



Patient traits shape health-care stakeholders' choices on how to best allocate life-saving care

Charles Crabtree ^{1,4}, John B. Holbein ^{2,4} and J. Quin Monson ^{3,4}

During global pandemics, health-care decision makers often face critical shortages of life-saving medical equipment. How do medical stakeholders prioritize which patients are most deserving of scarce treatment? We report the results of three joint experiments conducted in the United States in 2020, testing for biases in US physicians', citizens' and elected politicians' preferences for scarce ventilator distribution. We found that all stakeholders prioritized younger patients and patients who had a higher probability of surviving with ventilator access. When patients' survivability was tied, physicians prioritized patients from racial/ethnic minorities (that is, Asian, Black and Hispanic patients) over all-else-equal white patients, religious minorities (that is, Muslim patients) over religious majority group members (that is, Catholic patients) and patients of lower socio-economic status over wealthier patients. The public also prioritized Black and Hispanic patients over white patients but were biased against religious minorities (that is, Atheist and Muslim patients) relative to religious majority group members. Elected politicians were also biased against Atheist patients. Our effects varied by political party—with Republican physicians, politicians and members of the public showing bias against religious minority patients and Democratic physicians showing preferential treatment of racial and religious minorities. Our results suggest that health-care stakeholders' personal biases impact decisions on who deserves life-saving medical equipment.

When people are forced to make choices that have life-and-death consequences for others, to what extent do the deciders' individual biases come into play? This question was thrust to the forefront of public discussion during the COVID-19 outbreak that infected hundreds of millions of individuals around the world. During the outbreak, many nations faced critical shortages of life-saving medical equipment. One example of such a shortage was the stock of ventilators—medical devices that move breathable air into and out of patients' lungs. These devices are vitally important, as some patients infected with SARS-CoV-2 struggled to breathe on their own^{1–4}. For this reason, ventilators have been referred to as “the device[s] that become the decider between life and death” for COVID-19 patients⁵. (There were other treatments—such as oxygen therapy/extracorporeal membrane oxygenation—but many COVID-19 patients still received treatment from ventilators⁶.) However, in many locations around the world, the number of ventilators needed far surpassed the number of ventilators available^{7,8}. Although these shortages were especially acute in developing contexts^{9–13}, developed nations experienced shortfalls as well^{8,14}. At the beginning of the pandemic, for example, some estimates pegged the number of ventilators needed in the United States alone at several hundred thousand to as many as a million¹⁵; only somewhere between 60,000 and 160,000 ventilators were actually available in the country at that time. (Estimates of the number of ventilators available at a given period depend on whether one includes partially functioning machines or not¹⁶.) This led medical experts to issue direct and dire warnings that “there [were] not enough ventilators for patients with COVID-19”¹⁷ and that there was “no way to solve the problem before the disease reach[ed] full throttle.”¹⁷ Some areas faced fewer shortages than others¹⁸ (largely because many developed nations resorted to extreme tactics to build more ventilators or to improvisational means of finding substitutes

for them), but the threat of shortages of ventilators remained ever present^{8,18,19}. Under these circumstances, many were left to ask what we would do “if two COVID-19 victims need[ed] ventilators and just one [was] available.”²⁰

Under these emergent circumstances, many medical decision makers had to make truly excruciating choices about which patients were most deserving of scarce medical resources. Many of these choices involved decision makers having to choose between patients with similar likelihoods of survival. As a result, some physicians have called choices over the allocation of ventilators “the toughest [form of] triage”²¹—a truly “heart-wrenching choice”²¹ for any decision maker. Although some guidelines have been developed to govern how ventilators are allocated during the pandemic^{1,4,22–31}, many of them come to different conclusions about who should ethically have preference, and even when the guidelines do agree, all of them leave much room open for individual discretion by the key stakeholders in the decision-making process. Previous research in other social contexts has shown that administrative discretion provides decision makers with opportunities to rely on heuristics or shortcuts that advantage some groups over others^{32–39}. Given this, we can reasonably theorize that when several patients have the same chances of surviving, decision makers may rely on their individual biases—whether deeply embedded, unconscious or overt—in making their decisions about how to allocate scarce medical resources.

Some prior work has explored the extent to which discrimination occurs in health-care settings. This work ranges from studies that explore whether physicians perceive that certain types of patients have higher pain thresholds^{40–42}, to studies that explore the extent to which various groups are given preferential access to health resources^{43–57}, to studies that explore how much time physicians give to and how they communicate with their patients^{58–60}. A vast majority of these studies focus on the extent and nature of

¹Department of Government, Dartmouth College, Hanover, NH, USA. ²Frank Batten School of Leadership and Public Policy, University of Virginia, Charlottesville, VA, USA. ³Department of Political Science, Brigham Young University, Provo, UT, USA. ⁴These authors contributed equally: Charles Crabtree, John B. Holbein, J. Quin Monson. ✉e-mail: holbein@virginia.edu

racial and ethnic biases^{34,40–54}. The next most common are studies that examine socio-economic^{37,55,61–63}, age^{64–68} and religious^{36,69–72} biases in the health-care system.

Though a large body of research has explored the role of individual biases in decision-making processes, this literature has faced significant hurdles. First, though previous work has documented large observational disparities in access to health care and in the health outcomes that result downstream^{45–52}, it is difficult to attribute these patterns to discrimination alone. Given that many difficult-to-observe characteristics could be driving unequal outcomes (for example, economic inequality, cultural differences and institutional design), observational differences in individual health outcomes are not sufficient to definitively demonstrate discrimination. This methodological weakness is not avoided by using implicit association tests that look for subtle differences in click patterns in structured artificial tasks^{73–77}. Unfortunately, this research design too leaves open the possibility of endogeneity—wherein relationships attributable to discrimination are really driven by other unobserved factors. Furthermore, implicit association test measures themselves face distinct challenges in capturing individuals' tendencies to discriminate. For example, scholars have recently argued that implicit association tests often lack construct, external, and predictive validity^{78–86}.

Moreover, much of the work that has tried to account for endogeneity has looked solely in low-stakes environments. Many of these studies have used the so-called correspondence audit design to see whether, for example, decision makers discriminate in providing differential access to appointments for outpatient care^{32–39}. Much of the research that social scientists do on discrimination in life-and-death decisions has been limited by the fundamental biases that shape the data-generating process. For example, studies of discrimination patterns in policing often face issues with post-treatment/collider biases that arise from not observing the population of individuals that could have been exposed to treatment^{87,88}. This same problem—and others such as omitted variable bias—may also be present in the health domain. (We return to these methodological challenges in the Discussion and Methods.) This gap in the literature is unfortunate for many reasons but particularly given the many contexts for intense resource scarcity in the medical field. Because we know little about discrimination in many high-stakes situations, we can only theorize that it might be lower—because people might be more intentional about their choices and less likely to rely on subconscious biases—or higher—because people might be more likely to rely on crude heuristics under periods of threat. These emergent conditions and large inequities demand rigorous evidence that teases apart just how much of a role individual biases play in promoting health inequalities. Our work helps fill this important gap in the scientific literature.

Given the large hurdles that previous research has been unable to clear, there is a great deal of merit to using experimental designs to study discrimination in the health-care domain. Unfortunately, not many experiments have been run to study the extent and nature of discrimination in health care^{89,90}. As Currie et al. succinctly put it, “[experimental tests for discrimination] are still unusual in health care and existing studies often have very small samples”⁹¹.

In short, what we lack in the study of discrimination in the health-care domain is a study with an experimental design on a large sample of participants at various stages of the health-care chain that measures discrimination in a higher-stakes environment. In this paper, we push towards achieving these objectives. We use conjoint experiments to explore the extent and nature of individual racial, religious, age and socio-economic biases among American health-care decision makers and stakeholders—that is, the physicians who treat patients, the members of the public who receive these services and the public officials who play a large role in allocating scarce health-care resources—in the allocation of the limited stock of medical ventilators. We also test whether patients who would be

more likely to survive with a ventilator are more or less likely to receive access to that life-saving treatment. We chose to randomize this set of attributes because they are often considered to be important in making ventilator allocation decisions (for example, age and the probability of survival^{1,4,22–27}), have been shown to have biases in other social contexts (for example, race and religion^{33,34,36,92–94}), have well-documented inequities in the health domain^{45–52} and/or are the types of information that those in our sample know, or could find out, about potential medical patients.

Results

We ran three conjoint experiments to test for individual biases in decisions made by physicians, policymakers and the public. Conjoint experiments are commonly used in the social sciences to help understand how people value different attributes of possible choices^{95–103}. In our conjoint experiments, physicians ($N=1,523$), politicians ($N=839$) and members of the public ($N=2,080$) viewed fictional patient case files with the following characteristics randomized: patient survivability, age, socio-economic status (SES), race/ethnicity and religion. We kept our list of baseline variables small to ensure that our survey was not an undue burden on the respondents. The other questions focused on the demographic and political backgrounds of the participants in our study. They included fear about catching the coronavirus, fear about the people they care about catching the coronavirus, how closely they follow the news, approval of President Trump overall and in regard to the response to the pandemic, political ideology, age, race/ethnicity and political party. Doctors were also asked how many years they had practised medicine, whether they had treated patients who have had the COVID-19 virus, whether they prescribe life-saving COVID-related interventions, whether they had difficulty obtaining personal protective equipment, and about their own mental health state and their fear levels that their patients will contract the coronavirus. (For the full list of baseline characteristics we measured and for more details on our experimental design and the methods employed in our study, see the Methods.)

Do stakeholders give preference to patients with higher odds of survival? We started by testing whether our samples were sensitive to the patients' estimated probability of survival (if they were given a breathing assistance treatment) in assigning the use of scarce ventilators. This allowed us to explicitly test whether physicians, politicians and the public were responding to the core characteristic undergirding much, if not all, medical triage decisions—the idea that patients that are more likely to benefit from a scarce service should be given priority. After all, most medical groups agree that this should be an important deciding factor in the provision of ventilators. If our respondents were taking the task seriously and responding in a way that many key actors have deemed appropriate, we would expect that chances of survival would have a substantial effect on ventilator allocation decisions, given that virtually every decision-making criterion for allocating scarce medical resources focuses first on the probability of a patient surviving. (In the Discussion, we briefly return to a broader consideration of whether preferential treatment for patients deemed to have a higher probability of survival is actually universally desirable.)

Figure 1 shows the average effect of a patient having a high probability of survival (if they were given treatment from a ventilator) on the chances that the participants in our three samples chose to give the ventilator to that person. In the figure, the circles represent the coefficients, while the thick and thin bars show the 90% and 95% confidence intervals. If doctors, elected politicians and the public were responding to the chances that a patient would survive with the ventilator treatment, we would expect the coefficients to be positive and distinct from zero (that is, to the right of the grey dashed reference line).

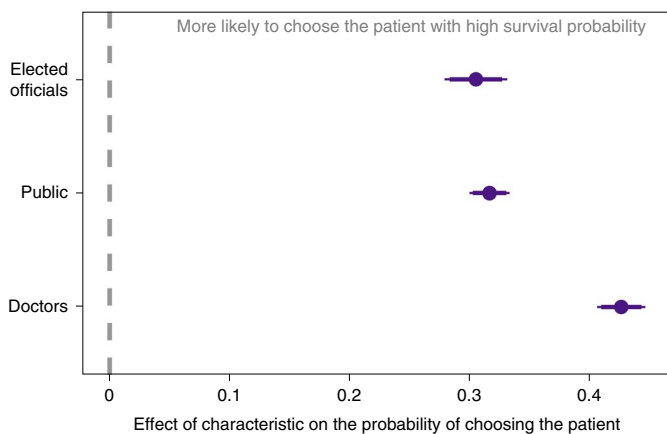


Fig. 1 | Patterns of physicians', politicians' and the public's allocation of ventilators based on patients' probability of survival (pooled individual samples). Coefficient plot for the effect of patient probability of survival (high (70%) versus low (10%)) if they have access to a ventilator on the probability of choosing that patient type. The circles represent the average marginal component effects, while the thick and thin bars show the 90% and 95% confidence intervals. Coefficients on the right side of the dashed line indicate a higher likelihood of choosing a patient with the given characteristic, all else being equal. As in all conjoint experiments, the unit of analysis is the respondent-comparison pair-individual profile. Pooled $N_{\text{docs}} = 12,596$; pooled $N_{\text{public}} = 18,962$; pooled $N_{\text{pols}} = 7,442$.

