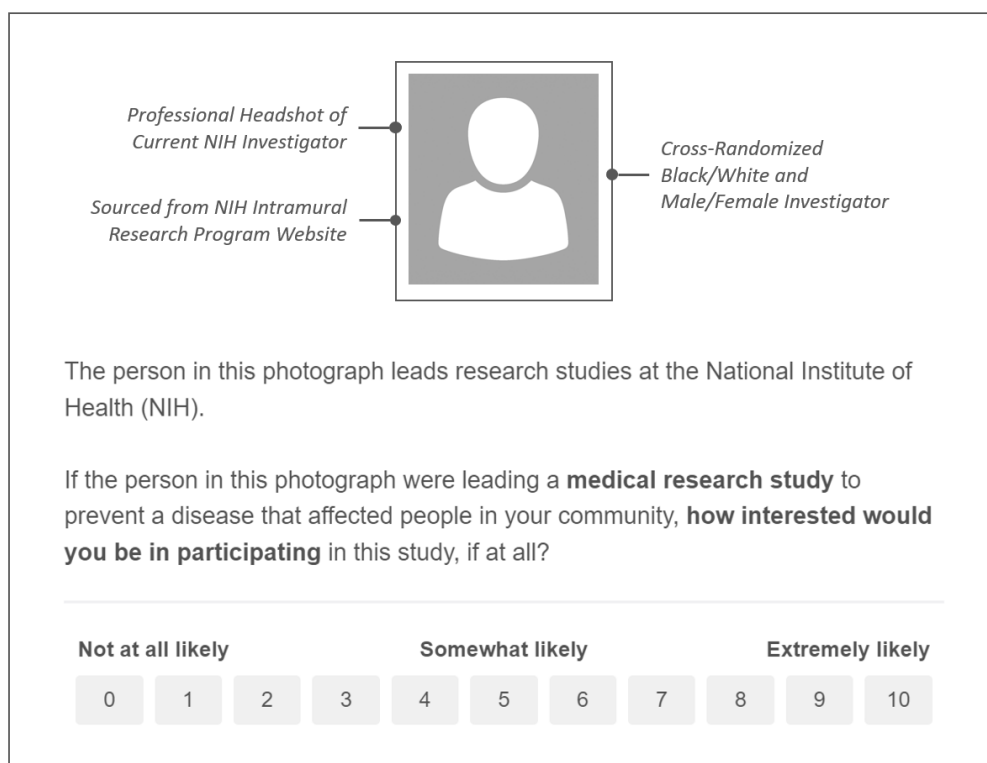
The goal of the experiment was to test whether racial concordance with clinical trial investigators affects patient willingness to participate in clinical studies. In addition, we sought to probe the mechanisms underlying a potential effect. We randomly assigned subjects to an image of an NIH investigator and elicited their willingness to participate in a future study as well as their perceptions of the principal investigator. This section describes our experimental design choices in detail.

III.A Study Flow, Description of Treatment, and Outcome Measurement

Study Flow: The overall study flow is presented in Appendix Figure B1. After providing informed consent, subjects answered screening questions (described in Section III.B). Eligible respondents were then randomized to the treatment. Following the intervention, we queried subjects on their willingness to participate in a trial led by the randomly assigned investigator — our primary outcome of interest. To assess potential mechanisms, we subsequently asked subjects to rate the investigator on various characteristics that prior literature suggests are important for persuasion (details below). Finally, we asked about the respondent’s current access to medical care, risk preferences, and included open-response questions about the study’s purpose.

Treatment Description: The intervention is shown in Figure 1. The race (Black/White) and sex (Male/Female) in the image of a current NIH investigator were cross-randomized. The investigator’s head shot was centered and an accurate description of their role as a lead investigator of the NIH was provided.

Figure 1: NIH Investigator Randomization



Notes: Depiction of the Randomized photo of an NIH investigator and the question relating to the primary outcome.

Selection of Investigator Images: Professional head shots of NIH intramural research program (IRP) investigators were sourced from pictures publicly available on the IRP website (NIH, 2022c). The IRP has over 12,000 principal investigators conducting research and is the world's largest biomedical research institution. The investigators are employed by the NIH directly. As Black investigators are less represented on the website and comprise less than 4 percent of NIH-funded researchers, we found fewer usable pictures for Black investigators than for White investigators (NIH, 2022b). We ultimately selected five pictures of Black males, seven of Black females, ten of White males, and eight of White females for the experiment.¹

To ensure the comparability and quality of investigator images, we followed a principled image selection process. First, we hand-selected pictures of investigators that were

¹Images used in the experiment are available to reviewers and other interested parties by specific request.

taken in front of a neutral background and that featured a neutral expression. Second, we used ML-generated ratings of the investigator photos to confirm that picture quality, investigator age, and the emotions conveyed by the NIH investigators were similar across demographic groups. The ML-generated investigator emotions were Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. We used the Python packages *deepface* (Serengil and Ozpinar, 2021) and *image-quality* (Ocampo, 2022) to generate these ratings. The image file size was used as an additional proxy for quality. ML-generated ratings of the investigator pictures were balanced across race (Appendix Table A1). Female investigators were generally rated more favorably (*i.e.*, rated as younger, happier, and less neutral) than their male counterparts (Appendix Table A2). As the same investigator images were shown to both male and female respondents, each image was part of both the sex concordant treatments, so that balance across treatments was assured.

Outcome Description: After viewing the image, participants were asked, “If the person in this photograph were leading a medical research study to prevent a disease that affected people in your community, how interested would you be in participating in this study, if at all?” We chose this generic language so that we could elicit responses from all subjects, not just those with a specific condition. Our primary outcome variable is the answer to this question measured on a ten-point Likert scale and standardized for analysis (see Appendix Table C1).

To probe mechanisms, we asked subjects to rate investigator age, education, trustworthiness, attractiveness, and relative quality. Investigator age was shown on an interval scale (25-34 years, 35-44 years, 45-54 years, 55-64 years, 65+ years), investigator education was measured by highest educational attainment (options ranged from less than a high school degree to completion of a professional degree) and attractiveness, quality, and trustworthiness were measured on five-point Likert scales. See Appendix Table C1 for additional details on the precise wording of questions, answer options, and coding for analysis. For our analysis, we standardized all perceived investigator characteristics ex-

cept education. Education was coded as an indicator variable for whether the respondent selected graduate degree (*i.e.*, M.D., J.D., or Ph.D.), though results are not sensitive to this recoding.

Demand Effects: The experiment was specifically designed to mitigate experimenter demand effects. As the respondents were only shown one investigator, it would have been difficult for them to discern the rationale of the study. Indeed, only 5.6 percent of respondents included race- or sex-related terms in response to the open-ended question “What do you think this study was about?”. We present a wordcloud of subject responses in Appendix Figure B2 with the dominant term being “health.” Nevertheless, we exclude such respondents in robustness checks (see Section IV.B).

III.B Recruitment

We recruited Black survey participants using CloudResearch in February 2022. CloudResearch is an online survey vendor frequently used in social science research. The platform relies on an aggregation of opt-in market research panels coupled with additional data quality checks and selection criteria (Chandler et al., 2019). We limited the sample to Black subjects given their under-representation in clinical trials and previous research showing the importance of concordance in Black populations (Alsan et al., 2021; Alsan and Eichmeyer, Forthcoming; Loeb et al., 2023). Upon completing the consent process, participants answered a set of questions to determine eligibility based on nationality (U.S.-born), self-identified race (non-Hispanic Black), and age (25-64). We also screened out respondents who failed our attention check. Approximately 27 percent of individuals who began the survey were screened out due to eligibility criteria or demographic quotas. Eligible respondents went on to answer basic demographic questions covering sex, income, education, state of birth, health insurance, employment, and existing medical conditions. Our pre-specified recruitment target of 300 participants was determined by power calculations based on responses from a pilot survey.

III.C Descriptive Statistics, Balance, and Attrition

Our experimental sample exceeded our target of 300 and consisted of 323 Black individuals who fulfilled our eligibility criteria (see Section III.B), passed our quality check, and completed the survey. Summary statistics are presented in Appendix Table A3. Our sample is broadly representative of the U.S. non-Hispanic Black population by age, sex, geography, education level, employment, and health insurance status (Appendix Table A3).² Respondents were, on average, 44 years old and about 40 percent had a high school degree or less education. The only characteristic that differed significantly from the U.S. population was employment status, with our survey participants less likely to hold formal employment. Respondent characteristics were balanced across race and sex concordant/discordant arms (see Appendix Tables A4 and A5). Attrition after randomization was low — among respondents who were exposed to the treatment (i.e., saw an image of an investigator), less than one percent did not complete the survey. In Appendix Table A6, we show that attrition was balanced across treatment arms.

IV Estimation and Results

IV.A Main Findings

Given the cross-randomized experimental design, we estimate the following equation:

$$Y_{si} = \beta_0 + \beta_j \text{Concordance}_{si}^j + X_s' \gamma + M_i' \delta + \epsilon_{si} \quad (1)$$

where s is the subject, i is the investigator image, and j signifies either racial or sex concordance. We first estimate separate equations for racial and sex concordance. In subsequent specifications, we simply control for respondent sex as we fail to find differences by sex concordance. Y_i is the standardized outcome capturing interest in participating in

²We used the 2019 American Community Survey and limited to non-Hispanic Black Americans with internet access and English proficiency.

a clinical trial. X_s is a vector of respondent covariates including age, squared age, education level, employment status, and an indicator for whether the respondent resides in the South. M_i is a vector of the ML-generated investigator image characteristics, such as predicted investigator emotion, age, and image quality, as well as image size.

