Pre-Analysis Plan: The Oakland Men's Health Disparities Study

Date: 3/21/2018

Principal Investigator: Marcella Alsan, Stanford University

Community Investigator: Owen Garrick, Bridge Clinical Research

Co-Investigator: Grant Graziani, University of California, Berkeley

Background

Racial disparities in health outcomes are a major policy concern in the United States. Across all population subgroups, black men experience earlier morbidity and mortality from preventable or manageable conditions. On average, black males live 4.2 fewer years relative to white males.[1] One potential solution to this problem is increasing minority representation in the health care workforce. Another is using financial incentives to encourage uptake of appropriate health care services. This research project has two main components — to examine the impact of patient-provider concordance and of financial incentives on the take-up of preventive health services by black men in the U.S.

Registration Timeline

The study was registered with the American Economic Association (AEA) on 7th October 2017. The trial entry can be found here: https://www.socialscienceregistry.org/trials/2497. On 21 March 2018 the PI uploaded a pre-analysis plan (PAP) that outlines the main hypotheses. At the time of submission, the project was transitioning all data from a secure cloud to a secure server and continuing data entry. The PI did not perform any analyses prior to the PAP being uploaded. Fieldwork recruitment commenced on 7th October 2017 with the clinic opening on 14th October 2017. The last day of the experiment was 20th January 2018. IRB clearance has been obtained from Stanford (#36255) and Berkeley (# 2016-02-8435) for this trial.

Experimental Design

The experimental design was as follows:

(1) Recruitment: Adult men who self-identified as African American were recruited at community sites in the Oakland, CA area. Most recruitment sites were black-owned barber shops, though some men were recruited from flea markets, churches and schools. Those who chose to enroll in the study completed a baseline survey regarding demographics, health status, history of health care utilization and trust. Survey completion was incentivized by haircut vouchers at the barbershops or $10–$20 in cash, depending on the venue. The participants received coupons for free cardiovascular health screenings to redeem at a clinic that was run on

[1] National Center on Health Statistics. "Life Expectancy by Race and Sex," 66(4). https://www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr66_04.pdf

Saturdays specifically for the purposes of the study. It was emphasized that the doctors would not be able to address other health concerns. Uber donated ride-sharing services to transport individuals to and from the clinic.

(2) Presentation and ex-ante demand: Subjects who presented at the clinic were randomized to a male study doctor and an incentive amount. After check-in, the subject was placed in a private patient room. At that point, the subject was given an unconditional cash payment for presentation at the clinic ($50). Afterwards, the subject was introduced to their assigned doctor through a tablet device, which included a generic description of the doctor's background and his actual photo. Subjects were then asked to select which of five preventative health care services they would like to receive from their assigned doctor. One of the services — a flu shot — was a surprise service the subjects learned about on presentation, at which point they were randomly offered a financial incentive ($0, $5, $10) conditional on selecting to receive the vaccination.

(3) Ex-post demand: After selecting the desired services, the assigned doctor met with the subject. Study doctors were provided with a list of each subject's ex-ante demand; however, they were instructed to encourage participants to obtain all of the services. The doctors were informed that, due to restrictions on the malpractice coverage purchased for the study, they were not able to provide services outside those in the screening. The ex-post demand (the demand after interacting with the assigned doctor) was also recorded. Subjects were asked to fill out a satisfaction survey after the encounter.

Specific Aims/Hypotheses Tested

The specific aims of the study are as follows:

H1.) To determine whether ex-ante demand for preventive health screening services, including an unanticipated influenza vaccination, is higher among black male subjects randomized to a racially concordant doctor.

H2.) To determine whether ex-post demand for preventive health screening services, including an unanticipated influenza vaccination, is higher among black male subjects randomized to a concordant doctor.

H3.) To determine whether subjects revise their choices after meeting with a concordant doctor (i.e. the difference between ex-post and ex-ante demand).