In fact, this is just what we saw. Figure 1 shows that physicians who prescribe treatments to patients, politicians who make key health policy decisions and the public that receives these treatments (and who often influence the decisions of doctors and elected politicians) were—as we would expect if these groups are adhering to common medical guidelines—highly sensitive to a patient's probability of survival in assigning life-saving medical equipment. Patients with a high probability of survival (that is, 70%) were much more likely to be prescribed a ventilator than patients with a low probability of survival (that is, 10%). All else being equal, doctors were 42.6 percentage points more likely to choose the high-probability-of-survival patient for the treatment ($\beta_{\text{d.f.}} = 1,522$; $P < 0.001$; $\beta = 42.6$ percentage points; 95% confidence interval, (40.6, 44.6)). Though less sensitive to this treatment condition, politicians ($\beta_{\text{d.f.}} = 838$; $P < 0.001$; $\beta = 30.6$ percentage points; 95% confidence interval, (28.0, 33.2)) and the public ($\beta_{\text{d.f.}} = 2,079$; $P < 0.001$; $\beta = 31.7$ percentage points; 95% confidence interval, (30.0, 33.4)) were still quite sensitive to a patient's probability of survival. (For the estimates used in Fig. 1, see Supplementary Table 4; for formal statistical tests of the difference in response to survivability across our three samples, see Supplementary Table 19; for the full set of P values adjusted for multiple comparisons that we employ throughout our results—that is, the Westfall–Young 'stepdown' adjusted P values¹⁰⁴—see Supplementary Table 15.) This difference in the prescription patterns of physicians versus politicians and the public was consistent with physicians being more attuned to the criteria frequently emphasized in medical training, as we might expect given their extensive training that reinforces this core principle of medical triage. However, these results also suggested that all three samples are making decisions as we would expect if they were taking this hypothetical prompt seriously.

Do stakeholders respond to other patient characteristics? We next turn to whether physicians, politicians and the public prescribe life-saving medical treatment on the basis of other patient characteristics outside of the estimated probability of survival. If the three

groups in our sample were making decisions solely on the basis of a patient's chances of survival, we would expect that all other characteristics manipulated in our conjoint experiment—age, SES, religion and race/ethnicity—would have no effect on the odds that a patient was chosen. If the respondents were treating young and old, rich and poor, religious majorities (that is, Protestants and Catholics) and religious minorities (that is, Muslims and Atheists), and racial/ethnic majorities (that is, non-Hispanic white people) and racial/ethnic minorities (that is, Black, Hispanic or Latino, and Asian people) the same—or simply responding to our conditions randomly—we would expect to see no differences across these dimensions.

In fact, this is not what we observed. Figure 2 shows the effect of patients' age, SES, religion and race/ethnicity—independent of their chances of survival—on the chances that our pooled samples prescribed them a ventilator. (Supplementary Tables 5–7 provide the estimates for Fig. 2, along with the corresponding Bayes factors (BFs) for these estimates; Supplementary Table 15 provides the P values adjusted for multiple comparisons that we mention below.)

When we pool all of our groups together within our three samples (that is, not breaking down our models by individual characteristics of the respondents), across all three groups, older patients were less likely to be chosen independent of their chances of survival (if given a ventilator). Interestingly, this age penalty was largest among doctors. All else being equal, physicians were 9.4 percentage points ($\beta_{\text{d.f.}} = 1,522$; two-tailed $P < 0.001$; $\beta = -9.4$ percentage points; 95% confidence interval, (-11.4, -7.4)) less likely to give a ventilator to an older patient (holding constant their probability of survival). This age penalty was present among the public ($\beta_{\text{d.f.}} = 2,079$; two-tailed $P < 0.001$; $\beta = -3.9$ percentage points; 95% confidence interval, (-5.4, -2.4)) and among elected politicians ($\beta_{\text{d.f.}} = 838$; two-tailed $P < 0.001$; $\beta = -5.4$ percentage points; 95% confidence interval, (-7.6, -2.9)), but these effects were statistically and substantively smaller than among doctors. (For formal statistical tests of differences in the age penalty across our three samples, see Supplementary Table 19.)

There were some interesting patterns in how our three samples treated religious minorities. Doctors as a whole were equally likely to prescribe a life-saving ventilator to a religious minority (that is, to a Muslim or Atheist) as to a religious majority (that is, to a Catholic) (Muslim patient: $\beta_{\text{d.f.}} = 1,522$; two-tailed $P = 0.63$; $\beta = 1.9$ percentage points; 95% confidence interval, (-0.7, 4.4); $\text{BF}_{[-2.0, 2.0]} = 0.036$; Atheist patient: $\beta_{\text{d.f.}} = 1,522$; two-tailed $P = 0.71$; $\beta = 1.3$ percentage points; 95% confidence interval, (-1.9, 3.3); $\text{BF}_{[-2.0, 2.0]} = 0.013$). However, in our pooled individual samples of politicians and the public, we found evidence that respondents were less likely to provide ventilators to patients from religious minority groups. With all other factors about the patient held constant, politicians were 6.5 percentage points ($\beta_{\text{d.f.}} = 838$; two-tailed $P < 0.001$; $\beta = -6.5$ percentage points; 95% confidence interval, (-9.7, -3.2)) less likely to give a scarce ventilator to an Atheist and 4.5 percentage points ($\beta_{\text{d.f.}} = 838$; two-tailed $P < 0.030$; $\beta = -4.5$ percentage points; 95% confidence interval, (-7.8, -1.3)) less likely to give one to a Muslim. Similarly, the public were 4.7 percentage points ($\beta_{\text{d.f.}} = 2,079$; two-tailed $P < 0.001$; $\beta = -4.7$ percentage points; 95% confidence interval, (-6.6, -2.7)) less likely to choose an Atheist patient and 5.3 percentage points ($\beta_{\text{d.f.}} = 2,079$; two-tailed $P < 0.001$; $\beta = -5.3$ percentage points; 95% confidence interval, (-7.2, -3.3)) less likely to choose a Muslim patient. Supplementary Table 19 confirms that the public and elected officials were statistically and substantively more likely to be punitive against Muslim and Atheist patients than physicians were.

Other patterns emerged regarding the race/ethnicity of the patient. Physicians were more likely to prescribe a ventilator to a racial/ethnic minority than to a non-Hispanic white patient, all else being equal. That is, doctors were 4.1 percentage points more likely to prescribe a ventilator to an Asian patient ($\beta_{\text{d.f.}} = 1,522$; two-tailed

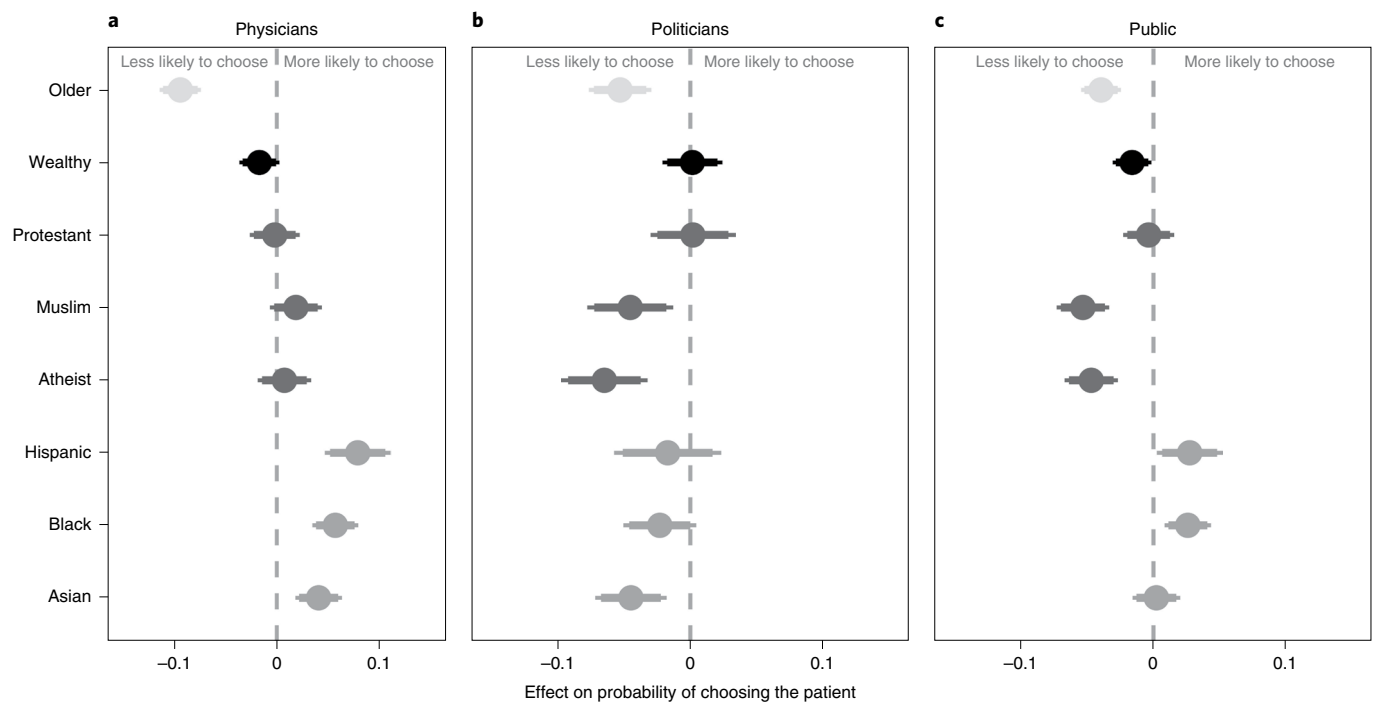


Fig. 2 | Patterns of physicians', politicians' and the public's allocation of ventilators to patients based on their age, SES, religion and race/ethnicity (pooled individual samples). **a–c.** Coefficient plots for the effects of patient characteristics (listed on the vertical axis) on the probability of choosing that patient type for a scarce ventilator (on the horizontal axis) among doctors (**a**), politicians (**b**) and the public (**c**). The circles represent the average marginal component effects, while the thick and thin bars show the 90% and 95% confidence intervals. Coefficients on the left side of the dashed line indicate a lower likelihood of choosing a patient with the given characteristics, all else being equal; those on the right indicate a higher likelihood of choosing a patient with the given characteristics, all else being equal. The baseline categories are young, low SES, Catholic and white. $N_{\text{docs}}=12,606$; $N_{\text{public}}=18,962$; $N_{\text{pol}}=7,442$.

$P < 0.001$; $\beta = 4.1$ percentage points; 95% confidence interval, (1.8, 6.3)), 5.7 percentage points more likely to prescribe a ventilator to a Black patient ($\beta_{\text{d.f.}} = 1,522$; two-tailed $P < 0.001$; $\beta = 5.7$ percentage points; 95% confidence interval, (3.5, 7.9)) and 7.9 percentage points more likely to prescribe a ventilator to a Hispanic or Latino patient ($\beta_{\text{d.f.}} = 1,522$; two-tailed $P < 0.001$; $\beta = 7.9$ percentage points; 95% confidence interval, (4.7, 11.1)). Among the public, there is evidence of preferential treatment of Black patients ($\beta_{\text{d.f.}} = 2,079$; two-tailed $P = 0.02$; $\beta = 2.6$ percentage points; 95% confidence interval, (0.8, 4.3)) but no statistical evidence of differential treatment of Hispanic patients ($\beta_{\text{d.f.}} = 2,079$; two-tailed $P = 0.09$; $\beta = 2.7$ percentage points; 95% confidence interval, (0.3, 5.2); $\text{BF}_{[-2.0, 2.0]} = 0.415$) or Asian patients ($\beta_{\text{d.f.}} = 2,079$; $\beta = 0.2$ percentage points; 95% confidence interval, (-1.6, 2.0); $\text{BF}_{[-2.0, 2.0]} = 0.005$) relative to white patients. Among elected politicians, there is no evidence of preferential treatment of Hispanic patients ($\beta_{\text{d.f.}} = 838$; two-tailed $P = 0.79$; $\beta = -1.7$ percentage points; 95% confidence interval, (-5.7, 2.3); $\text{BF}_{[-2.0, 2.0]} = 0.149$) or Black patients ($\beta_{\text{d.f.}} = 838$; two-tailed $P = 0.44$; $\beta = -2.3$ percentage points; 95% confidence interval, (-5.0, 0.4); $\text{BF}_{[-2.0, 2.0]} = 0.254$). In fact, both of the coefficients for these groups are negative (though not statistically significant). However, elected politicians do seem to discriminate against Asian patients. All else being equal, elected politicians are 4.5 percentage points ($\beta_{\text{d.f.}} = 838$; two-tailed $P = 0.02$; $\beta = -4.5$ percentage points; 95% confidence interval, (-7.2, -1.8)) less likely to recommend prescribing a ventilator to a patient of Asian descent than to a non-Hispanic white patient. (For formal tests of differences in responsiveness to racial/ethnic minority patients across our three samples, see Supplementary Table 19.)

