Measuring Bias Against Doctors in a Patient Analog Experiment (Lucid) (#12916)

Created: 07/25/2018 12:54 PM (PT)
Shared: 09/22/2018 06:29 PM (PT)

1) Have any data been collected for this study already?
No, no data have been collected for this study yet.

2) What's the main question being asked or hypothesis being tested in this study?
Whether the race and gender of a simulated doctor causes participants in an analogue patient experiment to give lower satisfaction ratings and report less confidence in their diagnosis. The main (null) hypothesis being tested is that there are no differences across four treatment arms that randomize the race and gender of a putative doctor.

3) Describe the key dependent variable(s) specifying how they will be measured.
Primary outcome measures:
• How confident are you that this doctor made the correct diagnosis? [not at all confident (1) to completely confident (5)]
• How confident are you that this doctor recommended the correct treatment plan? [not at all confident (1) to completely confident (5)]
• Which diagnosis do you think is more likely to be correct? [the doctor's diagnosis (1); online symptom checker (2)]
• Would you ask the doctor to perform additional diagnostic tests? (Such as the CT scan recommended by the Symptom Checker). [definitely not (1); probably not (2); might or might not (3); probably (4); definitely (5)]
2) Satisfaction.
• What number would you use to rate your care during this emergency room visit? [0 (worse possible care) to 10 (best possible care)]
• Would you recommend this doctor to your friends and family? [definitely not (1); probably not (2); might or might not (3); probably (4); definitely (5)]
Secondary outcome measures:
1) Warmth and Competence of Doctor (6-item warmth-competence scale)
2) Willingness to file a complaint. Willing to sue if a diagnostic mistake was made [both: definitely not (1); probably not (2); might or might not (3); probably (4); definitely (5)]

4) How many and which conditions will participants be assigned to?
This is a 2x2 experimental design. Subjects will be randomly assigned to one of four possible conditions: Black Female, Black Male, White Female, White Male. Within each condition, subjects are randomly assigned 1 of 10 possible putative doctors. We have a total of 40 putative doctors, 10 for each condition. For example, 10 Black Female doctors, etc. These are, in fact, actors selected from the Chicago Face Database. Given that the vast majority of Emergency Physicians in the United States are white men, the White Males condition serves as the "control".

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
For all enumerated outcome variables, we will use a linear regression of the outcome on treatment assignment (with White Male as the omitted category), with robust standard errors. In addition to the basic "difference in means" estimator, we will also conduct regression analyses with covariate adjustment. In addition to standard demographic covariates, we also include a measure of trust in doctors, self-assessed health, insurance coverage status, recent emergency department visit experience, the gender and race of the respondent's primary care doctor (if applicable), and whether subjects have unpaid medical bills. Previous results indicated that trust in doctors is the most prognostic covariate.
We also plan to conduct the following sub-group analyses to estimate conditional average treatment effects:
1) By sex (male or female).
2) Age (individuals born before 1965 versus those born in 1965 or after)
3) Race (black or white)
4) Prejudice. We have a 4-item measure of racial prejudice (Stereotype Endorsement) and an 5-item measure of gender prejudice (from the Ambivalent Sexism Inventory)
We also plan to conduct exploratory analyses of treatment effect heterogeneity using Bayesian Additive Regression Trees (BART).

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We exclude respondents who are not in the United States. The following exclusion criteria are also applied: 1) subjects who are currently pregnant are ineligible; 2) subjects who have had cancer or currently have cancer are ineligible; 3) subjects who have had abdominal surgery are ineligible.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We will collect 1600 observations in total for a target of 0.80 power to estimate a Minimum Detectable Effect (MDE) of 0.10 standard units for the

primary outcome variables. The MDE calculations are based on the results from the previous experiment.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

This is a replication of a study conducted on MTurk (AsPredicted #9068). As in previous study, all exclusion criteria are applied *before* the study begins so that excluded respondents are deemed ineligible to participate before any study data are collected.

AsPredicted
Pre-Registration made easy

Measuring Bias Against Doctors in a Patient Analog Experiment (#9068)

Created: 03/13/2018 04:42 PM (PT)
Shared: 09/29/2018 12:03 PM (PT)

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
The main question is whether the race and gender of a simulated doctor causes participants in an analogue patient experiment to give lower satisfaction ratings and report less confidence in their diagnosis. The main (null) hypothesis being tested is that there are no differences in confidence or satisfaction ratings across four treatment arms that randomize the race and gender of a putative doctor.

3) Describe the key dependent variable(s) specifying how they will be measured.
• Primary outcome measures:
1) Confidence in doctor's diagnosis:
• We will combine the following visual analogue scales into an index (sum and divide by 2):
– How confident are you that this doctor made the correct diagnosis? [0-100]
– How confident are you that this doctor recommended the correct treatment plan? [0-100]
• Which diagnosis do you think is more likely to be correct? [the doctor's diagnosis (1); online symptom checker (2)]
• Would you ask the doctor to perform additional diagnostic tests? (Such as the CT scan recommended by the Symptom Checker). [yes (1); no (0)]
2) Satisfaction.
• What number would you use to rate your care during this emergency room visit? [0-100]
• Would you recommend this doctor to your friends and family? [5-point likert scale]
• Secondary outcome measures:
1) Warmth and Competence of Doctor (7-item warmth-competence scale)
2) Fairness of the cost of an ER visit.
• You would be charged about $350 for this emergency department visit. How fair do you think this charge is? [0-100]

4) How many and which conditions will participants be assigned to?
This is a 2x2 experimental design. Subjects will be randomly assigned to one of four possible conditions: Black Female, Black Male, White Female, White Male. Within each condition, subjects are randomly assigned 1 of 10 possible putative doctors. We have a total of 40 putative doctors, 10 for each condition. For example, 10 Black Female doctors, etc. These are, in fact, actors selected from the Chicago Face Database. Given that the vast majority of Emergency Physicians in the United States are white men, the White Males condition serves as the "control".

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
For all enumerated outcome variables, we will use a linear regression of the outcome on treatment assignment (with White Male as the omitted category), with robust standard errors. In addition to the basic "difference in means" estimator, we will also conduct regression analyses with covariate adjustment. In addition to standard demographic covariates, we also include a trust in doctor's measure, self-assessed health, insurance coverage status, recent emergency department visit experience, the gender and race of the respondent's primary care doctor (if applicable), and whether subjects have unpaid medical bills. Preliminary results indicated that trust in doctors is the most prognostic covariate for the outcomes.
We also plan to conduct the following sub-group analyses to estimate conditional average treatment effects:
1) By sex (male or female).
2) Age (individuals born before 1965 versus those born in 1965 or after)
3) Race (black or white), but it's unlikely that we will successfully recruit many black respondents given the demographic characteristics of the MTurk subject pool.
4) Prejudice. We have a 4-item measure of racial prejudice (Stereotype Endorsement) and a 4-item measure of gender prejudice (Hostile Sexism).
We also plan to conduct exploratory analyses of treatment effect heterogeneity using Bayesian Additive Regression Trees (BART).

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
We exclude MTurk respondents who are not in the United States. The following exclusion criteria are also applied: 1) subjects who are currently pregnant are ineligible; 2) subjects who have had cancer or currently have cancer are ineligible; 3) subjects who have had abdominal surgery are ineligible.

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We will collect 1600 observations in total. This was approximated based on the estimates from the pilot study, with a target of 0.80 power.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

We conducted a pilot study (n = 198) to confirm Javascript code is working as expected, and to estimate completion time for pricing of the full study. We have also written the main cleaning and analysis scripts that will be used for the study and tested them using these pilot data. These pilot data have also been used to make power calculations.