Table 1 displays our main results. Panel (A) reports findings for racial concordance and Panel (B) reports results for sex concordance. Column (1) includes no controls except an indicator for treatment status. We find that survey respondents randomized to an image of a racially concordant NIH investigator are 0.293 standard deviation units more likely to participate in a clinical trial led by the investigator compared with a standardized mean of -0.147 among respondents assigned to a racially discordant investigator. The same pattern does not hold for Panel (B). We do not find detectable effects of sex concordance on participation. Moving across the columns, we add controls for respondent and ML-rated image characteristics. The final column includes both sets of characteristics and is our preferred specification. Race concordant NIH investigators increase willingness to participate by 0.350 standard deviation units. Model fit increases from an R-squared of 0.022 to an R-squared of 0.116. Random assignment to a sex concordant NIH investigator does not significantly affect interest in study participation in any specification.

Table 1: The Effect of NIH Investigator Race or Sex Concordance on Clinical Trial Participation Interest

Panel A: Racial Concordance				
	(1)	(2)	(3)	(4)
Racial Concordance	0.293 (0.110)	0.294 (0.109)	0.322 (0.134)	0.350 (0.132)
Control Mean	-0.147	-0.147	-0.147	-0.147
Panel B: Sex Concordance				
	(1)	(2)	(3)	(4)
Sex Concordance	-0.013 (0.112)	-0.037 (0.111)	-0.001 (0.111)	-0.010 (0.108)
Control Mean	0.006	0.006	0.006	0.006
Observations	323	323	323	323
Respondent Covariates	No	Yes	No	Yes
Image Covariates	No	No	Yes	Yes

Notes: OLS estimates of the willingness to participate in a future clinical trial led by race (Panel A) or sex concordant (Panel B) NIH investigators. Outcomes are standardized to have a mean of 0 and standard deviation of 1. Column (1) includes only an indicator for treatment. Column (2) adds respondent characteristics (age, squared age, education, employment, and South/non-South region). Column (3) adds ML-rated image characteristics (predicted investigator emotion, age, image quality, image size), and Column (4) includes both. Robust standard errors are shown in parentheses.

Next, we explore potential underlying mechanisms for the effect of investigator racial concordance. To do so, we use Equation 1, placing investigator characteristics on the left-hand side (Table 2). On average, subjects perceived race concordant investigators as younger and rated them more positively on all characteristics, including 0.495 standard deviation units more qualified, 0.408 standard deviation units more attractive, and 0.309 standard deviation units more trustworthy (Panel A). Sex concordance does not significantly affect perceived image characteristics (Panel B).

Table 2: Racial and Sex Concordance on Perceived of Investigator Characteristics

Panel A: Racial Concordance					
	Age	Education	Quality	Attractiveness	Trustworthiness
	(1)	(2)	(3)	(4)	(5)
Racial Concordance	-0.365 (0.124)	0.131 (0.066)	0.495 (0.134)	0.408 (0.139)	0.309 (0.138)
Control Mean	0.061	0.410	-0.163	-0.175	-0.132
Panel B: Sex Concordance					
Sex Concordance	0.150 (0.107)	-0.076 (0.055)	-0.044 (0.112)	-0.122 (0.111)	-0.073 (0.113)
Control Mean	-0.049	0.494	0.022	0.050	0.031
Observations	323	323	323	323	323
Respondent Covariates	Yes	Yes	Yes	Yes	Yes
Image Covariates	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimates of ratings of investigator characteristics by race (Panel A) or sex (Panel B) concordance. In Column (2), the outcome is an indicator for completion of a graduate degree. Outcomes in Column (1) and Columns (3)-(5) are standardized to have a mean of 0 and standard deviation of 1. All specifications include respondent characteristics (age, squared age, sex, education, employment, and South/non-South region) and ML-rated image characteristics (predicted investigator emotion, age, image quality, image size). Robust standard errors are in parentheses.

We then investigate whether perceived investigator characteristics partially account for our estimated racial concordance effect. In Figure 2, we plot the coefficient estimates from our main specification (hereafter, Model 0) and alternative specifications. Recall that Model 0 regresses clinical trial participation on racial concordance as well as respondent and ML-rated investigator image characteristics (Table 1, Column (4)). In Models 1-5, we iterate through respondents' ratings of perceived investigator characteristics, adding trustworthiness, attractiveness, quality, education, and age to regression models individually (Panel A).³ In Model 6, we saturate the model with *all* perceived investigator characteristics (Panel B). In Panel (C), we plot the R-squared, demonstrating how model fit varies with the inclusion of different perceived investigator characteristics.

We find that perceived investigator trustworthiness is important for unpacking the racial concordance effect. The inclusion of trustworthiness reduces the racial concordance

³See Appendix Table A7 for a correlation matrix of investigator features.

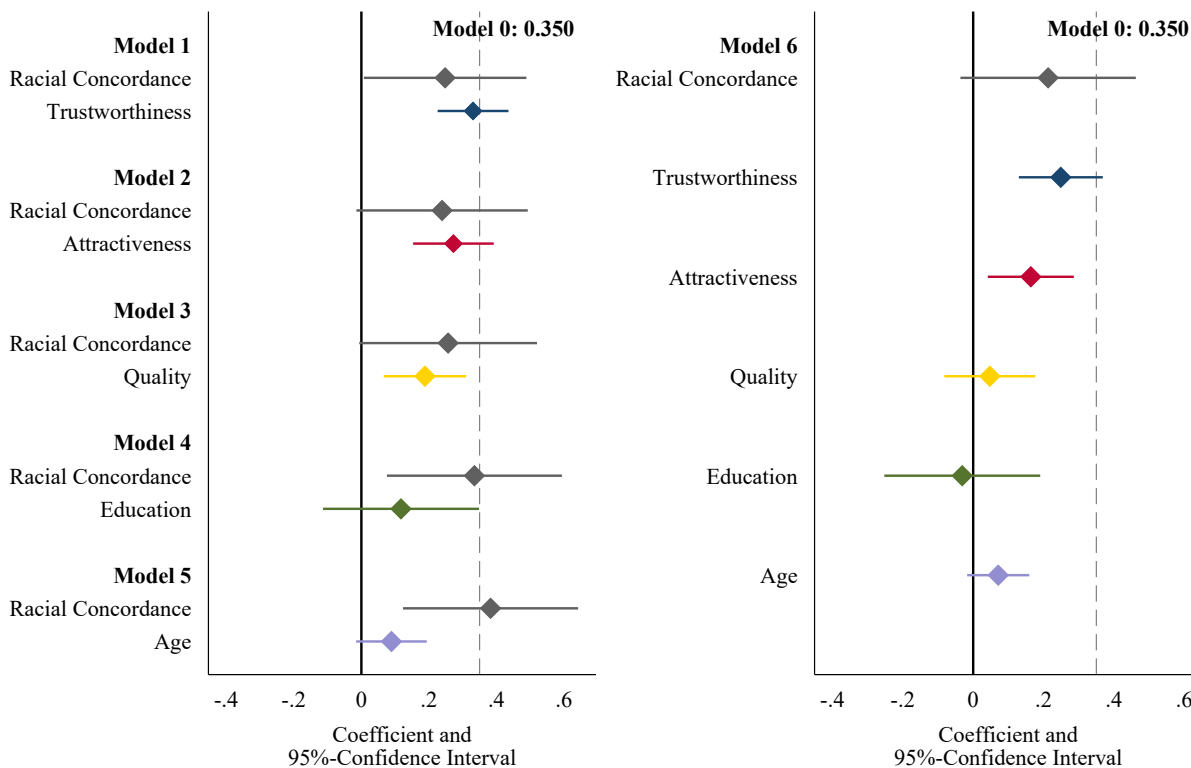
effect by 0.103 standard deviation units and increases the fit of the model from an R-squared of 0.116 to an R-squared of 0.218 (Figure 2, Model 0 and Model 1). We can reject the null hypothesis that the coefficients are the same in the two models (p-value = 0.026). Similarly, the inclusion of either attractiveness or quality causes a decrease of similar magnitude but less significance in the racial concordance effect, reducing the estimated coefficient by roughly 0.110 and 0.093 standard deviation units, respectively (Figure 2, Model 2 and Model 3). Again, we can reject the null hypothesis that the coefficients are the same across the models (Model 0 vs. Model 2: p-value = .005 and Model 0 vs. Model 3: p-value = 0.016). They also both increase model fit but to a lesser degree than trustworthiness (Figure 2, Panel C). Perceived investigator education and age have a negligible impact.

In a multivariate regression of clinical trial participation on treatment and *all* perceived investigator characteristics (Model 6), the R-squared rises to 0.244 (Panel C). In this specification, racial concordance increases interest in future trial participation by 0.213 standard deviation units, compared to 0.350 standard deviation units in Model 0. Among all covariates, trustworthiness has the largest and most significant effect at 0.249 standard deviation units. Taken together, the results suggest that specific investigator characteristics — namely trustworthiness, followed by attractiveness and quality — explain at least part of the effect of racial concordance.