Finally, regarding the wealth of the patient, the public showed no differential treatment of wealthy persons ($\beta_{\text{d.f.}} = 2,079$; $P = 0.090$; $\beta = -1.6$ percentage points; 95% confidence interval, (-3.1, -0.1);

$\text{BF}_{[-2.0, 2.0]} = 0.067$). Moreover, doctors and elected politicians showed no statistically significant evidence for a bias based on wealth (elected politicians: $\beta_{\text{d.f.}} = 838$; two-tailed $P = 0.95$; $\beta = 0.2$ percentage points; 95% confidence interval, (-2.1, 2.4); $\text{BF}_{[-2.0, 2.0]} = 0.016$; doctors: $\beta_{\text{d.f.}} = 1,522$; two-tailed $P = 0.11$; $\beta = -1.7$ percentage points; 95% confidence interval, (-3.7, 0.3); $\text{BF}_{[-2.0, 2.0]} = 0.082$). Overall, our results point towards patient wealth not playing a very large role in the chances of being chosen to have a scarce ventilator.

Differences in response based on survivability being tied or not. Did the participants in our study behave differently when the probabilities of the two patients surviving were tied as opposed to when they were not? Answering this question is vitally important, as, in the real world, decision makers have to make decisions both when they have two patients with similar odds of survival and when patients' survivability differs²⁰. Examining this potential treatment effect heterogeneity, then, adds to the construct and ecological validity of our estimates. Before providing these results, we note that even when the probabilities of survival are tied, the participants need not exhibit bias along the dimensions that we study. If the participants were not biased when patients' probabilities of survival were tied, we would expect to see null effects for our other patient characteristics; after all, even when patients' probabilities of survival were tied, the respondents could pick patients randomly rather than on the basis of their individual characteristics.

Figures 3 and 4 show our results broken down by whether the patients' probabilities of survival were tied (see Supplementary Tables 8, 9 and 15 for the output associated with these figures). In practice, we observed that differential treatment based on patient traits occurs mostly when patients' probabilities of survival are tied—that is, when doctors face the most pressure in making decisions about whom to allocate scarce medical resources.

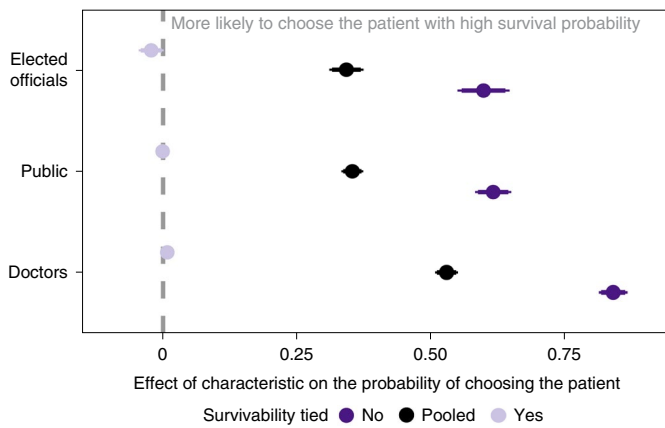


Fig. 3 | Patterns of physicians', politicians' and the public's allocation of ventilators based on patients' probability of survival (by probability of survival tied or not). Coefficient plot for the effect of patient probability of survival (high (70%) versus low (10%)) if they have access to a ventilator on the probability of choosing that patient type. The circles represent the effect estimates, while the thick and thin bars show the 90% and 95% confidence intervals. The light purple plotted coefficients denote when the survival estimates are tied and is, as expected, statistically and substantively zero. Coefficients on the left side of the dashed line indicate a lower likelihood of choosing a patient with the given characteristic, all else being equal; those on the right indicate a higher likelihood of choosing a patient with the given characteristic, all else being equal. As in all conjoint experiments, the unit of analysis is the respondent-comparison pair–individual profile. Pooled $N_{\text{docs}} = 12,596$; pooled $N_{\text{public}} = 18,962$; pooled $N_{\text{pols}} = 7,442$.

In this scenario, physicians were 8.4 percentage points more likely to choose an Asian patient (over a white patient; $\beta_{\text{d.f.}} = 1,161$; two-tailed $P < 0.001$; $\beta = 8.4$ percentage points; 95% confidence interval, (4.1, 12.8)), 6.9 percentage points more likely to choose a Black patient ($\beta_{\text{d.f.}} = 1,161$; two-tailed $P < 0.001$; $\beta = 6.9$ percentage points; 95% confidence interval, (2.6, 11.2)), 12.6 percentage points more likely to choose a Hispanic or Latino patient ($\beta_{\text{d.f.}} = 1,161$; two-tailed $P < 0.001$; $\beta = 12.6$ percentage points; 95% confidence interval, (5.5, 19.6)), 6.7 percentage points more likely to choose a Muslim patient (over a Catholic; $\beta_{\text{d.f.}} = 1,161$; two-tailed $P = 0.030$; $\beta = 6.7$ percentage points; 95% confidence interval, (1.3, 12.1)), 6.1 percentage points more likely to choose a patient of lower SES ($\beta_{\text{d.f.}} = 1,161$; two-tailed $P < 0.001$; $\beta = 6.1$ percentage points; 95% confidence interval, (1.9, 10.2)) and 24.7 percentage points less likely to choose an older patient ($\beta_{\text{d.f.}} = 1,161$; two-tailed $P < 0.001$; $\beta = -24.7$ percentage points; 95% confidence interval, (-29.0, -20.4)). All other effects were not statistically significant, as can be seen in Supplementary Table 9.

Among our pooled sample of members of the public, there were also signs of bias when the probabilities of patients' survival were tied. However, the responses of members of the public were somewhat different from those of physicians. (For a full statistical comparison of how our three samples responded when the probabilities of survival were tied, see Supplementary Table 20.) When the probabilities of survival were tied, members of the public were 5.1 percentage points more likely to choose a Black patient ($\beta_{\text{d.f.}} = 1,935$; two-tailed $P < 0.030$; $\beta = 5.1$ percentage points; 95% confidence interval, (1.5, 8.8)), 8.7 percentage points more likely to choose a Hispanic or Latino patient (over a white patient; $\beta_{\text{d.f.}} = 1,935$; two-tailed $P = 0.01$; $\beta = 8.7$ percentage points; 95% confidence interval, (3.4, 14.0)), 7.6 percentage points less likely to choose a Muslim patient ($\beta_{\text{d.f.}} = 1,935$; two-tailed $P < 0.001$; $\beta = -7.6$ percentage points; 95% confidence interval, (-3.3, -11.8)),

8.2 percentage points less likely to choose an Atheist patient ($\beta_{\text{d.f.}} = 1,935$; two-tailed $P < 0.001$; $\beta = -8.2$ percentage points; 95% confidence interval, (-3.8, 12.6)) and 9.7 percentage points less likely to choose an older patient ($\beta_{\text{d.f.}} = 1,935$; two-tailed $P < 0.001$; $\beta = -9.7$ percentage points; 95% confidence interval, (-6.5, 12.9)). All other effects were not statistically significant, as can be seen in Supplementary Table 9.

Our pooled sample of elected politicians also showed signs of bias when the probabilities of patients' survival were tied. When the probabilities of survival were tied, elected politicians were biased against Atheist patients ($\beta_{\text{d.f.}} = 756$; two-tailed $P = 0.01$; $\beta = -11.1$ percentage points; 95% confidence interval, (-18.0, -4.1)) and older patients ($\beta_{\text{d.f.}} = 756$; two-tailed $P < 0.001$; $\beta = -12.9$ percentage points; 95% confidence interval, (-18.0, -8.0)). However, all other effects were not statistically significant, as can be seen in Supplementary Table 9.

Figure 3 confirms that, as we would expect, when patients' probabilities of survival differ, our participants were more likely to choose the patient with a high probability survival and less likely to rely on other factors. This can be seen in Supplementary Table 8. Taken together, these results suggest that when patients have similar likelihoods of survival, differential treatment occurs.

Treatment effect heterogeneity. Did our main effects mask substantial heterogeneity across the characteristics of the respondents in our sample? None of these tests of treatment effect heterogeneity were specified a priori. As such, these should be treated as post hoc examinations for treatment effect heterogeneity. To be thorough, however, we test the entire universe of possible treatment effects across all of our baseline variables. As we show below, respondents' political affiliation—measured by their political party, ideology, approval of (then) President Donald Trump or approval of President Trump's handling of the COVID-19 pandemic—was the only construct that consistently showed signs of treatment effect heterogeneity.

Interestingly, all three samples (physicians, politicians and the public) responded to the signal of survivability. Figure 5 shows this visually. The corresponding estimates for this figure can be found in Supplementary Table 10.

Figure 6 shows the effects of the other characteristics we tested broken down by respondents' political parties. (For the results broken down by both survivability and political party, see Supplementary Tables 16–18.) Here, each panel is a different subset of our samples—Republicans (top left), Democrats (top right) and Independents (bottom left). The bottom right panel tests for differences between Republicans and Democrats. (For the corresponding estimates in table form, see Supplementary Tables 11–15.) Republican physicians showed consistent signs of bias against religious minorities (that is, Muslims and Atheists) in the prescribing of life-saving medical equipment. All else about the patient being held constant, Republican doctors were 8.1 percentage points less likely to prescribe a needed ventilator to an Atheist patient ($\beta_{\text{d.f.}} = 255$; two-tailed $P = 0.03$; $\beta = -8.1$; 95% confidence interval, (-14.2, -2.0)). Republican physicians were likewise 7.4 percentage points less likely to prescribe a ventilator to a Muslim patient ($\beta_{\text{d.f.}} = 255$; two-tailed $P = 0.04$; $\beta = -7.4$; 95% confidence interval, (-13.3, -1.6)). Republican elected politicians were 9.3 percentage points less likely to give a scarce ventilator to an Atheist patient ($\beta_{\text{d.f.}} = 274$; two-tailed $P < 0.001$; $\beta = -9.3$; 95% confidence interval, (-14.8, -3.9)) and 8.8 percentage points less likely to give a ventilator to a Muslim patient ($\beta_{\text{d.f.}} = 274$; two-tailed $P < 0.001$; $\beta = -8.8$; 95% confidence interval, (-14.3, -3.3)). Republican members of the public were 8.3 percentage points less likely to give a scarce ventilator to an Atheist patient ($\beta_{\text{d.f.}} = 614$; two-tailed $P < 0.001$; $\beta = -8.3$; 95% confidence interval, (-12, -4.7)) and 12.1 percentage points less likely to give a ventilator to a Muslim patient ($\beta_{\text{d.f.}} = 614$; two-tailed $P < 0.001$;