H4.) To assess whether the impact of having a concordant doctor and/or subsidy on the ex-ante take-up, ex-post take-up, and delta between ex-ante and ex-post take-up is greater among various groups with like characteristics. Potential heterogeneous effects are as follows:

A) The effects may differ for those who are more mistrustful of the medical system at baseline. This will be proxied for by baseline scores on medical mistrust questions.
B) The effects may differ for those already engaged with health care and/or those with major medical conditions versus those who are less engaged.
C) The effects may differ for those who have different demographic backgrounds, such as age, education, income and marital status.
D) The effects may differ for those who experienced greater hassle costs during the choice process (ex ante or ex post) — such as longer wait times or clinic

encounters at certain times of day, which may impact outcomes due to fatigue or hunger.[2]

E) The effects may differ for those who were recruited from certain locations, and for those who were told to come to recruitment venues through social networks vs. those who naturally encountered enumerators in the field.

H5.) To determine whether financial incentives increase take-up of the flu shot, and to assess whether incentives function as substitutes or complements to concordance in influencing demand for the flu shot.

H6.) To assess whether subjects report higher satisfaction when assigned to a concordant doctor or when assigned to a higher financial incentive.

H7.) To assess whether doctors spend a longer time with or field more questions from concordant patients.

H8.) To assess whether more nuanced categorizations of race and its constructs influence subject take-up.

A) To assess whether there are differential effects on take-up when black doctors are compared to different types of non-black doctors.[3]
B) To assess whether perceived race vs. self-reported race by doctors has differential effects on take-up.[4]

Because we are testing multiple hypotheses, we will use techniques that limit the false discovery rate.[5]

Econometric Specifications

Our main sample of interest will be the subjects who self-identify as black men, visit the clinic and meet with a study doctor (i.e. subjects that complete the Informed Consent, Baseline Field Survey, Clinic Tablet Survey, and the Clinic Encounter Form).[6] The primary test of interest is evaluating the effects of concordance between patient and physician on preventative health care choices. Our core specification evaluates treatment effects for **H1** and **H2** as follows:

[2] For instance, see work by Danziger, Levav, and Avnaim-Pesso (2011) demonstrating that judges are more lenient at different times of day (particularly after a food break). Danziger, S., J. Levav, and L. Avnaim-Pesso. 2011. "Extraneous Factors in Judicial Decisions," *Proceedings of the Natural Academy of Sciences*, 108(17): 6889–92.
[3] In pilot focus group work, black men ranked doctors (based on photos) they would be most likely to trust with a clear preference for concordant doctors, yet some other race/ethnicity groups ranked higher than others.
[4] In a companion survey, a separate sample of men that share characteristics with our study population are assessing photos of the study doctors.
[5] Benjamini, Y., A. Krieger, and D. Yekutieli. 2006. "Adaptive Linear Step-Up Procedures That Control the False Discovery Rate," *Biometrika*, 93: 491–507. Anderson, M. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedaian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103(484): 1481–95.
[6] If a subject changes identify (race/gender/ethnicity) after enrollment (i.e. they no longer self-identify as a black male) they will be excluded from the analysis.

$$Y_i = \alpha + \beta Concordance_{ij} + \sum_k \gamma^k Incentive_i^k + X_{ijc}'\Omega + \varepsilon_i, \qquad \text{Equation(1)}$$

where i is subject, j is doctor and c is clinic. Y is a measure of the choice of preventative services (see the "Outcome Measures" tab of the spreadsheet for further details). $Concordance$ measures whether the race of the doctor matches the race of the subject (i.e. black). Extensions of this definition use multi-dimensional measures of concordance combining race, age and education (see the "Concordance Measures" tab for further details). $Incentive$ is an indicator variable for the $5 or $10 amounts — the $0 category will be omitted. X, included in some specifications, will be a set of background characteristics on the subject as well as details on the clinic such as time of day, date, waiting room congestion, staff characteristics, and seasonality since all can influence the subject's clinical experience and potentially the return for some of the preventative care services.[7] Including these variables in the analysis could increase the precision of our treatment effect estimates.