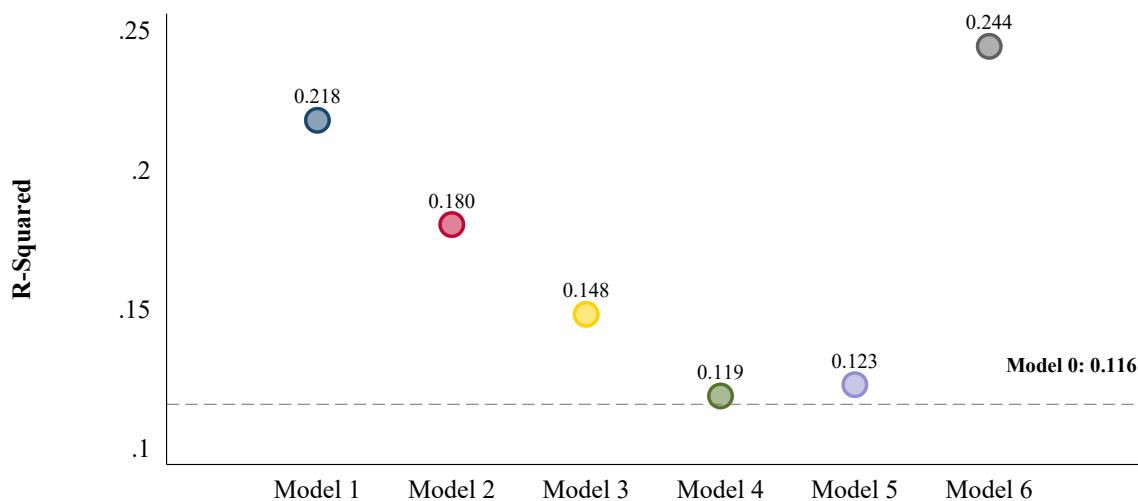
For completeness, we conduct the same exercise for sex concordance (Appendix Figure B3) despite not finding significant sex concordance effects in Model 0. While additions of investigator trustworthiness (Model 1), attractiveness (Model 2), and education (Model 4) flip the sign of the treatment coefficient, it remains small and insignificant across all specifications. We cannot reject the null that the treatment effect is the same across all models. Across all models, perceived trustworthiness is a positive, statistically significant, and robust predictor of trial participation intent (Appendix Tables A8 and A9).

Figure 2: The Effect of NIH Investigator Racial Concordance and Perceived Characteristics on Clinical Trial Participation Interest

(a) Coefficient Plot



(b) Model Fit Plot



Notes: Figure plots OLS estimates from the six separate models described in the text. The outcome across all specifications is the willingness to participate in a clinical study led by the investigator, which is standardized to have a mean of 0 and standard deviation of 1. Panel (A) plots estimates for specifications that include the race concordant treatment and a *single* perceived investigator characteristic. Panel (B) plots estimates for a model with race concordant treatment and *all* perceived investigator characteristics. All specifications include respondent characteristics (age, squared age, sex, education, employment, and South/non-South region) and ML-rated image characteristics (predicted investigator emotion, age, image quality, image size). 95% CIs are plotted and based on robust standard errors. Panel (C) plots the R-squared value for each respective model. Dashed lines represent the estimated racial concordance effect (Panels A and B) and R-squared value (Panel C) for our main specification (Model 0, Table 1, Panel A, Column (4)).

We use the method proposed in Gelbach (2016) to further explore the mechanisms underlying the effect of racial concordance on interest in clinical trial participation. The method builds on the omitted variable bias formula and decomposes the change in the main coefficient between the base model (Model 0) and a version of the model including the variables that are candidates for explaining the effect (Model 6). The coefficient on racial concordance in Model 0 is 0.350, which is reduced to 0.214 in Model 6. Therefore, the difference that is to be explained is 0.136. Table 3 shows that trustworthiness explains 0.076 units of the change, attractiveness explains 0.071 units, and quality explains 0.025 units. Including the age and education variables actually increased the portion of the racial concordance effect that is not explained by the included characteristics. This confirms that trustworthiness and attractiveness seem to be main drivers of the racial concordance effect with trustworthiness explaining 21.7 percent and attractiveness 20.3 percent of its effect (calculated by dividing the explained parts by the main coefficients in Model 0).

Table 3: Decomposition of the Effect of NIH Investigator Racial Concordance on Clinical Trial Participation Interest

	Model 0 (1)	Model 6 (2)	Explained (3)
Racial Concordance	0.350	0.214	0.136
Trustworthiness	Not Included	Included	0.076
Attractiveness	Not Included	Included	0.071
Quality	Not Included	Included	0.025
Age	Not Included	Included	-0.007
Education	Not Included	Included	-0.029

Notes: OLS estimates of the willingness to participate in a future clinical trial led by race concordant NIH investigators. Outcomes are standardized to have a mean of 0 and standard deviation of 1. Column (1) includes an indicator for race concordant treatment. Column (2) includes all perceived investigator characteristics as defined in text. Column (3) shows the difference between the effect of racial concordance in Column (1) and Column (2) in the first row. The remaining rows show a decomposition of this difference into changes driven by the respective covariates added in Column (2). The decomposition follows the approach proposed in Gelbach (2016). All specifications include respondent characteristics (age, squared age, sex, education, employment, and South/non-South region) and ML-rated image characteristics (predicted investigator emotion, age, image quality, image size).

IV.B Robustness Tests

We conduct several robustness checks to confirm the results. First, we conduct a further check for the presence of experimenter demand effects. To do this, we re-estimate the main effects, excluding the individuals who mentioned race- or sex-related terms in the open-text question about the study’s purpose. The results, reported in Appendix Table A10, show similar effects. A second concern is that results are driven by outlier images — this is a legitimate concern given the few images of Black investigators we have based on the limited availability on the NIH website. To test this, we regressed interest in clinical trial participation on investigator image fixed effects. We cannot reject the null that the coefficients for all racially concordant investigators are the same (F-statistic = 0.95, p-value

= 0.486) and that the coefficients for all discordant investigators are the same (F-statistic = 1.44, p-value = 0.117). Lastly, we check whether our estimated effect is robust to the inclusion of other subject characteristics that may explain trial participation, such as risk aversion, trust in others, and altruism. Appendix Table A11 shows the results when we include these respondent preference measures as controls. We fail to reject the null that the racial concordance estimates are the same as in the original specification (p-value = 0.967).

IV.C Limitations and Applicability of our Findings

Randomized evaluations have many strengths but also well-known weaknesses — such as weak applicability to real-world situations. In our study, we presented participants with a picture of an NIH investigator and elicited their interest in participating in a clinical trial led by the investigator. In practice, featuring investigator images during trial recruitment is uncommon. However, it is increasingly common to highlight the contribution of Black investigators when aiming to engender trust, boost Black enrollment in trials and/or take-up of novel medical products (*e.g.*, Dr. Kizzmekia Corbett and the COVID-19 vaccine) (CBS News, 2021). When considering extrapolation to other settings, it is important that all of the images shown in our study were of actual NIH investigators and that subjects were informed of this fact. Results may differ if images were of investigators in a less-reputable institution or if the institution was not specified; previous research has shown lack of information increases opportunities for implicit bias to play a role in decision-making (Doleac and Hansen, 2020; Laouénan and Rathelot, 2022).

An additional potential concern with our findings is that we used a stated instead of a revealed preference measure. Enrolling subjects in actual medical experiments was not feasible as it would have been logistically challenging given the limited number of Black investigators and the specific enrollment criteria for each trial. Yet, given the importance of clinical trial participation to biomedical research productivity *and* the fact that it re-

quires informed consent, stated willingness to participate is often used as an outcome (e.g., Moorcraft et al. (2016) and Pariera et al. (2017)). We searched the term “willingness to participate in clinical trials” in PubMed — and found that it has been the subject of more than 4500 publications in PubMed since 1990, many of which use stated “willingness to participate” as an outcome.⁴ The number of publications with this key phrase has steadily trended upward over time (Appendix Figure B4).

Insights from behavioral economics research support the use of stated preferences as proxies for revealed preference measures. Levitt and List (2007) show there can be a strong correlation between lab and real-world decisions. To assess these correlations in our setting, we leverage data collected by a nonprofit, Research!America, which annually surveys the U.S. public on their views regarding science and technology. We take advantage of a specific wave (2021) when the survey asked both about willingness to participate in a clinical trial as well as whether the respondent (or their family member) had enrolled in a trial in the past. The results, shown in Appendix Figure B5, demonstrate that 43 percent of respondents who stated that they were “very likely” to participate in a future clinical trial actually had participated in the past, compared to only 4 percent of respondents who stated they “would not participate”.

To further validate our use of stated willingness to participate, we also analyzed alternative outcomes that are plausibly associated with future clinical trial participation. First, we asked respondents, “Would you like to learn more about research studies led by this National Institutes of Health Scientist?” directly after our intervention and primary outcome elicitation. In addition, at the end of the survey, we allowed respondents to act by clicking a link that directed to them more information on clinical trial enrollment. We also added a question on publicly financing research adopted from Research!America: “Would you be willing to pay \$1 per week more in taxes if you were certain that all of the money

⁴PubMed search results for “willingness to participate in clinical trials” can be accessed here. It’s also the case that information and logistical hurdles pose key challenges to participation though are not our focus herein (Unger et al., 2020).

would be spent on additional medical research?” We find that the effect of racial concordance on these three alternative measures is generally positive but small, and insignificant (Appendix Table A12).⁵ This is likely because the outcomes were dichotomous, thus had low power, and were only indirectly related to the investigator treatment.