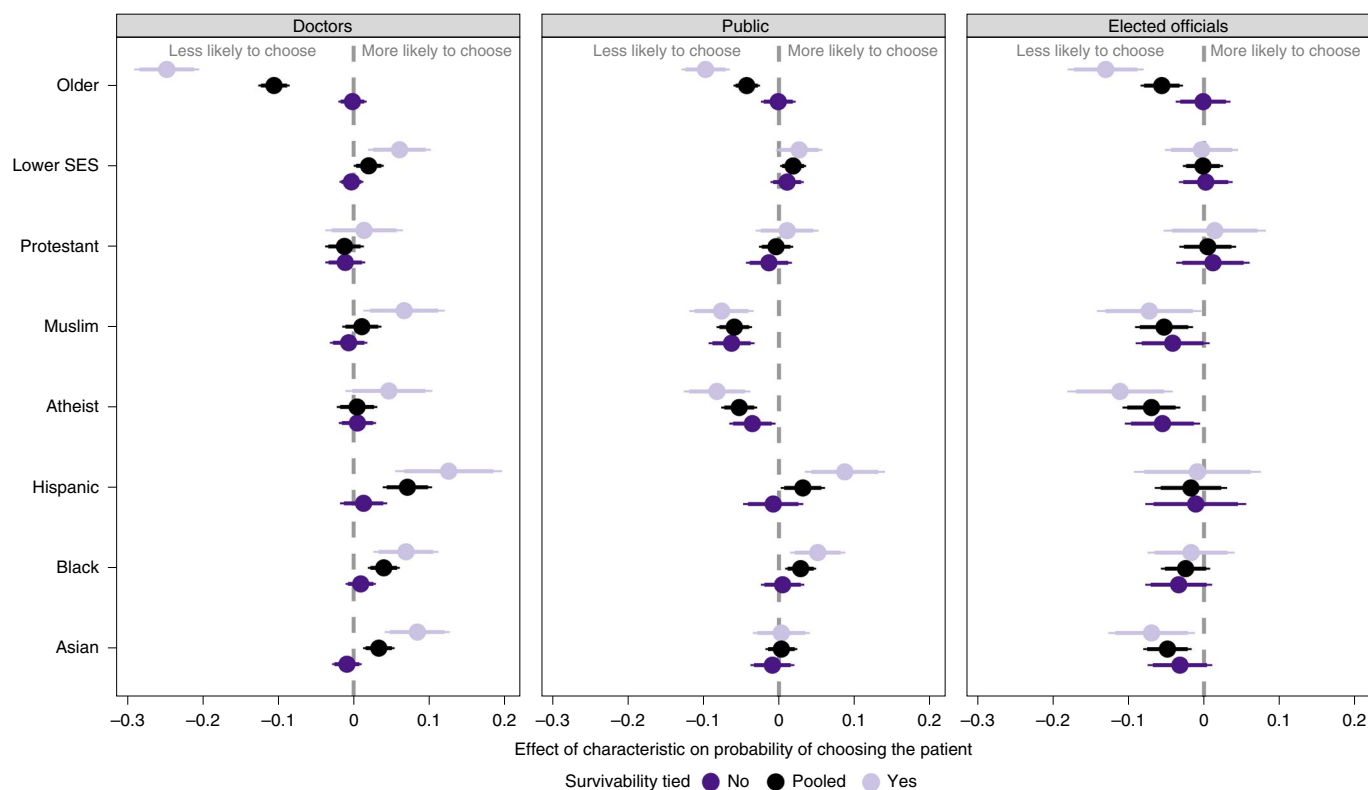


Fig. 4 | Patterns of physicians’, politicians’ and the public’s allocation of ventilators based on patient age, SES, religion and race/ethnicity (by probability of survival tied or not). Coefficient plots for the effects of patient characteristics (on the y axis) on the probability of choosing that patient type. The circles represent the effect estimates, while the thick and thin bars show the 90% and 95% confidence intervals. Coefficients on the left side of the dashed line indicate a lower likelihood of choosing a patient with the given characteristics, all else being equal; those on the right indicate a higher likelihood of choosing a patient with the given characteristics, all else being equal. As in all conjoint experiments, the unit of analysis is the respondent-comparison pair-individual profile. Docs, pooled, Asian $N=12,606$; docs, no, Asian $N=7,164$; docs, yes, Asian $N=5,432$; pub, no, Asian $N=9,758$; pub, yes, Asian $N=9,204$; pub, pooled, Asian $N=18,962$; pols, yes, Asian $N=3,572$; pols, pooled, Asian $N=7,442$; pols, no, Asian $N=3,870$; docs, no, Black $N=7,164$; docs, yes, Black $N=5,432$; docs, pooled, Black $N=12,606$; pub, pooled, Black $N=18,962$; pub, yes, Black $N=9,204$; pub, no, Black $N=9,758$; pols, yes, Black $N=3,572$; pols, pooled, Black $N=7,442$; pols, no, Black $N=3,870$; docs, no, Hispanic $N=7,164$; docs, yes, Hispanic $N=5,432$; docs, pooled, Hispanic $N=12,606$; pub, yes, Hispanic $N=9,204$; pub, pooled, Hispanic $N=18,962$; pub, no, Hispanic $N=9,758$; pols, pooled, Hispanic $N=7,442$; pols, yes, Hispanic $N=3,572$; pols, no, Hispanic $N=3,870$; docs, no, Atheist $N=7,164$; docs, pooled, Atheist $N=12,606$; docs, yes, Atheist $N=5,432$; pub, pooled, Atheist $N=18,962$; pub, yes, Atheist $N=9,204$; pub, no, Atheist $N=9,758$; pols, no, Atheist $N=3,870$; pols, yes, Atheist $N=3,572$; pols, pooled, Atheist $N=7,442$; docs, pooled, Muslim $N=12,606$; docs, yes, Muslim $N=5,432$; docs, no, Muslim $N=7,164$; pub, no, Muslim $N=9,758$; pub, pooled, Muslim $N=18,962$; pub, yes, Muslim $N=9,204$; pols, pooled, Muslim $N=7,442$; pols, yes, Muslim $N=3,572$; pols, no, Muslim $N=3,870$; docs, yes, Protestant $N=5,432$; docs, no, Protestant $N=7,164$; docs, pooled, Protestant $N=12,606$; pub, yes, Protestant $N=9,204$; pub, pooled, Protestant $N=18,962$; pub, no, Protestant $N=9,758$; pols, pooled, Protestant $N=7,442$; pols, yes, Protestant $N=3,572$; pols, no, Protestant $N=3,870$; docs, no, wealthy $N=7,164$; docs, pooled, wealthy $N=12,606$; docs, yes, wealthy $N=5,432$; pub, no, wealthy $N=9,758$; pub, yes, wealthy $N=9,204$; pub, pooled, wealthy $N=18,962$; pols, yes, wealthy $N=3,572$; pols, no, wealthy $N=3,870$; pols, pooled, wealthy $N=7,442$; docs, no, older $N=7,164$; docs, pooled, older $N=12,606$; docs, yes, older $N=5,432$; pub, pooled, older $N=18,962$; pub, yes, older $N=9,204$; pub, no, older $N=9,758$; pols, pooled, older $N=7,442$; pols, no, older $N=3,870$; pols, yes, older $N=3,572$.

$\beta = -12.1$; 95% confidence interval, $(-15.7, -8.5)$). Republican physicians and members of the public showed no evidence of discrimination on the basis of race or SES. These estimates can be found in full in Supplementary Table 11. Republican doctors did, however, show signs of bias against older patients ($\beta_{d.f.} = 255$; two-tailed $P < 0.001$; $\beta = -9.9$; 95% confidence interval, $(-14.5, -5.3)$.)

In comparison, Democratic doctors (but not elected officials or the public) showed favouritism towards Muslim patients ($\beta_{d.f.} = 459$; two-tailed $P < 0.001$; $\beta = 7.7$; 95% confidence interval, $(3.7, 11.7)$), Atheist patients ($\beta_{d.f.} = 459$; two-tailed $P < 0.001$; $\beta = 7.0$; 95% confidence interval, $(2.5, 11.4)$), Hispanic patients ($\beta_{d.f.} = 459$; two-tailed $P < 0.001$; $\beta = 12.6$; 95% confidence interval, $(7.4, 17.8)$) and Black patients ($\beta_{d.f.} = 459$; two-tailed $P = 0.03$; $\beta = 8.4$; 95% confidence interval, $(4.4, 12.3)$), over all-else-equal religious majority (that is, Catholic) and white patients. Democratic members of the public and elected politicians showed

no signs of bias for or against religious and racial/ethnic minorities (Supplementary Tables 12 and 15).

Figure 6d shows these differences across Republicans and Democrats. Republican physicians, elected politicians and members of the public consistently recommended providing less favourable treatment to Muslim patients and Hispanic patients. Republican physicians and members of the public recommended providing less favourable treatment to Atheist patients. (For these formal tests, see Supplementary Table 21.)

Interestingly, these biases among partisans of various backgrounds arose mainly when the two comparison patients had the same probability of survival. Again, these results suggest that when patients have similar odds of survival—as is often the case during pandemics and times of scarcity—key opportunities for differential treatment arise and are realized. The full set of estimates can be found in Supplementary Tables 12–14 and 16–18.

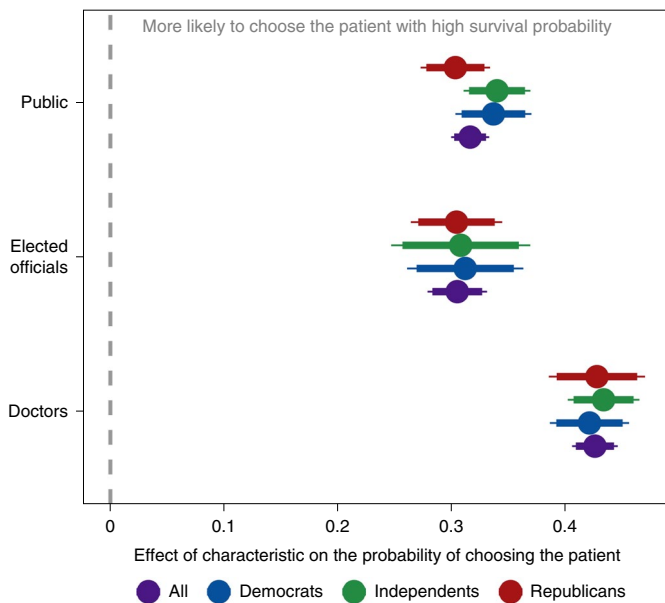


Fig. 5 | Patterns of physicians', politicians' and the public's allocation of ventilators based on patients' probability of survival (by party of the respondent). Coefficient plot for the effect of patient probability of survival (high (70%) versus low (10%)) if they have access to a ventilator on the probability of choosing that patient type. The estimates are broken down by sample (y axis) and by the political parties of the sample members (colours). The circles represent the effect estimates, while the thick and thin bars show the 90% and 95% confidence intervals. Coefficients on the right side of the dashed line indicate a higher likelihood of choosing a patient with the given characteristic, all else being equal. As in all conjoint experiments, the unit of analysis is the respondent-comparison pair-individual profile. Pooled $N_{\text{docs}} = 12,596$; pooled $N_{\text{public}} = 18,962$; pooled $N_{\text{pols}} = 7,442$; Republican $N_{\text{docs}} = 2,304$; Republican $N_{\text{public}} = 6,140$, Republican $N_{\text{pols}} = 2,740$; Democrat $N_{\text{docs}} = 4,024$; Democrat $N_{\text{public}} = 4,470$; Democrat $N_{\text{pols}} = 1,990$; Independent $N_{\text{docs}} = 4,450$; Independent $N_{\text{public}} = 5,630$; Independent $N_{\text{pols}} = 1,520$.

Some may wonder whether our treatment effects vary by any other characteristics not involving political leanings. The answer is that the clearest evidence of treatment heterogeneity, by far, is with the political leanings of the respondent. We provide all estimates of treatment effect heterogeneity in full in Supplementary Table 21. However, to summarize these findings, Fig. 7 shows the results from models that test for treatment effect heterogeneity across all of our baseline variables. (An analogue to Fig. 7 is Supplementary Fig. 6, which breaks the results down by each of the political variables and shows that treatment effect heterogeneity seems to be more prevalent along all of these dimensions than non-political dimensions.) Figure 7 shows the distribution of t -statistics from the tests of treatment effect heterogeneity across our eight treatments in the conjoint experiment. When considering all baseline variables, variables that captured respondents' political leanings (partisanship, ideology and Trump approval) showed the most signs of heterogeneity. In Fig. 7—across all three of our samples—political variables comprise most of the statistically significant tests of treatment effect heterogeneity (that is, the black dots in the right panels are more frequently outside the red dashed lines that demarcate statistically significant tests). Among elected officials, 74% of the significant tests of treatment effect heterogeneity are variables that measure the respondents' political leanings. Among the public, that number is 70%; among physicians, it is 61%. Also apparent is that political variables produce larger test statistics, on average, across all samples (that is, the distribution is flatter and wider in the right

panels than in the left panels). Among elected officials, the average t -statistic for political variables is 1.4, while for non-political variables it is only 0.89. Among physicians, the average t -statistic is 1.8 for political variables and 0.9 for non-political variables. Among the public, the average t -statistic is 2.2 for political variables and 1.04 for non-political variables.

Discussion

Here we reported results from three conjoint experiments designed to reveal the preferences of key stakeholders in the allocation of life-saving medical equipment. Our study included the physicians that prescribe life-saving medical equipment, the elected politicians that make key decisions about how to allocate public resources (which constrain the decisions of health-care workers on the ground) and the public that receives life-saving treatment and oversees the decisions of public officials and doctors. We chose a conjoint design that leveraged hypothetical patients 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 the prescription of ventilators across various demographic groups—would be a highly problematic way of testing for individual biases in this domain. With this research design, we could not hold all other factors about the patient constant. As a result, if we saw differential rates of ventilator prescriptions across various demographic groups, we would never know whether they were caused by those features or other difficult-to-observe characteristics. Second, datasets that track who receives a ventilator do not include the full pool of patients that could have received a ventilator; that is to say, they do not include those that—ultimately, for whatever reason—did not receive this treatment. This means that any analyses using these data would probably suffer from post-treatment/collider biases^{105–108}. This form of bias is unavoidable when one only has administrative records documenting those that actually received a public good or service. These biases can profoundly undermine our ability to draw inferences about social inequities in public service provision. For example, refs. ^{87,88,109} show that in an analogous domain—that of administrative records tracking police violence—conditioning on those who 'successfully' make their way into an administrative dataset completely undermines our ability to measure biases in service provision. The same could hold true in any use of health records that track only those who successfully gain access to a ventilator. Leaving out those who were in the pool to receive ventilators, but didn't, could bias any naive estimates of differential treatment in this domain. Finally, even if we were willing to set these two large methodological problems aside, it is practically impossible to find an administrative health-care dataset that has the degree of coverage we would want to answer our research question. No readily available dataset exists that tracks ventilator usage across all states in the United States. Even though these datasets exist in small, non-representative areas of the country, most of these records do not include all of the individual patient attributes that we need to test for our set of biases. Also, many of these datasets are protected by strict legal restrictions. For all of these reasons, conjoint experiments with hypothetical patients are, in our view, the best feasible way to look for biases in the allocation of scarce life-saving medical treatments.

