

### Seminar 1

The questions relate to the following papers:

*Questions (1) to (6).* Alsan, Marcella and Owen Garrick and Grant Graziani (2019) “Does Diversity Matter for Health? Experimental Evidence from Oakland” American Economic Review 2019, 109(12): 4071–4111. [Video summary.](#) [Paper.](#)

*Questions (7) to (11).* Pons, Vincent (2018) “Will a Five Minute Discussion Change your Mind? A Countrywide Experiment on Voter Choice in France.” American Economic Review 2018, 108(6): 1322–1363. [Video summary.](#) [Paper.](#)

**Before the tutorial: read the introduction of at least one of the papers, and watch the videos summaries for both. You should also try to answer the questions below.**

- 1) Figure 1 below shows the study design for the Alsan et al (2019) paper. The authors decided to randomize the main treatment (being assigned to a black doctor) *after* observing which individuals redeem the clinic coupon and complete the questionnaire. Can you cite two good reasons why they do so?

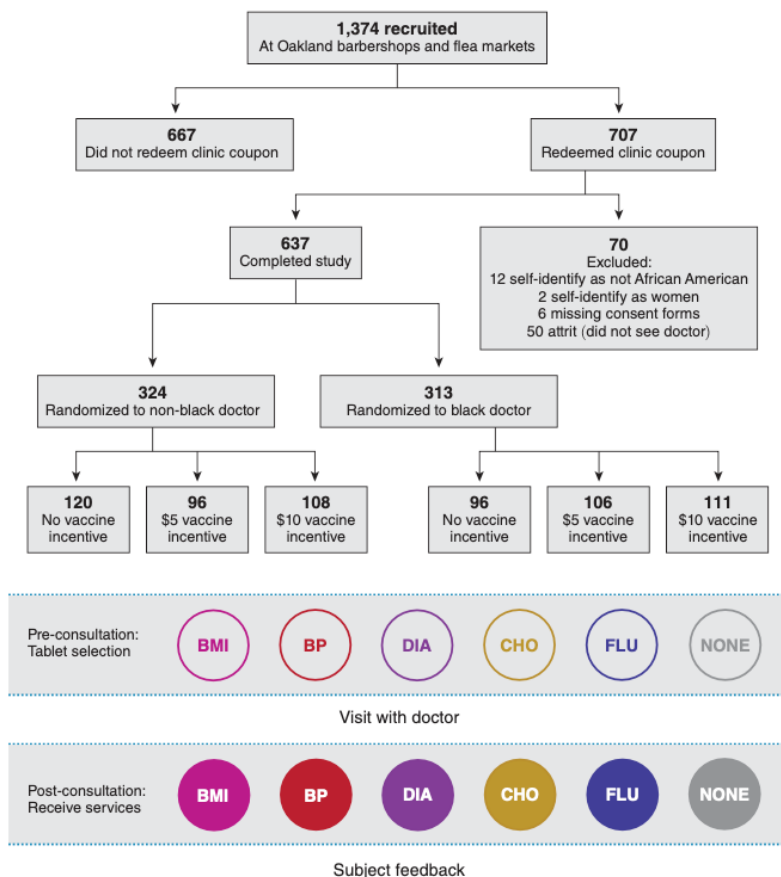


FIGURE 1. STUDY DESIGN AND FLOW

- 2) In Table 5 below, The outcome variable is the “delta  $\Delta$ ” i.e. difference for an individual between the post and pre- consultation demand for a specific test. Pre and post demand are 0/1 dummies. Why is it important, for the authors argument, to also analyse the effect of Age and Educational Proximity?