V Discussion

In a randomized survey experiment using photographs of NIH investigators, Black respondents’ interest in clinical trial participation increased when presented with a photograph of a Black investigator. We find that perceived investigator trustworthiness followed by attractiveness are the most important factors predicting interest in clinical trial participation using regression analysis and the Gelbach (2016) decomposition. We do not observe any effect of sex concordance on interest in participation. These findings suggest that racial concordance among principal investigators and potential research participants can improve enrollment of underrepresented racial minorities in medical research by projecting trustworthiness.

Our results are consistent with other literature emphasizing trustworthiness as a significant determinant of participation in the healthcare system (Jacobs et al., 2006; Warren et al., 2020). The historical and contemporaneous mistreatment and neglect of Black Americans, including but not limited to the infamous Tuskegee syphilis experiment, have negatively affected Black Americans’ trust in the medical profession (Alsan and Wanamaker, 2018). A 2021 Pew Research Center survey found that the majority of Black adults think that research misconduct is just as likely to occur today as it was in the past (Funk, 2022). Further, in a July 2017 survey comparing attitudes about clinical trials among White, Black, Asian, and Hispanic U.S. adults, 50 percent of Black respondents cited lack of trust as a barrier to participating in clinical trials and 50 percent believed that subjects are sometimes

⁵The three outcomes are also significantly positively correlated with our primary outcome, stated willingness to participate, validating data quality and consistency of respondents’ preferences.

included in clinical trials without being told (Research!America, 2017). Increasing trust among Black communities requires the medical profession to become more trustworthy (Warren et al., 2020). Part of building trust is increasing the opportunity for members of underrepresented groups to be in positions of authority, including as principal investigators and physicians.

While better representation of Black and other historically marginalized groups may be an effective tool for encouraging participation in clinical studies, it is not likely to be a comprehensive solution given the structural barriers to accessing medical care and obtaining information on trial opportunities as well as the time and financial burden imposed on participants (National Academies of Sciences and Medicine, 2022). Further research is needed to determine the importance of investigator characteristics relative to these other factors and to assess potential complementarities.

Bibliography

- Alegria, Mikhaila, Shawn Sud, Benjamin E Steinberg, Nan Gai, and Asad Siddiqui.** 2021. "Reporting of Participant Race, Sex, and Socioeconomic Status in Randomized Clinical Trials in General Medical Journals, 2015 vs 2019." *JAMA Network Open*, 4(5): e2111516–e2111516.
- Alsan, Marcella, and Marianne Wanamaker.** 2018. "Tuskegee and the Health of Black Men." *The Quarterly Journal of Economics*, 133(1): 407–455.
- Alsan, Marcella, and Sarah Eichmeyer.** Forthcoming. "Experimental Evidence on the Effectiveness of Non-experts for Improving Vaccine Demand." *American Economic Journal: Economic Policy*.
- Alsan, Marcella, Fatima Cody Stanford, Abhijit Banerjee, Emily Breza, Arun G Chandrasekhar, Sarah Eichmeyer, Paul Goldsmith-Pinkham, Lucy Ogbu-Nwobodo, Benjamin A Olken, Carlos Torres, et al.** 2021. "Comparison of Knowledge and Information-seeking Behavior after General COVID-19 Public Health Messages and Messages Tailored for Black and Latinx Communities: a Randomized Controlled Trial." *Annals of Internal Medicine*, 174(4): 484–492.
- Alsan, Marcella, Maya Durvasula, Harsh Gupta, Joshua Schwartzstein, and Heidi L Williams.** Forthcoming. "Representation and Extrapolation: Evidence from Clinical Trials." *The Quarterly Journal of Economics*.
- Alsan, Marcella, Owen Garrick, and Grant Graziani.** 2019. "Does Diversity Matter for Health? Experimental Evidence from Oakland." *American Economic Review*, 109(12): 4071–4111.
- Anderson, Steven R, Morgan Gianola, Jenna M Perry, and Elizabeth A Reynolds Losin.** 2020. "Clinician–Patient Racial/Ethnic Concordance Influences Racial/Ethnic Minority Pain: Evidence from Simulated Clinical Interactions." *Pain Medicine*, 21(11): 3109–3125.

- Artiga, Samantha, Jennifer Kates, Josh Michaud, and L Hill.** 2022. "Racial Diversity Within COVID-19 Vaccine Clinical Trials: Key Questions and Answers." <https://www.kff.org/racial-equity-and-health-policy/issue-brief/racial-diversity-within-covid-19-vaccine-clinical-trials-key-questions-and-answers/>, Accessed: 7/9/2022.
- Awidi, Muhammad, and Samer Al Hadidi.** 2021. "Participation of Black Americans in Cancer Clinical Trials: Current Challenges and Proposed Solutions." *JCO Oncology Practice*, 17(5): 265–271.
- Banerjee, Abhijit, Marcella Alsan, Emily Breza, Arun G Chandrasekhar, Abhijit Chowdhury, Esther Duflo, Paul Goldsmith-Pinkham, and Benjamin A Olken.** 2020. "Messages on COVID-19 Prevention in India Increased Symptoms Reporting and Adherence to Preventive Behaviors Among 25 Million Recipients with Similar Effects on Non-recipient Members of Their Communities." *NBER Working Paper No. 27496*.
- Basu, Anirban, and Kritee Gujral.** 2020. "Evidence Generation, Decision Making, and Consequent Growth in Health Disparities." *Proceedings of the National Academy of Sciences*, 117(25): 14042–14051.
- Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman.** 2010. "What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment." *The Quarterly Journal of Economics*, 125(1): 263–306.
- Brandt, Allan M.** 1987. *No Magic Bullet: A Social History of Venereal Disease in the United States Since 1880*. USA:Oxford University Press.
- CBS News.** 2021. "Meet the Black Female Scientist at the Forefront of COVID-19 Vaccine Development." <https://www.cbsnews.com/news/covid-19-vaccine-development-kizzmekia-corbett/>, Accessed: 9/6/2023.

- Chandler, Jesse, Cheskie Rosenzweig, Aaron J Moss, Jonathan Robinson, and Leib Litman.** 2019. "Online Panels in Social Science Research: Expanding Sampling Methods Beyond Mechanical Turk." *Behavior Research Methods*, 51: 2022–2038.
- Chu, James, Sophia L Pink, and Robb Willer.** 2021. "Religious Identity Cues Increase Vaccination Intentions and Trust in Medical Experts among American Christians." *Proceedings of the National Academy of Sciences*, 118(49): e2106481118.
- Cooper, Lisa A, Debra L Roter, Rachel L Johnson, Daniel E Ford, Donald M Steinwachs, and Neil R Powe.** 2003. "Patient-centered Communication, Ratings of Care, and Concordance of Patient and Physician Race." *Annals of Internal Medicine*, 139(11): 907–915.
- Deming, David J, Noam Yuchtman, Amira Abulafi, Claudia Goldin, and Lawrence F Katz.** 2016. "The Value of Postsecondary Credentials in the Labor Market: An Experimental Study." *American Economic Review*, 106(3): 778–806.
- Doleac, Jennifer L, and Benjamin Hansen.** 2020. "The Unintended Consequences of "Ban the Box": Statistical Discrimination and Employment Outcomes when Criminal Histories are Hidden." *Journal of Labor Economics*, 38(2): 321–374.
- FDA.** 2014. "FDA Action Plan to Enhance the Collection and Availability of Demographic Subgroup Data." <https://www.fda.gov/media/89307/download>, Accessed: 7/9/2022.
- FDA.** 2022. "Diversity Plans to Improve Enrollment of Participants from Underrepresented Racial and Ethnic Populations in Clinical Trials; Draft Guidance for Industry." <https://www.regulations.gov/document/FDA-2021-D-0789-0002>, Accessed: 7/9/2022.
- Frakes, Michael D, and Jonathan Gruber.** 2022. "Racial Concordance and the Quality of Medical Care: Evidence from the Military." *NBER Working Paper No. 30767*.
- Funk, Cary.** 2022. "Black Americans' Trust in Medical Scientists and Views about the Potential for Researcher Misconduct." <https://www.pewresearch.org/science/2022/04/>

07/black-americans-trust-in-medical-scientists-and-views-about-the-potential-for-researcher-misconduct/, Accessed: 7/9/2022.

Gelbach, Jonah B. 2016. “When Do Covariates Matter? And Which Ones, and How Much?” *Journal of Labor Economics*, 34(2): 509–543.

Ginther, Donna K, Walter T Schaffer, Joshua Schnell, Beth Masimore, Faye Liu, Laurel L Haak, and Raynard Kington. 2011. “Race, Ethnicity, and NIH Research Awards.” *Science*, 333(6045): 1015–1019.

Green, Angela K, Niti Trivedi, Jennifer J Hsu, Nancy L Yu, Peter B Bach, and Susan Chimonas. 2022. “Despite The FDA’s Five-Year Plan, Black Patients Remain Inadequately Represented In Clinical Trials For Drugs: Study Examines FDA’s Five-year Action Plan aimed at Improving Diversity in and Transparency of Pivotal Clinical Trials for Newly-approved Drugs.” *Health Affairs*, 41(3): 368–374.

Greenwood, Brad N, Rachel R Hardeman, Laura Huang, and Aaron Sojourner. 2020. “Physician–patient Racial Concordance and Disparities in Birthing Mortality for Newborns.” *Proceedings of the National Academy of Sciences*, 117(35): 21194–21200.