We found that decision makers (regardless of their baseline demographics) agree: deference should be given to patients that would be more likely to survive with treatment and to younger patients. Our results regarding survivability suggest that the core health-care principle of maximizing positive impact transcends the powerful pull of partisanship in the assignment of life-saving medical treatment. This is important because previous work has shown that politics shapes (nearly) all aspects of our lives^{110–114}. It is also notable given that our other results do not follow this pattern—politics fundamentally moderate how individuals respond to other

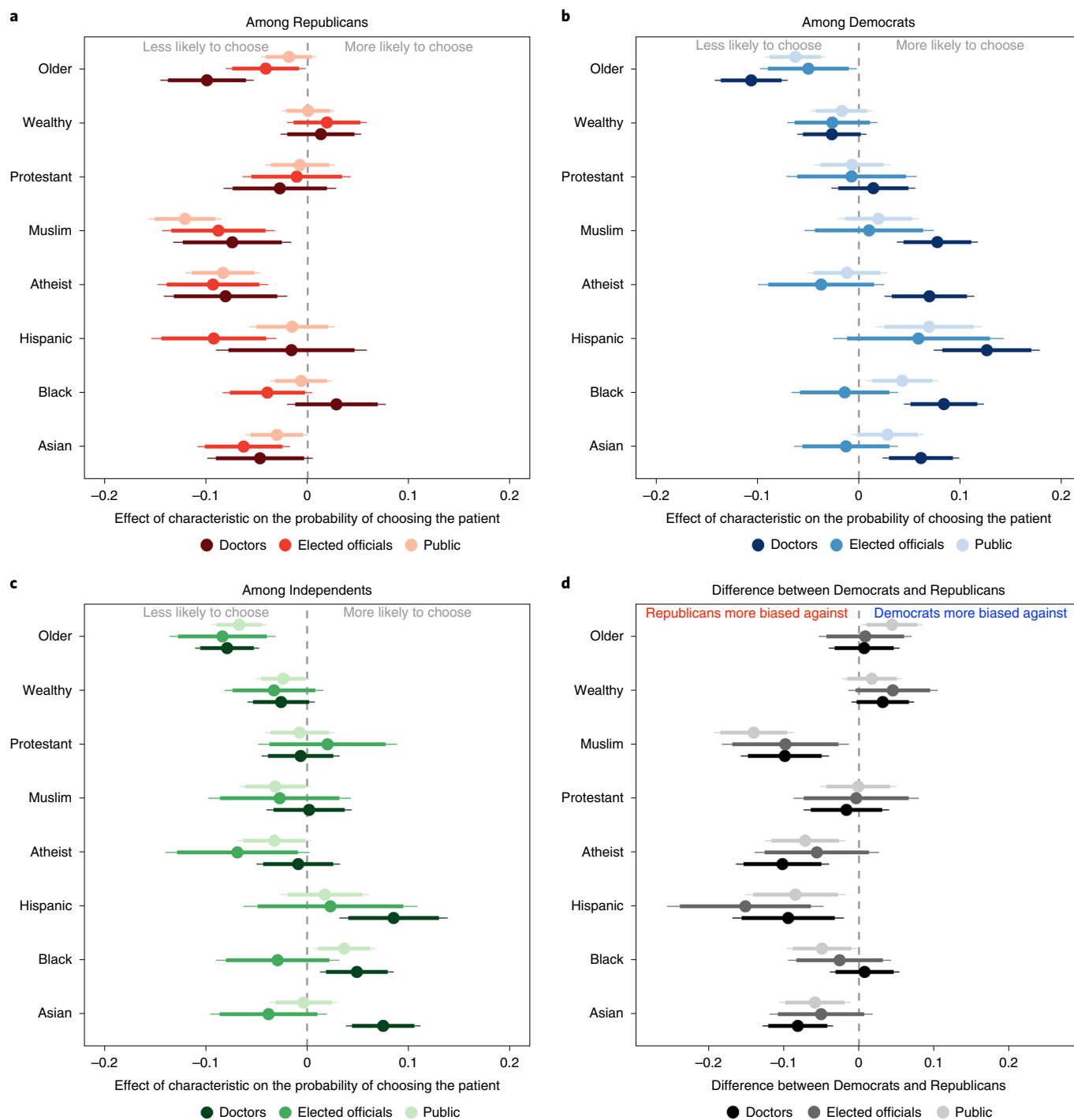


Fig. 6 | Patterns of physicians’, politicians’ and the public’s allocation of ventilators to patients based on their age, SES, religion and race/ethnicity (by party of the respondent). **a–d**, Coefficient plots for the effects of patient characteristics (listed on the vertical axis) on the probability of choosing that patient type for a scarce ventilator (on the horizontal axis) among Republicans (**a**), Democrats (**b**) and Independents (**c**). The circles represent the effects, while the thick and thin bars show the 90% and 95% confidence intervals. Coefficients on the left side of the dashed line indicate a lower likelihood of choosing a patient with the given characteristics, all else being equal; those on the right indicate a higher likelihood of choosing a patient with the given characteristics, all else being equal. The baseline categories are young, low SES, Catholic and white. As in all conjoint experiments, the unit of analysis is the respondent–comparison pair–individual profile. Republican $N_{docs} = 2,304$; Republican $N_{public} = 6,140$; Republican $N_{pols} = 2,740$; Democrat $N_{docs} = 4,024$; Democrat $N_{public} = 4,470$; Democrat $N_{pols} = 1,990$; Independent $N_{docs} = 4,450$; Independent $N_{public} = 5,630$; Independent $N_{pols} = 1,520$.

patient characteristics, when the probability of survival is held constant. What should we make of this preference for patients that are more likely to survive if they are given access to a ventilator? One way to interpret our results normatively is to conclude that all is

well in the medical space—that doctors (and members of the public and elected politicians) effectively discriminating against patients with a low probability of survival (or for those with a high probability) is what we would expect to happen in the real world. (After all,

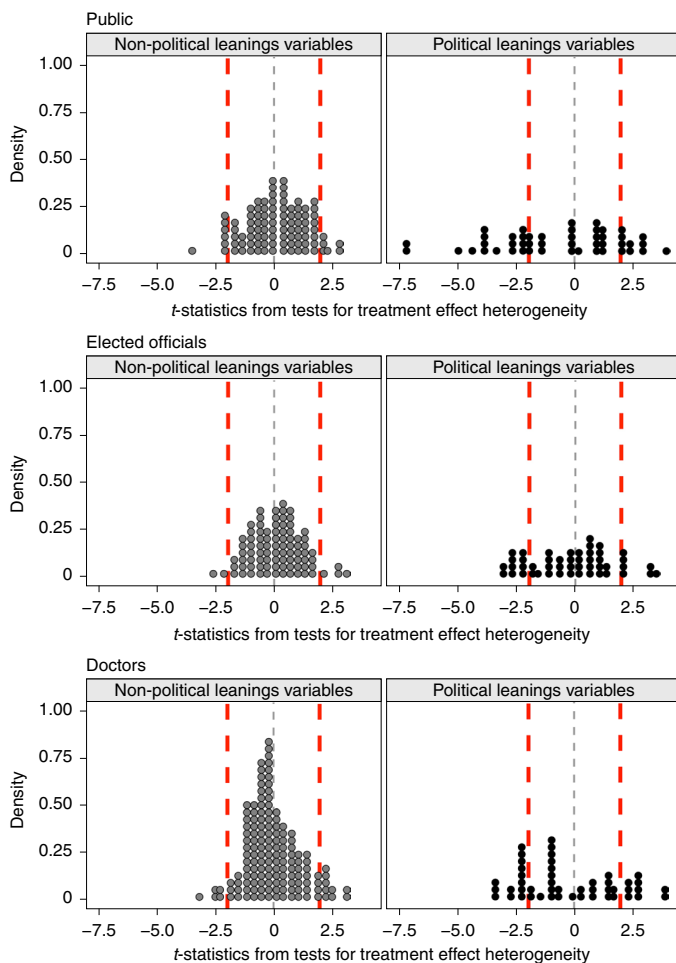


Fig. 7 | Tests for treatment effect heterogeneity, broken down by measures of respondents' political leanings and other baseline variables. The *t*-statistics are shown for estimates of treatment effect heterogeneity broken down by all of our baseline variables. The full list of baseline variables is included in the Methods. The political variables include political party, ideology, Trump approval and Trump approval relating to COVID-19. The grey dashed lines show 0, and the red dashed lines show $\text{abs}(1.96)$ values.

as we mentioned earlier, many have assumed that this is the case.) However, there may also be reasons to view this type of physician, politician and public behaviour at least somewhat warily. In the real world, subjective judgements of survivability may be strongly correlated with certain individual traits such as disability status, race and pre-existing conditions^{47–52}. One way of interpreting the results in Fig. 1 is that the decision to provide ventilators to individuals with a high likelihood of survival could, in part, be driven by discriminating against individuals who have traits correlated with survivability. Further research would do well to explore this question.

We also found that when survivability was tied, physicians (on average) showed evidence of prioritizing the allocation of ventilators to racial/ethnic minorities (Asian, Black, and Hispanic or Latino) over white patients, as well as to Muslim patients and patients of lower SES. Interestingly, physicians (on average) showed evidence of preferential treatment of Asian patients rather than bias against Asian patients, which has been shown in other contexts¹¹⁵. The general public (on average) also favoured Black and Hispanic or Latino patients but showed evidence of bias against religious minorities (Muslim or Atheist). Elected officials (on average) shared with the public the bias against religious minorities but also exhibited

evidence of bias against Asian patients. In short, when patients' probabilities of survival are similar—as was the case during the COVID-19 pandemic—there were ripe opportunities for decision makers' biases to come into play. These results suggest that during pandemics and times of scarcity, key opportunities for differential treatment of otherwise all-else-equal patients arise and are realized.

Republican and Democratic physicians, politicians and members of the public often disagreed about how to prioritize scarce medical equipment during the pandemic. Republicans and Democrats often treated racial/ethnic and religious minorities differently—Democratic physicians and members of the public (but not elected politicians) tended to support providing favourable treatment to racial/ethnic (Hispanic, Black and Asian) and religious minorities (Muslims and Atheists), and Republicans consistently showed bias towards religious-majority patients, with some evidence of bias against some racial/ethnic minorities, particularly among Republican elected politicians.

Our work has important practical and theoretical implications. From a theoretical perspective, our work speaks to the broader study of discrimination. Although a large experimental literature has studied discrimination, our work overcomes key hurdles in this area of inquiry. Much of the work in this area has used audit studies to examine discrimination in low-stakes environments—for example, requests for help in small tasks or in getting information^{32,34}. And much of the research that social scientists do on discrimination in life-and-death decisions (for example, policing) has been limited by the fundamental biases that shape the data-generating process^{87,88,109}. Our work does have some important limitations. First, the public and politician recruitment occurred in the spring of 2020, while the recruitment of physicians occurred in the fall of 2020. This means that the data we collected from physicians occurred at a different point in the pandemic than our data collection with members of the public and elected officials. It could be that public understanding of health-care priorities and patient treatment changed between these two survey points. We acknowledge this possibility and encourage future research in different temporal contexts to assess the stability of our findings over time.

Second, our design relied on hypothetical scenarios, which by definition places constraints on ecological validity. However, even though we are testing for biases with hypothetical patients, there are reasons to suspect that the results from our conjoint experiments correlate with real-world behaviour. Indeed, a growing body of research from various fields has shown that results from conjoint experiments strongly predict real-world behaviours^{95–103}. For example, Hainmueller and Hopkins show that conjoint experiments that elicit citizens' attitudes towards immigrants are strong predictors of citizens' behaviours in this domain⁹⁶. Likewise, Auerbach and Thachil show that conjoint experiments strongly overlap with actual political behaviours in the real world¹⁰². Conjoint experiments also show minimal evidence of satisficing or demand effects¹¹⁶, even when a large number of attributes are randomized^{100,101}.

Third, as our study was focused on a high-stakes pandemic setting, our discrimination estimates may not hold in lower-stakes environments or in high-stakes environments caused by other non-virus threats. Our study was not designed to explore potential biases in day-to-day interactions between physicians and patients, citizens and their neighbours, and politicians and their constituents. We think that much more work on these topics, particularly the first, is needed.

Fourth, our conjoint experiments may miss out on other potential factors that influence decision-making (for example, physical appearance). As in any study, we cannot identify all of the many causal factors that influence our dependent variable (that is, prescription patterns); however, our study does explore several key factors (for example, survivability, race/ethnicity, age and religion) regarding whether they influence behaviour in health-care provision.

Fifth, our experiments were not designed to tease out the many potential mechanisms behind the effect estimates we provide. In this paper, we focused on the first-order question of whether differential treatment exists. Future research would do well to examine why this differential treatment exists, which could be the product of a mix of potential attitudinal, behavioural and contextual mechanisms.

Finally, some of our treatments were relatively light touch. For example, our race/ethnicity signals came from variants of individuals' names. Given this, our discrimination estimates may be considered a lower bound.