To test **H3**, the differences in ex ante and ex post demand will be assessed according to the following specification:

$$\Delta Y_i = \alpha + \beta Concordance_{ij} + \sum_k \gamma^k Incentive_i^k + X_{ijc}'\Omega + \varepsilon_i, \quad \text{Equation(2)}$$

All variables are defined as in Equation (1) except the outcome of interest is now the difference between choices made in the ex ante and ex post periods.

To evaluate **H4**, X will be augmented to include interactions between the $Concordance$ measure and/or the $Incentive$ measure and baseline covariates obtained from subject surveys performed in the field. This modification tests whether effects are stronger for certain subgroups of the population. Alsan's prior work on the historical effect of the Tuskegee experiment suggests that the treatment might have a larger effect for groups that have lower education, lower income or higher mistrust at baseline.[8] For a complete list of heterogeneous effects and their particular pre-specified coding, please see the spreadsheet tab entitled "Heterogeneity." As noted in **H4D**, wait times could influence subject's engagement with the choices and the experiment overall. The study will collect various wait times in the clinical setting. The first wait time is from check-in at the door to check-in at the front desk (WT1). The second wait time will be from check-in at the desk to entry into a patient room (WT2). There will be a final time recorded when the subject exits the patient room after visiting with the doctor. The sum of the first two wait times measures time in the lobby. Wait times will be instrumented for by using the number of subjects who have checked into the clinic within a given time frame (i.e. 3 hours) and the number of doctors available at that clinic day. In addition to wait time from congestion, doctors vary in their time per patient. This can be instrumented for using the leave-one-out average wait time for a given

[7] Seasonality will influence flu since the flu shot might become less valuable over time as individuals have other opportunities to become vaccinated. On the other hand, if the flu season is particularly bad and is covered a lot in the media, demand could also increase over time.

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

doctor (omitting the index subject). Both wait times measures will be controlled for explicitly and interacted with the main treatment effects to assess how they influence take-up decisions.

To test **H5,** the incentive indicators will be interacted with concordance measures. If the marginal effect of incentives interacted with concordance is positive, this implies that incentives and concordance are complements. If, on the other hand, concordance interacted with incentives is negative and statistically significant, then that implies the two interventions are substitutes.

The specification to test **H6** is a modification of Equation (1) with the Y encompassing responses from the questionnaire that subjects completed after the visit with the study doctor. Similarly, **H7** replaces outcome measures with information collected by the doctor during the screening — including information on whether the subject asked them additional health questions and how long they spent with the subject. For **H8**, black doctors will be coded as an indicator separately from an indicator for white doctors and from an indicator specifically for Asian doctors. This will afford an examination of whether take-up differs across all three groups. Using a separate survey instrument on a sample of men who share study characteristics with the sample recruited from Oakland, perceptions of study doctors based on their photos will be collected. These data will allow for a test on whether self-assessed or perceived race matter more for the primary outcomes in **H1** and **H2**.

Additional Quality Checks

In addition to the evaluation plan detailed above and in the accompanying excel file, a series of standard quality checks will be performed on the data to ensure fidelity of the randomization and integrity of the design. First, a balance test will be performed to confirm that treatment (concordance or incentive amounts) does not predict baseline socio-demographic features of the subjects. Second, differential attrition across study arms will be assessed. Attrition in this context is defined as when subjects initiate a check-in at the clinic (and thus are randomized) but do not complete the study visit. Assuming that there is no differential attrition across arms, the main analytical sample will be those subjects who both came to the clinic and completed the visit with a study doctor. If there is evidence of selective attrition, then bounds will be used and/or the sample including those who attritted will be used as an alternative. Third, JPAL's replication service will be used prior to publication and pending IRB approval for the replication service to access the data.