Greenwood, Brad N, Seth Carnahan, and Laura Huang. 2018. “Patient–physician Gender Concordance and Increased Mortality among Female Heart Attack Patients.” *Proceedings of the National Academy of Sciences*, 115(34): 8569–8574.

Hill, Andrew, Daniel Jones, and Lindsey Woodworth. 2020. “Physician-patient Race-match Reduces Patient Mortality.” *Available at SSRN 3211276*.

Jacobs, Elizabeth A, Italia Rolle, Carol Estwing Ferrans, Eric E Whitaker, and Richard B Warnecke. 2006. “Understanding African Americans’ Views of the Trustworthiness of Physicians.” *Journal of General Internal Medicine*, 21: 642–647.

- Laouéan, Morgane, and Roland Rathelot.** 2022. "Can Information Reduce Ethnic Discrimination? Evidence from AirBnB." *American Economic Journal: Applied Economics*, 14(1): 107–132.
- Laurencin, Cato T, Camara P Jones, Lynne M Holden, Engineering National Academies of Sciences, Medicine, et al.** 2023. "Anti-Black Racism and the Mental Health of Black Men and Black Women in STEM." National Academies Press. Washington, DC.
- Levitt, Steven D., and John A. List.** 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World?" *Journal of Economic Perspectives*, 21(2): 153–174.
- Loeb, Stacy, Joseph E Ravenell, Scarlett Lin Gomez, Hala T Borno, Katherine Siu, Tatiana Sanchez Nolasco, Nataliya Byrne, Godfrey Wilson, Derek M Griffith, Rob Crocker, et al.** 2023. "The Effect of Racial Concordance on Patient Trust in Online Videos about Prostate Cancer: a Randomized Clinical Trial." *JAMA Network Open*, 6(7): e2324395–e2324395.
- Ludwig, Jens, and Sendhil Mullainathan.** 2023. "Machine Learning as a Tool for Hypothesis Generation." *NBER Working Paper No. 31017*.
- Manski, Charles F, John Mullahy, and Atheendar S Venkataramani.** 2023. "Using Measures of Race to Make Clinical Predictions: Decision Making, Patient Health, and Fairness." *Proceedings of the National Academy of Sciences*, 120(35): e2303370120.
- Moorcraft, Sing Yu, Cheryl Marriott, Clare Peckitt, David Cunningham, Ian Chau, Naureen Starling, David Watkins, and Sheela Rao.** 2016. "Patients' Willingness to Participate in Clinical Trials and Their Views on Aspects of Cancer Research: Results of a Prospective Patient Survey." *Trials*, 17(1): 1–12.

National Academies of Sciences, Engineering, and Medicine. 2022. "Recommendations for Improving Representation in Clinical Trials and Clinical Research." In *Improving Representation in Clinical Trials and Research: Building Equity for Women and Underrepresented Groups.* , ed. Kristen Bibbins-Domingo and Alex Helman, 133. Washington, DC:National Academies Press.

National Academies of Sciences, Engineering, and Medicine. 2023. "The Gatekeepers of STEMM: How Individual Bias and Inequality Persist and How STEM Professionals Can Help." In *Advancing Antiracism, Diversity, Equity, and Inclusion in STEMM Organizations: Beyond Broadening Participation.* , ed. Gilda A. Barabino, Susan T. Fiske, Layne A. Scherer and Emily A. Vargas, 165. Washington, DC:National Academies Press.

Nazha, Bassel, Manoj Mishra, Rebecca Pentz, and Taofeek K Owonikoko. 2019. "Enrollment of Racial Minorities in Clinical Trials: Old Problem Assumes New Urgency in the Age of Immunotherapy." *American Society of Clinical Oncology Educational Book*, 39: 3–10.

NIH. 2022a. "FAQs about Research Supplements to Promote Diversity in Health-related Research (Diversity Supplements)." <https://nida.nih.gov/about-nida/organization/offices/office-nida-director-od>, Accessed: 7/2/2022.

NIH. 2022b. "Number of NIH Principal Investigators Funded by Grant Mechanism and Race." <https://report.nih.gov/nihdatabook/report/306>, Accessed: 7/8/2022.

NIH. 2022c. "Principal Investigators." <https://irp.nih.gov/our-research/principal-investigators/name>, Accessed: 7/10/2022.

NIH. 2023. "Racial Disparities in NIH Funding." <https://diversity.nih.gov>, Accessed: 7/2/2022.

NIMHD. 2022. "National Institute of Minority Health and Health Disparities: History." <https://www.nimhd.nih.gov/about/overview/history/>, Accessed: 7/10/2022.

- Ocampo, Ricardo.** 2022. "Image-quality 1.2.7." <https://pypi.org/project/image-quality/>, Accessed: 6/30/2021.
- Pariera, Katrina L, Sheila T Murphy, Jingbo Meng, and Margaret L McLaughlin.** 2017. "Exploring Willingness to Participate in Clinical Trials by Ethnicity." *Journal of Racial and Ethnic Health Disparities*, 4: 763–769.
- Reardon, Sara.** 2023. "Alzheimer's Drug Trials Plagued by Lack of Racial Diversity." *Nature*, 620(7973): 256–257.
- Research!America.** 2017. "National Public Opinion Survey, Clinical Trials: Survey Data of Minority Populations." <https://www.researchamerica.org>, Accessed: 7/10/2022.
- Roberts, Dorothy.** 1997. *Killing the Black Body: Race, Reproduction, and The Meaning of Liberty*. New York:Vintage Books.
- 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.
- Serengil, Sefik Ilkin, and Alper Ozpinar.** 2021. "Hyperextended Lightface: A Facial Attribute Analysis Framework." In *2021 International Conference on Engineering and Emerging Technologies (ICEET)*. 1–4. Istanbul, Turkey.
- Takeshita, Junko, Shiyu Wang, Alison W Loren, Nandita Mitra, Justine Shults, Daniel B Shin, and Deirdre L Sawinski.** 2020. "Association of Racial/Ethnic and Gender Concordance between Patients and Physicians with Patient Experience Ratings." *JAMA Network Open*, 3(11): e2024583–e2024583.
- Unger, Joseph M, Dawn L Hershman, Cathee Till, Lori M Minasian, Raymond U Osarogiagbon, Mark E Fleury, and Riha Vaidya.** 2020. "'When Offered to Participate':

A Systematic Review and Meta-Analysis of Patient Agreement to Participate in Cancer Clinical Trials." *JNCI: Journal of the National Cancer Institute*, 113(3): 244–257.

Warren, Rueben C, Lachlan Forrow, David Augustin Hodge Sr, and Robert D Truog. 2020. "Trustworthiness Before Trust — Covid-19 Vaccine Trials and the Black Community." *New England Journal of Medicine*, 383(22): e121.

Washington, Harriet A. 2006. *Medical Apartheid: The Dark History of Medical Experimentation on Black Americans from Colonial Times to the Present*. New York:Doubleday.

Online Appendix for
Investigator Racial Diversity and Clinical Trial
Participation

by

Marcella Alsan, Romaine Campbell, Lukas Leister, and Ayotomiwa Ojo

Table of Contents

A	Appendix Tables	A.3
B	Appendix Figures	A.15
C	Survey Appendix	A.20

A Appendix Tables

Appendix Table A1: ML Ratings and Image Quality of NIH Investigator Photographs by Race

Variable	(1)		(2)		T-test
	Black Investigator N	Mean/SE	White Investigator N	Mean/SE	P-value (1)-(2)
Age	12	36.417 (1.811)	18	39.833 (1.837)	0.195
Angry	12	0.165 (0.383)	18	-0.242 (0.004)	0.291
Disgust	12	0.165 (0.383)	18	-0.236 (0.000)	0.299
Fear	12	0.119 (0.362)	18	-0.245 (0.022)	0.321
Happy	12	0.069 (0.304)	18	0.072 (0.219)	0.994
Sad	12	0.115 (0.360)	18	-0.244 (0.044)	0.327
Surprise	12	0.321 (0.449)	18	-0.264 (0.009)	0.199
Quality	12	0.206 (0.210)	18	-0.084 (0.306)	0.441
Size	12	-0.270 (0.326)	18	0.347 (0.198)	0.115

Notes: Machine learning generated ratings of NIH investigator pictures by race. Age, Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral are emotional ratings generated with the Python package Deepface (Serengil and Ozpinar, 2021). Quality is a rating of the image quality generated with the Python package Image-quality (Ocampo, 2022). Size is the file size in 5 bytes. All ratings are standardized to have a mean of 0 and standard deviation of 1, except for Age which is reported in years. Robust standard errors are used when comparing characteristics between the two groups.