Practically speaking, however, our work highlights the discrepancies and biases among key decision makers in the health-care domain. Narrowly, our results speak to biases that arise in the prescription patterns of ventilators. However, our results could also have implications for other medical scenarios where quantities are limited and decision makers have to make life-and-death choices (for example, vaccines or, more generally, other scarce therapies used to treat sick patients). Understanding how key actors in the health-care domain make decisions in these types of scenarios is vitally important given the current finite quantities of life-saving medical equipment and the high stakes. Ultimately, our results are indicative of the forces that shape individual decision-making during times of crisis. Our results suggest that when life-and-death decisions need to be made, decision makers—regardless of their background and training—may, in some instances, rely on the systemic biases that fundamentally govern human behaviour. When societies are unprepared for forced instances of scarcity—given a lack of preparation by policymakers or other societal agents—individual heuristics and biases reveal themselves in key decision-making processes.

We note that there remain real ethical ambiguities about who should be given preference in moments of scarcity. The biases that we document may reflect how the broader health-care system trains medical professionals, which is—in large part—outside the control of these individual professionals. Physicians are, after all, trained to make decisions based on health characteristics and receive limited, if any, training about how individual social biases may come about. This reflects itself in our results: when it comes to the health of a patient, physicians act consistently with their training. However, on other individual characteristics that doctors often know about patients, the same cannot be said. Moreover, given that the scarcity of life-saving medical equipment in the current pandemic is largely a function of public policy, we have some reason to lay these inequality-enhancing responses at the feet of elected politicians who control the response to the current pandemic and the preparation for future pandemics. In a system with scarce life-saving medical resources, individual biases play an important role.

Methods

Ethics. Our research was conducted with human participants and, as such, was approved beforehand by the Brigham Young University Institutional Review Board (E17512 and X19052). Our research complied with all relevant ethical regulations. Informed consent was obtained from all human participants. The participants were not compensated for their participation.

Samples and sampling methods. The overall response rates were 2.0% for the public, 5.7% for elected officials and 1% for physicians. Our end sample consisted of 2,080 members of the public, 839 elected politicians and 1,523 doctors that completed our experimental vignettes. Participants of all genders/sexes and ages were invited to be included in the study; for the breakdown of these characteristics, see Supplementary Tables 1–3. No statistical methods were used to predetermine sample sizes. We invited the population of elected politicians and physicians included in our lists to participate in our study. Among the public, our target was to invite 5,000 individuals per state to take the survey. Data collection for the public and elected officials ran from 1 April to 15 April 2020; data collection for doctors ran from 20 October to 12 November 2020.

We included physicians in our conjoint experiments because these are the front-line workers that make key decisions about which patients receive treatment. (As we show in the Supplementary Information (Supplementary Figs. 1 and 2), our results are the same regardless of whether we focus on physicians that prescribe

ventilators, meet with COVID patients and/or are simply in fields directly involved in or adjacent to the treatment of COVID patients.) We included elected politicians in our conjoint experiments given the key role that politicians play in health-care policy generally and in managing the stocks and flows of ventilators specifically. We included the public in our conjoint experiments for three reasons: first, the public serves as a useful benchmark because they are the (potential) recipients of life-saving medical care; second, because the behaviour of elected politicians is often, at its roots, partially driven by the preferences of the public^{117–121}; and third, because the behaviour of health-care providers may too be shaped by the preferences of the public. Importantly, physicians, politicians and members of the public usually form the committees tasked with making guidelines and decisions about who receives life-saving medical services^{14,22–31}, thus making them important bodies to study in the current context.

The samples for our three groups came from three separate sources. Our public sample was drawn from voter registration lists. It comprised randomly chosen individuals from a nationwide voter registration list compiled and maintained by the data and analytics firm DT Client Services LLC (commonly known as 'The Data Trust'), which has matched these individuals with their email addresses. Our elected official sample consisted of the population of state legislators, mayors and city councillors in the United States. Contact information for these political elites was scraped from publicly available sources via openstates.org (these data were collected and used in compliance with the terms of use for this platform) and collected from a comprehensive email list provided to us by the purveyors of the American Municipal Officials Survey²². Our physician sample was purchased from a data and analytics firm called Paragon Concepts that runs the website Doctordatabases.com, which scrapes health-care providers' websites and provides the contact information of physicians (this dataset also was purchased and used in compliance with the terms of use for this platform). To increase the ecological validity of our study, we restricted our physician sample to include only practitioners that could reasonably be expected to treat COVID-19 patients. For more information on our samples, see the Supplementary Information.

Design and analysis. Social scientists commonly use conjoint experiments to help understand how people value different attributes of possible choices^{95–103}. Conjoint experiments were 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 preferences⁹⁵. Under a conjoint design, a researcher shows a survey or lab respondent a series of products (in our case, patient case files) and randomizes a set of potential product attributes (in our case, the attributes of the patients in the case files). 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.

Each of our conjoint experiments started with some basic information about potential shortages in medical ventilators, followed by a request to evaluate a set of five pairwise comparisons of case files of individuals who needed a ventilator and to select the individual who should receive life-saving medical care. (This instructional text is included in the Supplementary Information.) This type of conjoint—a discrete choice experiment where individuals select from at least two competing individuals, products or services—has been shown to be preferred over alternatives if one desires to generalize from conjoint choices to real-world behaviour⁹⁵. We also used this type of design because it best mimics the real-world dilemma that many doctors have faced—that is, what should be done “if two COVID-19 victims need ventilators and just one is available?”²⁰.

As in all conjoint experiments, the unit of analysis was the respondent-choice profile. All respondents were provided with a full profile, since we had a limited number of attributes (that is, eight). The *N* reported in our models is the number of respondents multiplied by the number of pairwise choices (that is, five) and individuals within those pairs (that is, two). In the conjoint, individuals were fully randomized to patient profiles without blocking. Data collection and analysis were not performed blind to the conditions of the experiments; the investigators were not blinded to allocation during the experiments and outcome assessment. Listwise deletion was used for individuals who completed only part of the conjoint experiment; aside from this condition, however, no data were systematically excluded from the analyses. The data in our analyses met the assumptions of the statistical tests used; normality and equal variances were not formally tested.

In our experimental vignettes, we randomized the following characteristics: survivability, age, SES, race/ethnicity and religion. Specifically, we randomized patients' estimated chances of survival—either high (70% chance of survival) or low (10% chance of survival)—if given access to a ventilator. We chose this as a key characteristic given (1) the central role that this feature plays in many health-care decisions and (2) the fact that task forces assigned to make decisions often have access to and rely on estimated survival odds. Given its central role, we manipulated survivability so that we could have a manipulation check and test whether doctors, politicians and the public were attentive to the patient profiles and were responding in a way that was consistent with what we would expect given the weight given to survivability in contemporary discussions of the COVID pandemic. If our experiment were ecologically valid, we would expect that survivability would have a statistically and substantively meaningful

effect on respondents' choices. Another reason that we chose to manipulate survivability was that it allowed us to account for a potentially important form of bias in our experimental design. Specifically stating the patients' probability of survival allowed us to account for possible differential biases in the perceptions of survivability based on other individual characteristics. This was important because previous research has shown that people may mistakenly infer differences in the probability of survival on the basis of (among other things) patients' racial/ethnic or disability status^{123–126}. While this means that our (necessary) manipulation check blocked a potential channel for discrimination (that is, one that arises from misperceptions about survivability that differ across races/ethnicities and individual disabilities), it left open the possibility of a direct, and arguably more pernicious, form of discrimination (that is, one that originates from those in our samples inherently valuing the lives of minorities less than those of white individuals, for example). This allowed us to estimate a vitally important quantity of interest—the extent to which physicians, politicians and the public exhibit a preference for discrimination independent of a key form of statistical discrimination^{127–129}.

In a similar vein, we also manipulated the age of the patient (younger versus older). In addition to the probability of survival and age, we randomized a set of other attributes that prior work on discrimination in the United States (including the health-care domain) has suggested might influence the quality of care—in this case, the provision of life-saving ventilators—that doctors, members of the public and elected politicians would provide to an individual. These characteristics included SES (high or low), religion (Catholic, Protestant, Atheist or Muslim) and race/ethnicity (non-Hispanic white, Hispanic or Latino, Black, or Asian). Following previous practice, we signalled SES through patient occupation (a construction worker versus an owner of several international construction firms). As previous research has established, one's occupation is the strongest and most consistent signal of social class⁵⁷. Drawing inspiration from the contemporaneous methodological approaches in the audit study literature^{33,34,130}, we used racially distinctive names that break clearly along demographic lines. Though names may constitute, in some respects, a bundle of sticks¹³¹, other ways of signalling race/ethnicity, such as providing subjects with photographs, may magnify this problem. We signalled age based on how long the person had been working in their randomized profession. We signalled religion by simply stating the patient's religion. We held the gender of the participant (male) constant to ensure that gender was not confounding our results.

All of the attributes in our conjoint designs were information that doctors (and elected officials and the public) can and often do know about their patients (and constituents/neighbours). Hence, all of these attributes could be used as heuristics of who deserves medical treatment. While a patient's religious affiliation is probably the least readily accessible characteristic that we study, there are still reasons to explore whether physicians, politicians and the public discriminate on the basis of religion. Many hospitals ask patients their religious preferences, making it possible that doctors would also know this information. Moreover, many religious traditions have overt religious symbols (such as clothing and jewellery) that would be easily observable in a variety of social contexts (in a hospital, interacting with constituents or in social situations with one's neighbours). Furthermore, given that many religions have unique standards and preferences for health care, patients may be required to divulge their religion to their caretakers. Even in cases where religious affiliation is not observable, the participants in our study may learn patients' religious leanings from one-on-one social interactions with them.

We chose to test for biases in ventilator prescription for several reasons. First, we chose ventilators as opposed to other medical interventions because (1) there have been actual shortages of these in many geographic areas during the current pandemic, (2) many individuals have worried about shortages in other geographic areas, and (3) even in areas where there have been abundant ventilators, this was only the case after a shortage was seen and extensive resources were mobilized. This resulted in various decision-making units (comprising citizens, elected politicians and physicians) coming together to establish recommended best practices in prescribing ventilators. This increased the ecological validity of our task. Second, unlike other medical services, ventilator prescriptions were, at least in part, at the discretion of those decision makers in our sample. Another alternative to ventilators that we might have used was hospital beds, which were also in short supply during the COVID-19 pandemic. However, given underlying ethical principles related to not turning away patients, space inside a hospital is often determined not by heuristics but by who arrives first. Using beds, then, would not provide a good test for potential biases in prescription patterns. Finally, though there are different stages of disease progression and treatments—such as standard oxygen therapy—many COVID-19 patients still receive treatment from ventilators⁵, and with many unknowns about the new variants of the virus that causes COVID-19, many more still may require ventilator treatments in the future. (We did not use vaccines as our medical treatment to study because the vaccine had not been released at the time of the study.)

In our analyses, we focused on main effects (that is, among our entire samples) and on those broken down by all of the self-reported baseline characteristics that we collected in our study. None of these tests of treatment effect heterogeneity were specified a priori. As such, these should be treated as post hoc examinations for

treatment effect heterogeneity. To be thorough, however, we test the entire universe of possible treatment effects across all of our baseline variables.

We kept our list of baseline variables small to ensure that our survey was not an undue burden on the respondents. These questions focused on the demographic and political background of the participants in our study. The physician sample contained a few extra baseline variables that measure physicians' experiences during the pandemic, as we desired to have these descriptive quantities for future research.

Our baseline variables for physicians included these questions:

1. How many years have you been practicing medicine? (1–30+ years)
2. Have you treated or consulted patients who have (or are suspected to have) the COVID-19 virus? (Yes/no)
3. Do you have admitting privileges at a hospital? (Yes/no/does not apply to me)
4. In your normal work responsibilities, would you ever play a role in prescribing life support resources to a patient? (Yes/no)
5. To what extent do you agree or disagree with the following statement: I have had trouble obtaining personal protective equipment? (Strongly agree–strongly disagree)
6. Overall how would you rate your current mental health? (Excellent–poor)
7. I am afraid that I will get the coronavirus. (Strongly agree–strongly disagree)
8. I am afraid that people I care about will get the coronavirus. (Strongly agree–strongly disagree)
9. I am afraid that my patients will get the coronavirus. (Strongly agree–strongly disagree)
10. How closely do you follow the news? (Very closely–not closely at all)
11. Do you approve or disapprove of President Trump's overall job performance thus far? (Strongly approve–strongly disapprove)
12. Do you approve or disapprove of President Trump's handling of the coronavirus outbreak? (Strongly approve–strongly disapprove)
13. Generally speaking, do you consider yourself to be a(n): Republican; Democrat; Independent, or Unaffiliated; Something else
14. Generally speaking, would you describe your political views as: (Very liberal–very conservative)
15. Which of the following best describes how you think of yourself? (Male/female/in another way—please specify if you wish:)
16. What is your age? (Under 18–over 65)
17. Are you: American Indian/Native American, Asian, Black/African American, Hispanic/Latino, White/Caucasian, Pacific Islander and/or Other (please specify)?

The public and elected officials were asked questions 7, 8 and 10–17. All questions about patients were omitted from the surveys of the public and elected officials, as these would not have made sense to these samples.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The survey data for this article are available on the Open Science Framework at <https://doi.org/10.17605/OSF.IO/TNEQJ>.

Code availability

The supporting code for this article is available on the Open Science Framework at <https://doi.org/10.17605/OSF.IO/TNEQJ>.

Received: 9 March 2021; Accepted: 14 December 2021;

Published online: 24 February 2022

References

1. Truog, R. D., Mitchell, C. & Daley, G. Q. The toughest triage—allocating ventilators in a pandemic. *N. Engl. J. Med.* **382**, 1973–1975 (2020).
2. Iyengar, K., Bahl, S., Vaishya, R. & Vaish, A. Challenges and solutions in meeting up the urgent requirement of ventilators for COVID-19 patients. *Diabetes Metab. Syndr.* **14**, 499–501 (2020).
3. Beitler, J. R. et al. Ventilator sharing during an acute shortage caused by the COVID-19 pandemic. *Am. J. Respir. Crit. Care Med.* **202**, 600–604 (2020).
4. White, D. B. & Lo, B. A framework for rationing ventilators and critical care beds during the COVID-19 pandemic. *JAMA* **323**, 1773–1774 (2020).
5. Toussaint, K. These good Samaritans with a 3D printer are saving lives by making new respirator valves for free. *Fast Company* (16 March 2020).
6. Patel, B. K., Kress, J. P. & Hall, J. B. Alternatives to invasive ventilation in the COVID-19 pandemic. *JAMA* **324**, 43–44 (2020).
7. Ranney, M. L., Griffith, V. & Jha, A. K. Critical supply shortages—the need for ventilators and personal protective equipment during the COVID-19 pandemic. *N. Engl. J. Med.* **382**, e41 (2020).
8. Hollingsworth, H. Ventilator shortage as Missouri virus hospitalizations spike. *Associated Press* (2021).

9. McMahon, D. E., Peters, G. A., Ivers, L. C. & Freeman, E. E. Global resource shortages during COVID-19: bad news for low-income countries. *PLoS Negl. Trop. Dis.* **14**, e0008412 (2020).
10. Kumar, P. et al. Management of potential ventilator shortage in India in view of on-going COVID-19 pandemic. *Indian J. Anaesth.* **64**, S151–S152 (2020).
11. Krishnamoorthy, V., Vavilala, M. S. & Mock, C. N. The need for ventilators in the developing world: an opportunity to improve care and save lives. *J. Glob. Health* **4**, 010303 (2014).
12. Maclean, R. & Marks, S. 10 African countries have no ventilators. That's only part of the problem. *New York Times* (2020).
13. Naidoo, P. Vaccine and ventilator shortages show need for African free trade. *Bloomberg* (2021).
14. Miller, J. UK faces 'massive shortage' of ventilators—Swiss manufacturer. *Reuters* (2020).
15. Fink, S. Worst-case estimates for US coronavirus deaths. *New York Times* (18 March 2020).
16. *Stockpiling, Availability in the US* (Johns Hopkins Center for Health Security, 2020).
17. Kliff, S., Satariano, A., Silver-Greenberg, J. & Kulish, N. There aren't enough ventilators to cope with the coronavirus. *New York Times* (26 March 2020).
18. Fink, S. The rationing of a last-resort COVID treatment. *New York Times* (2021).
19. *Medical Device Shortages During the COVID-19 Public Health Emergency* Technical Report (US Food and Drug Administration, 23 December 2020).
20. Gandhi, R. & Patel, A. What if two COVID-19 victims need ventilators and just one is available. *Scientific American* (2020).
21. Beall, A. The heart-wrenching choice of who lives and dies. *BBC* (29 April 2020).
22. Bazerman, M., Bernhard, R., Greene, J. D., Huang, K. & Barak-Corren, N. How should we allocate scarce medical resources? *Harvard Business Review* (2020).
23. Jin, L., Huang, Y., Liang, Y. & Zhang, Q. Who gets the ventilator? Moral decision making regarding medical resource allocation in a pandemic. *J. Assoc. Consum. Res.* <https://doi.org/10.1086/711734> (2021).
24. Frakt, A. Who should be saved first? Experts offer ethical guidance. *New York Times* (29 March 2020).
25. Cohen, I. G., Crespo, A. M. & White, D. B. Potential legal liability for withdrawing or withholding ventilators during COVID-19: assessing the risks and identifying needed reforms. *JAMA* **323**, 1901–1902 (2020).
26. Piscitello, G. M. et al. Variation in ventilator allocation guidelines by US state during the coronavirus disease 2019 pandemic: a systematic review. *JAMA Netw. Open* **3**, e2012606 (2020).
27. Biddison, E. L. D. et al. Too many patients...a framework to guide statewide allocation of scarce mechanical ventilation during disasters. *Chest* **155**, 848–854 (2019).
28. Foggatt, T. Who gets a ventilator? *New Yorker* (20 April 2020).
29. Klein, A. Who will get ventilators in a COVID-19 crisis? *New Sci.* **245**(3276), 12.
30. *Americans Are Divided by Religion on Who Should Get Critical Care if There Is a Shortage of Ventilators* Technical Report (Pew Research, 24 April 2020).
31. White, D. B., Katz, M. H., Luce, J. M. & Lo, B. Who should receive life support during a public health emergency? Using ethical principles to improve allocation decisions. *Ann. Intern. Med.* **150**, 132–138 (2009).
32. Quillian, L., Pager, D., Hexel, O. & Midtbøen, A. H. Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proc. Natl Acad. Sci. USA* **114**, 10870–10875 (2017).
33. Butler, D. M. & Broockman, D. E. Do politicians racially discriminate against constituents? A field experiment on state legislators. *Am. J. Polit. Sci.* **55**, 463–477 (2011).
34. Costa, M. How responsive are political elites? A meta-analysis of experiments on public officials. *J. Exp. Polit. Sci.* **4**, 241–254 (2017).
35. Hassell, H. J., Holbein, J. B. & Miles, M. R. There is no liberal media bias in which news stories political journalists choose to cover. *Sci. Adv.* **6**, eaay9344 (2020).
36. Pfaff, S., Crabtree, C., Kern, H. L. & Holbein, J. B. Do street-level bureaucrats discriminate based on religion? A large-scale correspondence experiment among American public school principals. *Public Adm. Rev.* **81**, 244–259 (2020).
37. Carnes, N. & Holbein, J. Do public officials exhibit social class biases when they handle casework? Evidence from multiple correspondence experiments. *PLoS ONE* **14**, e0214244 (2019).
38. Crabtree, C. *Measuring and Explaining Discrimination*. PhD thesis, Univ. of Michigan (2019).
39. Gaddis, S. M. *Audit Studies: Behind the Scenes with Theory, Method, and Nuance* Vol. 14 (Springer, 2018).
40. Hoffman, K. M., Trawalter, S., Axt, J. R. & Oliver, M. N. Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proc. Natl Acad. Sci. USA* **113**, 4296–4301 (2016).
41. Trawalter, S., Hoffman, K. M. & Waytz, A. Racial bias in perceptions of others' pain. *PLoS ONE* **7**, e48546 (2012).
42. Trawalter, S. & Hoffman, K. M. Got pain? Racial bias in perceptions of pain. *Soc. Pers. Psychol. Compass* **9**, 146–157 (2015).
43. Ashton, C. M. et al. Racial and ethnic disparities in the use of health services. *J. Gen. Intern. Med.* **18**, 146–152 (2003).
44. Schulman, K. A. et al. The effect of race and sex on physicians' recommendations for cardiac catheterization. *N. Engl. J. Med.* **340**, 618–626 (1999).
45. Kawachi, I., Subramanian, S. & Almeida-Filho, N. A glossary for health inequalities. *J. Epidemiol. Community Health* **56**, 647–652 (2002).
46. Graham, H. *Understanding Health Inequalities* (McGraw-Hill Education, 2009).
47. Chin, N. M. & Harris, J. *Examining How Crisis Standards of Care May Lead to Intersectional Medical Discrimination Against COVID-19 Patients* Technical Report (The Center for Public Representation, 16 February 2021).
48. Webb, B. J. et al. Simple scoring tool to estimate risk of hospitalization and mortality in ambulatory and emergency department patients with COVID-19. Preprint at *medRxiv* <https://doi.org/10.1101/2021.02.22.21252171> (2021).
49. Higginson, I. J. & Costantini, M. Accuracy of prognosis estimates by four palliative care teams: a prospective cohort study. *BMC Palliat. Care* **1**, 1 (2020).
50. Cheon, S. et al. The accuracy of clinicians' predictions of survival in advanced cancer: a review. *Ann. Palliat. Med.* **5**, 22–29 (2016).
51. White, N., Reid, F., Harris, A., Harries, P. & Stone, P. A systematic review of predictions of survival in palliative care: how accurate are clinicians and who are the experts? *PLoS ONE* **11**, e0161407 (2016).
52. Hibberd, C. S. & Quan, G. M. Accuracy of preoperative scoring systems for the prognostication and treatment of patients with spinal metastases. *Int. Sch. Res. Notices* **2017**, 1320684 (2017).
53. Schwartz, L. M., Woloshin, S. & Welch, H. G. Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. *N. Engl. J. Med.* **341**, 279–283 (1999).
54. Sharma, R., Mitra, A. & Stano, M. Insurance, race/ethnicity, and sex in the search for a new physician. *Econ. Lett.* **137**, 150–153 (2015).
55. Olah, M. E., Gaisano, G. & Hwang, S. W. The effect of socioeconomic status on access to primary care: an audit study. *CMAJ* **185**, E263–E269 (2013).
56. Si, Y., Zhou, Z., Su, M. & Chen, X. Revisiting gender gap in quality of health care in urban China: a standardised patient audit study. *Lancet* **394**, S25 (2019).
57. Kugelmass, H. 'Sorry, I'm not accepting new patients': an audit study of access to mental health care. *J. Health Soc. Behav.* **57**, 168–183 (2016).
58. Hagiwara, N. et al. Racial attitudes, physician–patient talk time ratio, and adherence in racially discordant medical interactions. *Soc. Sci. Med.* **87**, 123–131 (2013).
59. Hagiwara, N., Slatcher, R. B., Eggly, S. & Penner, L. A. Physician racial bias and word use during racially discordant medical interactions. *Health Commun.* **32**, 401–408 (2017).
60. Street, R. L. Jr, Gordon, H. & Haidet, P. Physicians' communication and perceptions of patients: is it how they look, how they talk, or is it just the doctor? *Soc. Sci. Med.* **65**, 586–598 (2007).
61. Bernheim, S. M., Ross, J. S., Krumholz, H. M. & Bradley, E. H. Influence of patients' socioeconomic status on clinical management decisions: a qualitative study. *Ann. Fam. Med.* **6**, 53–59 (2008).
62. Franks, P. & Fiscella, K. Effect of patient socioeconomic status on physician profiles for prevention, disease management, and diagnostic testing costs. *Med. Care* **40**, 717–724 (2002).
63. O'Malley, M. S. et al. The association of race/ethnicity, socioeconomic status, and physician recommendation for mammography: who gets the message about breast cancer screening? *Am. J. Public Health* **91**, 49–54 (2001).
64. Greene, M. G., Adelman, R., Charon, R. & Hoffman, S. Ageism in the medical encounter: an exploratory study of the doctor–elderly patient relationship. *Lang. Commun.* **6**, 113–124 (1986).
65. Ayalon, L. & Tesch-Römer, C. *Contemporary Perspectives on Ageism* (Springer Nature, 2018).
66. Robb, C., Chen, H. & Haley, W. E. Ageism in mental health and health care: a critical review. *J. Clin. Geropsychol.* **8**, 1–12 (2002).
67. Butler, R. N. Ageism: a foreword. *J. Soc. Issues* **36**, 8–11 (1980).
68. Gunderson, A., Tomkowiak, J., Menachemi, N. & Brooks, R. Rural physicians' attitudes toward the elderly: evidence of ageism? *Qual. Manage. Healthc.* **14**, 167–176 (2005).
69. Curlin, F. A., Lantos, J. D., Roach, C. J., Sellergren, S. A. & Chin, M. H. Religious characteristics of US physicians. *J. Gen. Intern. Med.* **20**, 629–634 (2005).
70. Daaleman, T. P. & Nease, D. E. Jr Patient attitudes regarding physician inquiry into spiritual and religious issues. *J. Fam. Pract.* **39**, 564–569 (1994).