Appendix Table A2: ML Ratings and Image Quality of NIH Investigator Photographs by Sex

Variable	(1)		(2)		T-test P-value (1)-(2)
	Female Investigator N	Mean/SE	Male Investigator N	Mean/SE	
Age	15	36.200 (1.654)	15	40.733 (1.977)	0.090
Angry	15	-0.239 (0.008)	15	0.080 (0.307)	0.307
Disgust	15	-0.224 (0.012)	15	0.073 (0.307)	0.341
Fear	15	-0.242 (0.025)	15	0.043 (0.290)	0.336
Happy	15	0.534 (0.116)	15	-0.392 (0.289)	0.006
Sad	15	-0.281 (0.012)	15	0.080 (0.290)	0.224
Surprise	15	0.119 (0.362)	15	-0.180 (0.065)	0.423
Quality	15	0.041 (0.186)	15	0.023 (0.364)	0.964
Size	15	-0.018 (0.268)	15	0.219 (0.251)	0.524

Notes: Machine learning generated ratings of NIH investigator pictures by sex. Age, Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral are emotional ratings generated with the Python package Deepface (Serengil and Ozpinar, 2021). Quality is a rating of the image quality generated with the Python package Image-quality (Ocampo, 2022). Size is the file size in 5 bytes. All ratings are standardized to have a mean of 0 and standard deviation of 1, except for Age which is reported in years. Standard errors are reported in parentheses. Robust standard errors are used when comparing ratings between the two groups.

Appendix Table A3: Summary Statistics and Representativeness of Survey Respondents

Variable	(1) ACS		(2) Survey		T-test P-value (1)-(2)
	N	Mean/SE	N	Mean/SE	
Male	126,963	0.483 (0.002)	323	0.467 (0.028)	0.579
Age 25 to 44	126,963	0.546 (0.002)	323	0.505 (0.028)	0.136
Region South	126,963	0.563 (0.002)	323	0.576 (0.028)	0.651
HS Degree or Less	126,963	0.401 (0.002)	323	0.402 (0.027)	0.971
Health Insurance	126,963	0.866 (0.001)	323	0.830 (0.021)	0.082
Employed	126,963	0.740 (0.002)	323	0.563 (0.028)	0.000
F-test of joint significance (F-stat)					6.559
F-test, number of observations					127,286

Notes: Survey refers to our survey sample. ACS refers to U.S.-born non-Hispanic Black Americans between 25-64 years of age in the 2019 American Community Survey, a nationally representative survey. We limited the ACS sample to respondents who are English-speaking and report having computer and internet access, as these were additional requirements for participating in our study. Robust standard errors are used when comparing ratings between the two samples.

Appendix Table A4: Balance of Respondent Characteristics by Race Concordant Treatment

Variable	(1)		(2)		T-test P-value (1)-(2)
	Race Discordant N	Mean/SE	Race Concordant N	Mean/SE	
Male	161	0.472 (0.039)	162	0.463 (0.039)	0.871
Age 25 to 39	161	0.360 (0.038)	162	0.407 (0.039)	0.385
Age 40 to 54	161	0.379 (0.038)	162	0.321 (0.037)	0.277
Age 55+	161	0.261 (0.035)	162	0.272 (0.035)	0.828
HS Degree or Less	161	0.385 (0.038)	162	0.420 (0.039)	0.527
Some College	161	0.416 (0.039)	162	0.432 (0.039)	0.773
Bachelor's Degree	161	0.199 (0.032)	162	0.148 (0.028)	0.231
Lower Income	161	0.484 (0.040)	162	0.414 (0.039)	0.201
Middle Income	161	0.360 (0.038)	162	0.438 (0.039)	0.153
Higher Income	161	0.155 (0.029)	162	0.148 (0.028)	0.859
Employed	161	0.516 (0.040)	162	0.611 (0.038)	0.084
Any Medical Condition	161	0.453 (0.039)	162	0.414 (0.039)	0.472
Region South	161	0.596 (0.039)	162	0.556 (0.039)	0.461
Democrat	161	0.590 (0.039)	162	0.623 (0.038)	0.540
Republican	161	0.031 (0.014)	162	0.049 (0.017)	0.404

Notes: Balance table of respondent characteristics by race concordant NIH investigator treatment. Income brackets are defined with the following cutoffs: household income below \$30k (Lower Income), between \$30k and \$75k (Middle Income), and above \$75k (Higher Income). Any Medical Condition is an indicator for whether the respondent reported having a medical condition that requires regular care. Robust standard errors are used when comparing characteristics between the two groups.

Appendix Table A5: Balance of Respondent Characteristics by Sex Concordant Treatment

Variable	(1)		(2)		T-test P-value (1)-(2)
	Sex Discordant N	Mean/SE	Sex Concordant N	Mean/SE	
Male	164	0.457 (0.039)	159	0.478 (0.040)	0.711
Age 25 to 39	164	0.390 (0.038)	159	0.377 (0.039)	0.813
Age 40 to 54	164	0.384 (0.038)	159	0.314 (0.037)	0.190
Age 55+	164	0.226 (0.033)	159	0.308 (0.037)	0.094
HS Degree or Less	164	0.378 (0.038)	159	0.428 (0.039)	0.365
Some College	164	0.451 (0.039)	159	0.396 (0.039)	0.319
Bachelor's Degree	164	0.171 (0.029)	159	0.176 (0.030)	0.899
Lower Income	164	0.482 (0.039)	159	0.415 (0.039)	0.230
Middle Income	164	0.366 (0.038)	159	0.434 (0.039)	0.213
Higher Income	164	0.152 (0.028)	159	0.151 (0.028)	0.970
Employed	164	0.591 (0.039)	159	0.535 (0.040)	0.304
Any Medical Condition	164	0.378 (0.038)	159	0.491 (0.040)	0.041
Region South	164	0.579 (0.039)	159	0.572 (0.039)	0.900
Democrat	164	0.628 (0.038)	159	0.585 (0.039)	0.429
Republican	164	0.037 (0.015)	159	0.044 (0.016)	0.735

Notes: Balance table of respondent characteristics across sex concordant NIH investigator treatment. Income brackets are defined with the following cutoffs: household income below \$30k (Lower Income), between \$30k and \$75k (Middle Income), and above \$75k (Higher Income). Any Medical Condition is an indicator for whether the respondent reported having a medical condition that requires regular care. Robust standard errors are used when comparing characteristics between the two groups.

Appendix Table A6: Attrition Post-Randomization by Treatment

	Attrition	
	(1)	(2)
Racial Concordance	-0.012 (0.009)	
Sex Concordance		0.000 (0.009)
Control Mean	0.012	0.006
Observations	325	325

Notes: OLS estimates of attrition post-randomization by race concordant (Column 1) or sex concordant (Column 2) NIH investigator treatment. Robust standard errors are in parentheses.

Appendix Table A7: Correlation of Perceived Investigator Characteristics

	Trustworthiness (1)	Attractiveness (2)	Quality (3)	Education (4)	Age (5)
Trustworthiness	1.000				
Attractiveness	0.402	1.000			
Quality	0.499	0.229	1.000		
Education	0.214	0.099	0.343	1.000	
Age	0.037	0.048	-0.033	0.010	1.000

Notes: Correlation matrix of respondent ratings of investigator characteristics.

Appendix Table A8: The Effect of NIH Investigator Racial Concordance on Clinical Trial Participation Interest: Adding Perceived Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Racial Concordance	0.248 (0.122)	0.239 (0.129)	0.257 (0.134)	0.335 (0.132)	0.383 (0.132)	0.213 (0.127)
Trustworthiness	0.331 (0.053)					0.249 (0.061)
Attractiveness		0.273 (0.061)				0.164 (0.062)
Quality			0.188 (0.062)			0.047 (0.066)
Education				0.117 (0.117)		-0.031 (0.113)
Age					0.089 (0.053)	0.071 (0.045)
Control Mean	-0.147	-0.147	-0.147	-0.147	-0.147	-0.147
R-squared	0.218	0.180	0.148	0.119	0.123	0.244
Observations	323	323	323	323	323	323
Respondent Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Image Covariates	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimates of the willingness to participate in a future clinical trial led by race concordant NIH investigators. Outcomes are standardized to have a mean of 0 and standard deviation of 1. All columns include an indicator for treatment. Column (1) includes standardized perceived trustworthiness. Column (2) includes standardized perceived attractiveness. Column (3) includes standardized perceived quality. Column (4) includes a binary for perceived graduate education attainment. Column (5) includes standardized age group. Column (6) includes all of the covariates from Columns (1)-(5). All specifications include respondent characteristics (age, squared age, sex, education, employment, and South/non-South region) and ML-rated image characteristics (predicted investigator emotion, age, image quality, image size). Robust standard errors are in parentheses.

Appendix Table A9: The Effect of NIH Investigator Sex Concordance on Clinical Trial Participation Interest: Adding Perceived Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Sex Concordance	0.015 (0.102)	0.025 (0.104)	-0.001 (0.106)	0.001 (0.108)	-0.020 (0.109)	0.022 (0.102)
Trustworthiness	0.344 (0.053)					0.248 (0.061)
Attractiveness		0.291 (0.059)				0.178 (0.061)
Quality			0.210 (0.060)			0.061 (0.066)
Education				0.150 (0.118)		-0.019 (0.113)
Age					0.065 (0.052)	0.057 (0.044)
Control Mean	0.006	0.006	0.006	0.006	0.006	0.006
R-squared	0.208	0.171	0.139	0.103	0.102	0.238
Observations	323	323	323	323	323	323
Respondent Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Image Covariates	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS estimates of the willingness to participate in a future clinical trial led by sex concordant NIH investigators. Outcomes are standardized to have a mean of 0 and standard deviation of 1. All columns include an indicator for treatment. Column (1) includes perceived trustworthiness. Column (2) includes standardized perceived attractiveness. Column (3) includes standardized perceived quality. Column (4) includes a binary for perceived graduate education attainment. Column (5) includes standardized age. Column (6) includes all of the covariates from Columns (1)-(5). All specifications include respondent characteristics (age, squared age, sex, education, employment, and South/non-South region) and ML-rated image characteristics (predicted investigator emotion, age, image quality, image size). Robust standard errors are in parentheses.

Appendix Table A10: Robustness Check: Dropping Respondents with Race/Sex Mentions

	(1)	(2)
Racial Concordance	0.350 (0.132)	0.244 (0.132)
Control Mean	-0.147	-0.128
Observations	323	302
Respondent Covariates	Yes	Yes
Image Covariates	Yes	Yes

Notes: OLS estimates of the willingness to participate in a future clinical trial led by race concordant NIH investigators. Outcomes are standardized to have a mean of 0 and standard deviation of 1. Column (1) is our main specification. In Column (2), we drop all respondents who mention race or sex when asked “What do you think this study is about?” Specifically, we dropped responses that include “race”, “racial”, “black”, “white”, “gender”, “minority”, “ethnicity” and “African”. All specifications include respondent characteristics (age, squared age, sex, education, employment, and South/non-South region) and ML-rated image characteristics (predicted investigator emotion, age, image quality, image size). Robust standard errors are in parentheses.

Appendix Table A11: Robustness Check: Racial Concordance Effect and Respondent Characteristics

	(1)	(2)
Racial Concordance	0.350 (0.132)	0.353 (0.114)
Willingness to Trust		-0.029 (0.092)
Altruism		0.396 (0.065)
Risk Aversion		0.180 (0.070)
Control Mean	-0.147	-0.147
Observations	323	323
Respondent Covariates	Yes	Yes
Image Covariates	Yes	Yes

Notes: OLS estimates of the willingness to participate in a future clinical trial led by race concordant NIH investigators. Outcomes are standardized to have a mean of 0 and standard deviation of 1. The covariates of Altruism and Risk Aversion are also standardized to have a mean of 0 and standard deviation of 1. Willingness to Trust is an indicator for whether the respondent thinks most people can be trusted. All specifications include respondent characteristics (age, squared age, sex, education, employment, and South/non-South region) and ML-rated image characteristics (predicted investigator emotion, age, image quality, image size). Robust standard errors are in parentheses.

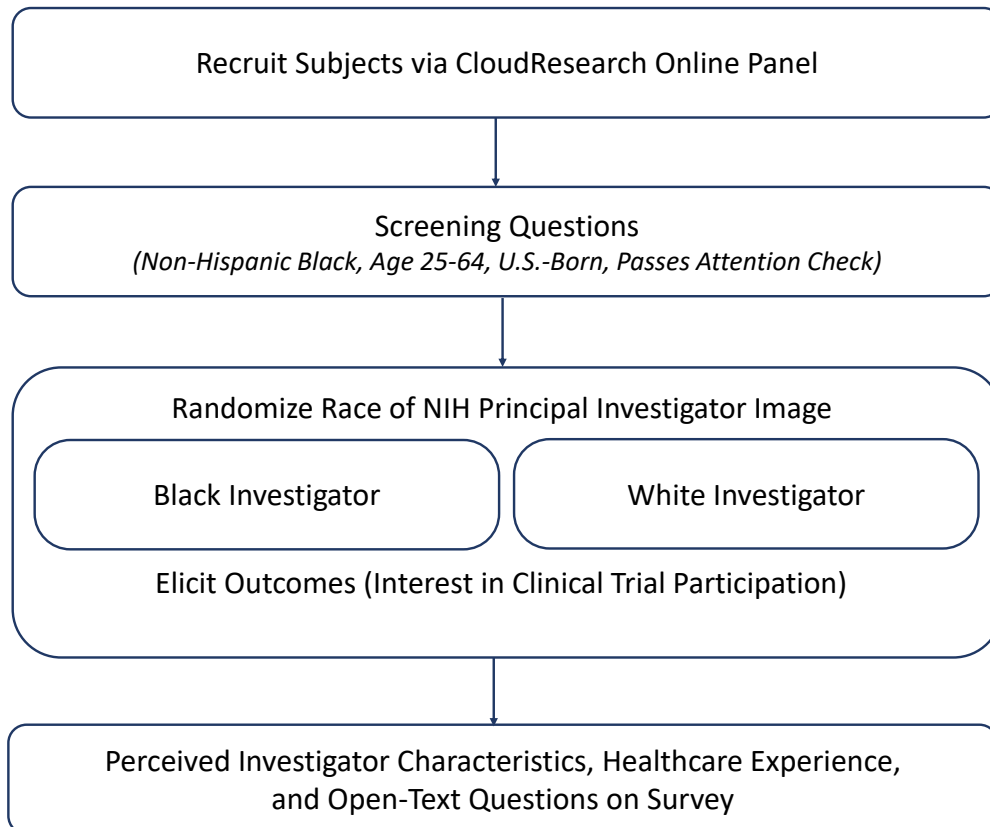
Appendix Table A12: The Effect of NIH Investigator Racial
Concordance on Alternative Measures of Clinical Trial
Participation

	Request Information		WTP for Research		Click Link	
	(1)	(2)	(3)	(4)	(5)	(6)
Racial Concordance	0.020 (0.050)	0.010 (0.063)	0.047 (0.055)	-0.009 (0.067)	0.006 (0.033)	0.006 (0.033)
Control Mean	0.708	0.708	0.422	0.422	0.068	0.068
Observations	323	323	323	323	323	323
Respondent Covariates	No	Yes	No	Yes	No	Yes
Image Covariates	No	Yes	No	Yes	No	Yes

Notes: OLS estimates of the effect of race concordant treatment on alternative measures of clinical trial participation. In Columns (1)-(2), the outcome is an indicator for whether the respondent answered yes to the question “Would you like to learn more about research studies led by this National Institutes of Health Scientist?” In Columns (3)-(4), the outcome is an indicator for whether the respondent answered yes to the question “Would you be willing to pay \$1 per week more in taxes if you were certain that all of the money would be spent on additional medical research?” In Columns (5)-(6), the outcome is an indicator for whether the respondent clicked on a link with more information on how to enroll in clinical trials. The Column (2), Column (4), and Column (6) specifications include respondent characteristics (age, squared age, sex, education, employment, and South/non-South region) and ML-rated image characteristics (predicted investigator emotion, age, image quality, image size). Robust standard errors are in parentheses.