71. Assouline, M., Gilad, S. & Ben-Nun Bloom, P. Discrimination of minority welfare claimants in the real world: the effect of implicit prejudice. *J. Public Adm. Res. Theory* **32**, 75–96 (2021).
72. Jilke, S., Van Dooren, W. & Rys, S. Discrimination and administrative burden in public service markets: does a public–private difference exist? *J. Public Adm. Res. Theory* **28**, 423–439 (2018).
73. Dehon, E. et al. A systematic review of the impact of physician implicit racial bias on clinical decision making. *Acad. Emerg. Med.* **24**, 895–904 (2017).
74. Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* **74**, 1464–1480 (1998).
75. Johnson, T. J. et al. Comparison of physician implicit racial bias toward adults versus children. *Acad. Pediatr.* **17**, 120–126 (2017).
76. Blair, I. V. et al. Clinicians' implicit ethnic/racial bias and perceptions of care among black and Latino patients. *Ann. Fam. Med.* **11**, 43–52 (2013).
77. Johnson, T. J. et al. The impact of cognitive stressors in the emergency department on physician implicit racial bias. *Acad. Emerg. Med.* **23**, 297–305 (2016).
78. Blanton, H., Jaccard, J., Gonzales, P. M. & Christie, C. Decoding the implicit association test: implications for criterion prediction. *J. Exp. Soc. Psychol.* **42**, 192–212 (2006).
79. Tetlock, P. E. & Mitchell, G. Implicit bias and accountability systems: what must organizations do to prevent discrimination? *Res. Organ. Behav.* **29**, 3–38 (2009).
80. Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J. & Tetlock, P. E. Using the IAT to predict ethnic and racial discrimination: small effect sizes of unknown societal significance. *J. Pers. Soc. Psychol.* **108**, 562–571 (2015).
81. Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J. & Tetlock, P. E. Predicting ethnic and racial discrimination: a meta-analysis of IAT criterion studies. *J. Pers. Soc. Psychol.* **105**, 171–192 (2013).
82. Blanton, H., Jaccard, J., Strauts, E., Mitchell, G. & Tetlock, P. E. Toward a meaningful metric of implicit prejudice. *J. Appl. Psychol.* **100**, 1468–1481 (2015).
83. Blanton, H., Jaccard, J. & Burrows, C. N. Implications of the implicit association test d-transformation for psychological assessment. *Assessment* **22**, 429–440 (2015).
84. Blanton, H. et al. Strong claims and weak evidence: reassessing the predictive validity of the IAT. *J. Appl. Psychol.* **94**, 567–582 (2009).
85. Arkes, H. R. & Tetlock, P. E. Attributions of implicit prejudice, or 'would Jesse Jackson 'fail' the implicit association test?'. *Psychol. Inq.* **15**, 257–278 (2004).
86. Ottaway, S. A., Hayden, D. C. & Oakes, M. A. Implicit attitudes and racism: effects of word familiarity and frequency on the implicit association test. *Soc. Cogn.* **19**, 97–144 (2001).
87. Knox, D., Lowe, W. & Mummolo, J. Administrative records mask racially biased policing. *Am. Polit. Sci. Rev.* **114**, 619–637 (2020).
88. Knox, D. & Mummolo, J. Making inferences about racial disparities in police violence. *Proc. Natl Acad. Sci. USA* **117**, 1261–1262 (2020).
89. Burgess, D., Van Ryn, M., Dovidio, J. & Saha, S. Reducing racial bias among health care providers: lessons from social–cognitive psychology. *J. Gen. Intern. Med.* **22**, 882–887 (2007).
90. Ferguson, W. J. & Candib, L. M. Culture, language, and the doctor–patient relationship. *FMCH Publ. Presentations* **61**, 353–361 (2002).
91. Currie, J., Lin, W. & Meng, J. Addressing antibiotic abuse in China: an experimental audit study. *J. Dev. Econ.* **110**, 39–51 (2014).
92. Kalkan, K. O., Layman, G. C. & Uslaner, E. M. 'Sands of others'? Attitudes toward Muslims in contemporary American society. *J. Polit.* **71**, 847–862 (2009).
93. Karpowitz, C. F., Monson, J. Q. & Patterson, K. D. Who's in and who's out: the politics of religious norms. *Polit. Relig.* **9**, 508–536 (2016).
94. Lajevardi, N. *Outsiders at Home: The Politics of American Islamophobia* (Cambridge Univ. Press, 2020).
95. Hainmueller, J., Hangartner, D. & Yamamoto, T. Validating vignette and conjoint survey experiments against real-world behavior. *Proc. Natl Acad. Sci. USA* **112**, 2395–2400 (2015).
96. Hainmueller, J. & Hopkins, D. J. The hidden American immigration consensus: a conjoint analysis of attitudes toward immigrants. *Am. J. Polit. Sci.* **59**, 529–548 (2015).
97. Jenke, L., Bansak, K., Hainmueller, J. & Hangartner, D. Using eye-tracking to understand decision-making in conjoint experiments. *Polit. Anal.* **29**, 75–101 (2021).
98. Ding, M., Grewal, R. & Liechty, J. Incentive-aligned conjoint analysis. *J. Mark. Res.* **42**, 67–82 (2005).
99. Adams, S. A., Font, X. & Stanford, D. All aboard the corporate socially and environmentally responsible cruise ship: a conjoint analysis of consumer choices. *Worldw. Hosp. Tour. Themes* **9**, 31–43 (2017).
100. Bansak, K., Hainmueller, J., Hopkins, D. J. & Yamamoto, T. Beyond the breaking point? Survey satisficing in conjoint experiments. *Polit. Sci. Res. Methods* **9**, 53–71 (2021).
101. Bansak, K., Hainmueller, J., Hopkins, D. J. & Yamamoto, T. The number of choice tasks and survey satisficing in conjoint experiments. *Polit. Anal.* **26**, 112–119 (2018).
102. Auerbach, A. M. & Thachil, T. How clients select brokers: competition and choice in India's slums. *Am. Polit. Sci. Rev.* **112**, 775–791 (2018).
103. Druckman, J. N. & Green, D. P. *Cambridge Handbook of Advances in Experimental Political Science* (Cambridge Univ. Press, 2021).
104. Westfall, P. & Young, S. *Resampling-Based Multiple Testing: Examples and Methods for P-value Adjustment* Vol. 279 (John Wiley and Sons, 1993).
105. Elwert, F. & Winship, C. Endogenous selection bias: the problem of conditioning on a collider variable. *Annu. Rev. Sociol.* **40**, 31–53 (2014).
106. Greenland, S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* **14**, 300–306 (2003).
107. Cole, S. R. et al. Illustrating bias due to conditioning on a collider. *Int. J. Epidemiol.* **39**, 417–420 (2010).
108. Montgomery, J. M., Nyhan, B. & Torres, M. How conditioning on posttreatment variables can ruin your experiment and what to do about it. *Am. J. Polit. Sci.* **62**, 760–775 (2018).
109. Zhao, Q., Keele, L. J., Small, D. S. & Joffe, M. M. A note on post-treatment selection in studying racial discrimination in policing. *Am. Polit. Sci. Rev.* (in the press).
110. Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. & Westwood, S. J. The origins and consequences of affective polarization in the United States. *Annu. Rev. Polit. Sci.* **22**, 129–146 (2019).
111. Jensen, C. & Petersen, M. B. The deservingness heuristic and the politics of health care. *Am. J. Polit. Sci.* **61**, 68–83 (2017).
112. Hartman, E. & Hidalgo, F. D. An equivalence approach to balance and placebo tests. *Am. J. Polit. Sci.* **62**, 1000–1013 (2018).
113. Lakens, D., Scheel, A. M. & Isager, P. M. Equivalence testing for psychological research: a tutorial. *Adv. Methods Pract. Psychol. Sci.* **1**, 259–269 (2018).
114. Lakens, D. Equivalence tests: a practical primer for *t* tests, correlations, and meta-analyses. *Soc. Psychol. Pers. Sci.* **8**, 355–362 (2017).
115. Schild, L. et al. 'Go eat a bat, Chang!': an early look on the emergence of sinophobic behavior on web communities in the face of COVID-19. Preprint at *arXiv* <https://doi.org/10.1145/3442381.3450024> (2021).
116. Mummolo, J. & Peterson, E. Demand effects in survey experiments: an empirical assessment. *Am. Polit. Sci. Rev.* **113**, 517–529 (2019).
117. Healy, A. & Malhotra, N. Retrospective voting reconsidered. *Annu. Rev. Polit. Sci.* **16**, 285–306 (2013).
118. Morgan, K. J. & Campbell, A. L. *The Delegated Welfare State: Medicare, Markets, and the Governance of Social Policy* (Oxford Univ. Press, 2011).
119. Mettler, S. *The Submerged State: How Invisible Government Policies Undermine American Democracy* (Univ. of Chicago Press, 2011).
120. Clinton, J. D. & Sances, M. W. The politics of policy: the initial mass political effects of Medicaid expansion in the States. *Am. Polit. Sci. Rev.* **112**, 167–185 (2018).
121. Hersh, E. D. & Goldenberg, M. N. Democratic and Republican physicians provide different care on politicized health issues. *Proc. Natl Acad. Sci. USA* **113**, 11811–11816 (2016).
122. Butler, D. M., Volden, C., Dynes, A. M. & Shor, B. Ideology, learning, and policy diffusion: experimental evidence. *Am. J. Polit. Sci.* **61**, 37–49 (2017).
123. Edwards, C. L., Fillingim, R. B. & Keefe, F. Race, ethnicity and pain. *Pain* **94**, 133–137 (2001).
124. Van Ryn, M., Burgess, D., Malat, J. & Griffin, J. Physicians' perceptions of patients' social and behavioral characteristics and race disparities in treatment recommendations for men with coronary artery disease. *Am. J. Public Health* **96**, 351–357 (2006).
125. Dominitz, J. A., Samsa, G. P., Landsman, P. & Provenzale, D. Race, treatment, and survival among colorectal carcinoma patients in an equal-access medical system. *Cancer* **82**, 2312–2320 (1998).
126. Penner, L. A. et al. The effects of oncologist implicit racial bias in racially discordant oncology interactions. *J. Clin. Oncol.* **34**, 2874 (2016).
127. Schwab, S. Is statistical discrimination efficient? *Am. Econ. Rev.* **76**, 228–234 (1986).
128. Altonji, J. G. & Pierret, C. R. Employer learning and statistical discrimination. *Q. J. Econ.* **116**, 313–350 (2001).
129. Guryan, J. & Charles, K. K. Taste-based or statistical discrimination: the economics of discrimination returns to its roots. *Econ. J.* **123**, F417–F432 (2013).
130. Gaddis, S. M. How black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies. *Sociol. Sci.* **4**, 469–489 (2017).
131. Sen, M. & Omar, W. Race as a bundle of sticks: designs that estimate effects of seemingly immutable characteristics. *Annu. Rev. Polit. Sci.* **19**, 499–522 (2016).

Acknowledgements

Financial support came from the National Science Foundation (grant no. SES-1657821, J.H.). The funders had no role in study design, data collection and analysis, decision to

publish or preparation of the manuscript. We thank D. Butler and A. Dynes for providing access to the contact information for local municipal officials. We thank the Brigham Young University College of Family Home and Social Sciences. We acknowledge feedback from scholars at Brigham Young University; Dartmouth College; the University of Virginia; Waseda University; the Japanese Society for Quantitative Political Science; the Japanese Association for Political Economy; the Research Institute of Economy, Trade, and Industry; and the Tokyo Foundation for Policy Research. We also thank (in alphabetical order) L. Argyle, L. Baldez, J. Barabas, M. Barber, R. Block, M. Costa, J. Druckman, M. Golder, S. Golder, J. Goodliffe, J. Gubler, M. Hansen, D. Hawkins, M. Herron, Y. Horiuchi, J. Houle, R. Johnson, C. Karpowitz, Y. Kobayashi, C. Krewson, D. Lacy, M. Lipscomb, D. Magleby, C. Mahoney, I. Mbiti, B. Nyhan, J. Preece, I. Solomon, S. Tello-Trillo, C. Volden and T. Yamamoto for their feedback.

Author contributions

C.C., J.B.H. and J.Q.M. all contributed equally to this paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41562-021-01280-9>.

Correspondence and requests for materials should be addressed to John B. Holbein.

Peer review information *Nature Human Behaviour* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We run three conjoint experiments on samples of the public, elected officials, and physicians that test for various social and demographic biases in the prescription patterns of ventilators to hypothetical patients.
Research sample	Random sample of registered voters who responded to our survey, population of physicians who responded to our survey, population of state and local officials who responded to our survey.
Sampling strategy	Random sample of 5,000 people per state from the voter registration lists. Population of physicians. Population of state and local officials.
Data collection	Conjoint experiments were embedded in a survey distributed through Qualtrics.
Timing	The public and politician recruitment occurred in the Spring of 2020, while the recruitment of physicians occurred in the Fall of 2020 as listed in the paper.
Data exclusions	List-wise deletion was employed. Otherwise all data were included as described in the manuscript.
Non-participation	The overall response rate for the public was 2.0%. The response rate for elected officials was 5.7%. The response rate among physicians was 1%.
Randomization	All of the characteristics included in the conjoint experiment were fully randomized across conditions.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	See above
Recruitment	All were invited to take a survey as outlined in the Manuscript and Supporting Informaton
Ethics oversight	Our project was approved by the Brigham Young University Institutional Review Board (E17512 and X19052)

Note that full information on the approval of the study protocol must also be provided in the manuscript.