B Appendix Figures

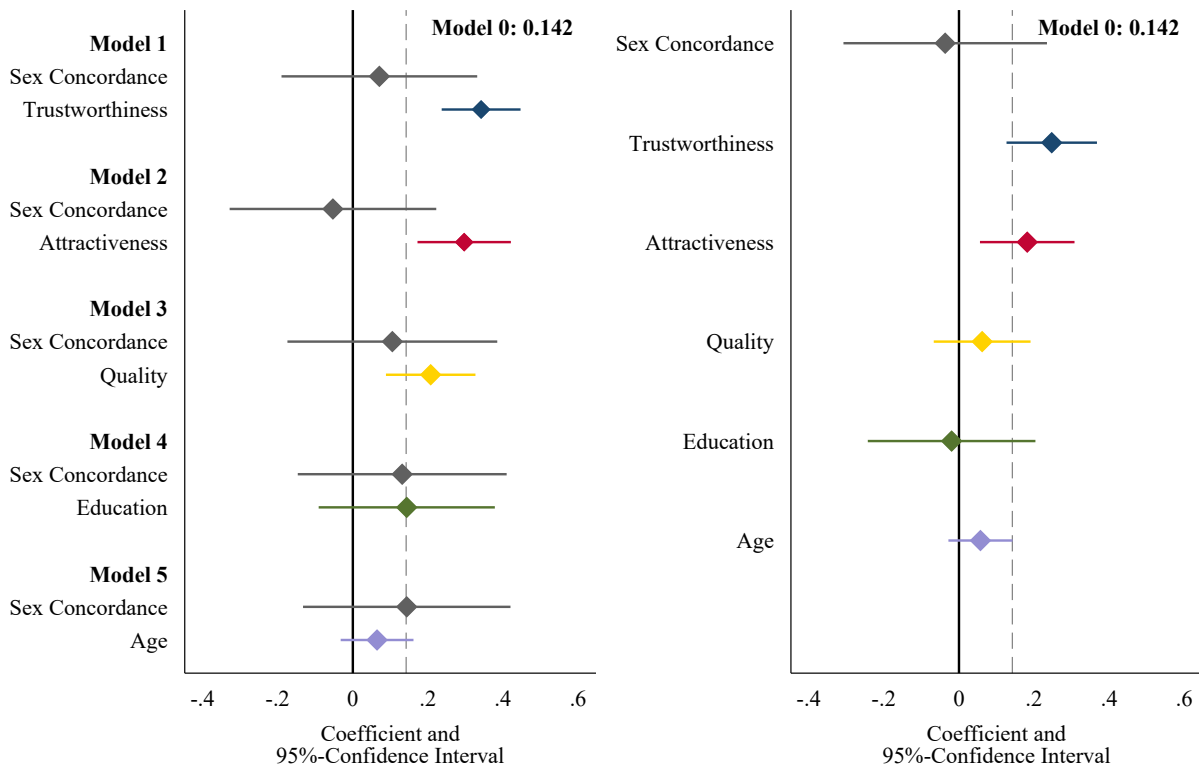
Appendix Figure B1: Study Flow



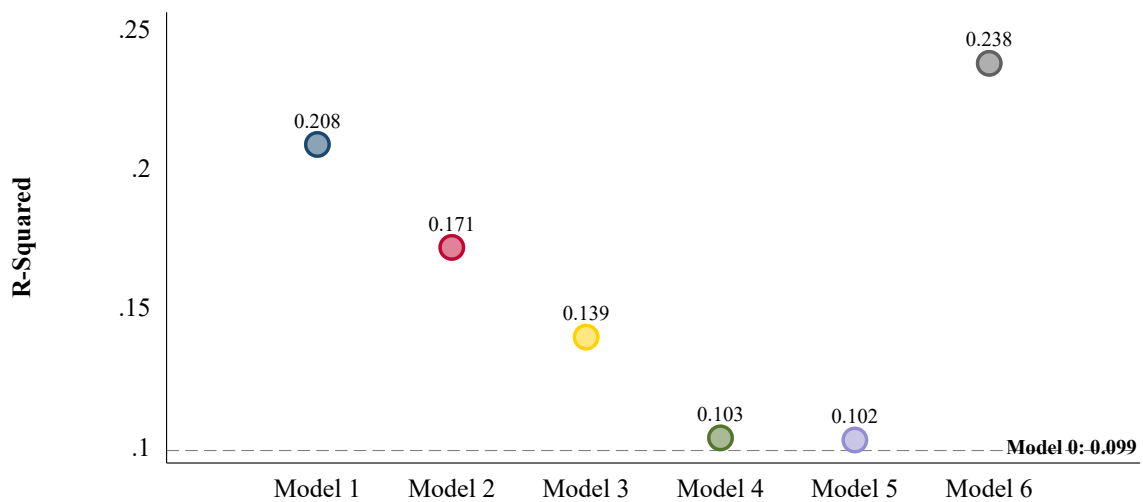
Notes: : Overview of the study.

Appendix Figure B3: The Effect of NIH Investigator Sex Concordance and Perceived Characteristics on Clinical Trial Participation Interest

(a) Coefficient Plot

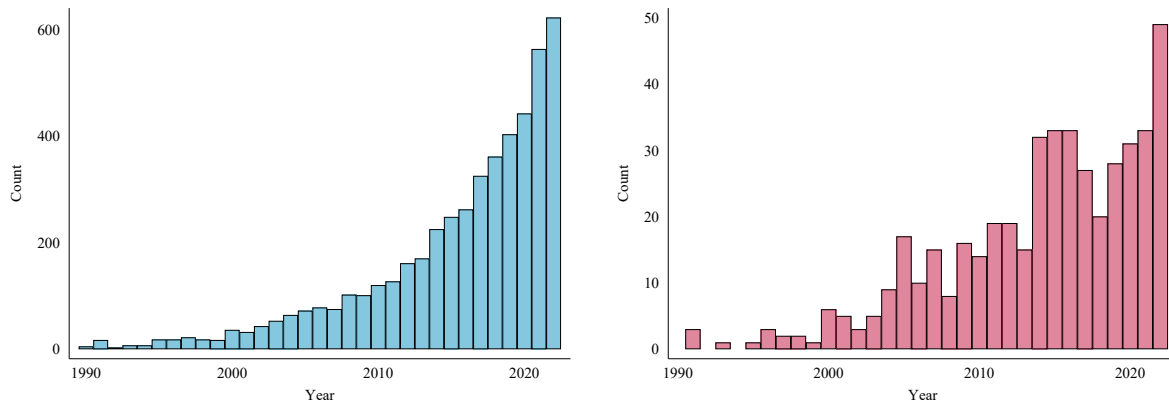


(b) Model Fit Plot



Notes: Figure plots OLS estimates from six separate models described in the text. The outcome across all specifications is the willingness to participate in a clinical study led by the investigator and is standardized to have a mean of 0 and standard deviation of 1. Panel (A) plots estimates for specifications that include the sex concordant treatment and a *single* perceived investigator characteristic. Panel (B) plots estimates for a model with sex concordant treatment and *all* perceived investigator characteristics. All specifications include respondent characteristics (age, squared age, sex, education, employment, and South/non-South region) and ML-rated image characteristics (predicted investigator emotion, age, image quality, image size). 95% CIs are plotted and based on robust standard errors. Panel (C) plots the R-squared value for each respective model. Dashed lines represent the estimated sex concordance effect (Panels A and B) and R-squared value (Panel C) for our main univariate specification (Model 0).

Appendix Figure B4: PubMed Publications on Willingness to Participate in Clinical Trials

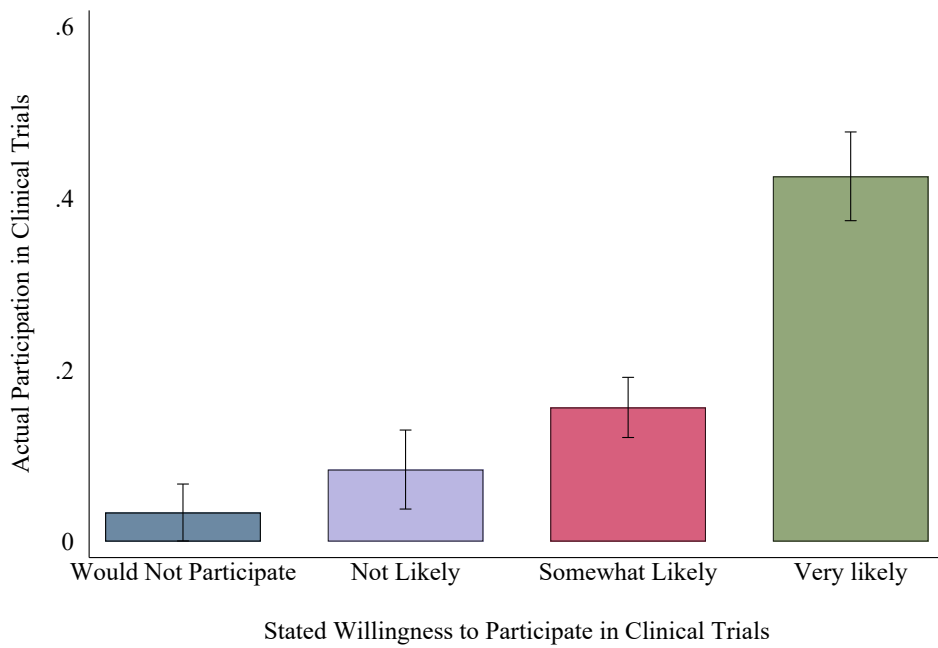


(a) Total Publications

(b) Race/Ethnicity Publications

Notes: Figure plots yearly the number of peer-reviewed publications containing the phrase "willingness to participate in clinical trials [or research]" on the PubMed search engine from 1990-2022. Panel (A) plots the total number of hits. Panel (B) restricts to publications that mention "race/racial", "ethnicity/ethnic" or "Black".

Appendix Figure B5: Stated Willingness to Participate versus Actual Participation in Clinical Trials



Notes: Figure plots actual clinical trial participation by respondents, stated willingness to participate in a clinical trial. Actual willingness is an indicator for whether the respondent answered “yes” to the question “Have you or anyone in your family ever participated in clinical trials?” Stated Willingness is measured by the question “If your doctor found a clinical trial for you and recommended you join, how likely would you be to participate in a clinical trial?” measured on a four-point Likert scale. 95% CIs are plotted. Data is drawn from the 2021 Research!America poll on a representative sample of U.S. adults (N=1209).

C Survey Appendix

Link to Survey: https://harvard.az1.qualtrics.com/jfe/form/SV_dnAytd4GphkUzc

Screeners Include:

- Race (Non-Hispanic Black)
- Age (25-24)
- Born in the United States
- Passes attention check

Appendix Table C1: Outcome Measures Question Wording

Outcome Name	Question Text	Response Options	Coded Options
Primary Outcome			
Willingness to Participate	If the person in this photograph were leading a medical research study to prevent a disease that affected people in your community, how interested would you be in participating in this study, if at all?	On a scale of 0: Not at all likely to 10: Extremely likely	Standardized to have a mean of 0 and standard deviation of 1
Secondary Outcomes			
Education	What is the highest degree or level of schooling that you think the person completed?	1: Less than a high school diploma, 2: High school diploma or equivalent, 3: Some college but no degree, 4: Associate's degree, 5: Bachelor's degree, 6: Master's degree, 7: Professional or doctoral degree	Created binary variable equal to 0 if 6 is selected and equal to 1 if other response is selected
Quality	How qualified do you think this person is compared to other scientists who work at the National Institutes of Health?	1: Much less qualified than other scientists, 2: Somewhat less qualified than other scientists, 3: Neither more or less qualified than other scientists, 4: Somewhat more qualified than other scientists, 5: Much more qualified than other scientists	Standardized to have a mean of 0 and standard deviation of 1
Trustworthiness	How trustworthy do you think this person looks?	1: Not at all trustworthy, 2: Somewhat untrustworthy, 3: Neither trustworthy or untrustworthy, 4: Somewhat trustworthy, 5: Extremely trustworthy	Standardized to have a mean of 0 and standard deviation of 1
Age	How old do you think this person looks?	1: 25-34 years old, 2: 35-44 years old, 3: 45-54 years old, 4: 55-64 years old, 5: 65 years or more	Standardized to have a mean of 0 and standard deviation of 1
		1: Not at all attractive, 2: Somewhat unattractive, 3: Neither	Standardized to have