TABLE 5—DEMAND FOR PREVENTIVES WITH ALTERNATIVE CONCORDANCE MEASURES

X =	Age, 5 years		Age, 10 years		Education	
	(1)	(2)	(3)	(4)	(5)	(6)
X	0.008 {0.030}	-0.005 {0.039}	0.015 {0.026}	-0.019 {0.034}	0.002 {0.052}	-0.024 {0.098}
X × black doctor		0.008 {0.053}		0.037 {0.048}		-0.018 {0.118}
Black doctor		0.165 {0.051}		0.153 {0.059}		0.157 {0.057}
Observations	620	620	620	620	556	556

*Notes:* Table reports OLS estimates of equation (1). The outcome is the delta share of the invasive screenings. Columns 1 and 2 explore age concordance (i.e., doctor and subject born within 5 years of each other), columns 3 and 4 examine concordance within a wider age window (i.e., doctor and subject born within 10 years of each other), and columns 5 and 6 explore concordance across educational attainment (i.e., subject has at least a bachelor of arts degree). Indicators for incentive levels are included but not reported. Robust standard errors clustered at the doctor level in curly brackets.

- 3) Write down the regression equation for Column (2).

- 4) In interpreting column (2) select the correct answers. Note that the average of the outcome variable is 0.01.

a. For every additional 5 years of age of the doctor, there is a decrease of 0.5 percentage points of the demand for invasive tests.

b. Meeting a black doctor increases the demand for invasive tests, to a similar extent whether or not the age difference is larger than 5 years.

c. Meeting a black doctor increases the demand for invasive tests by 0.16%.

d. Conditional on pre-consultation demand, meeting a black doctor increases post-consultation demand by 0.16 percentage points.

e. Meeting a black doctor increases the difference between pre and post consultation demand by 16 percentage points.

d. There is no significant effect of Age proximity on demand for invasive tests.

- 5) Provide similar interpretations for column (6).

6) Food for thought (Extra): What policy recommendations would you formulate following this study ?

7) In Pons (2018), the author's randomization is at the precinct (voting district) level. One strategy could be to take all precincts in France and randomly assign them to control and treatment groups. Instead, the author randomizes precincts within territories (a territory is defined as a set of contiguous municipalities sharing a common zip code), and within categories of "potential win". Potential win is a variable that allows to classify precincts according to whether one can expect the door-to-door campaign to have an effect on vote share.

7a) How is such a randomization strategy called? What are its advantages, in general and in this particular context?

7b) As a researcher, how would you go about to compute the "potential win" variable?

8) The author describes the main specifications as follows:

First, I show the effect of a precinct being *assigned to the treatment group* (the intent-to-treat effect of the campaign), using the following OLS specification:

$$(1) \quad Y_i = \alpha_1 + \beta_1 T_i + X_i' \lambda_1 + \sum_s \delta_{i1}^s + \epsilon_{i1},$$

where  $Y_i$  is the outcome in precinct  $i$ ,  $T_i$  is a dummy equal to one if the precinct was assigned to the treatment group and zero if it was assigned to the control group,  $\delta_{i1}^s$  are strata fixed effects, and  $X_i$  is a vector of controls.

Secondly, I evaluate the effect of a precinct being actually *allocated to canvassers* (a local average treatment effect) with the following specification:

$$(2) \quad Y_i = \alpha_2 + \beta_2 A_i + X_i' \lambda_2 + \sum_s \delta_{i2}^s + \epsilon_{i2},$$

where  $A_i$  is a dummy equal to one if the precinct was allocated to the canvassers and zero otherwise, and is instrumented with  $T_i$  as shown in the following first-stage equation:

$$(3) \quad A_i = a + b T_i + X_i' \lambda + \sum_s \delta_i^s + \nu_i,^{15}$$

TABLE 3—FIRST STAGE

	No control	With controls					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.5652 (0.0137)	0.5233 (0.0173)	0.5247 (0.0172)	0.5238 (0.0172)	0.5241 (0.0173)	0.5245 (0.0172)	0.5240 (0.0172)
Strata fixed effects	X	X	X	X	X	X	X
Control for past outcome and PO		X	X	X	X	X	X
Additional controls		X	X	X	X	X	X
2007 outcome controlled for		Voter turnout, round 1	Voter turnout, round 2	Voter turnout, average	Vote share Royal, round 1	Vote share Royal, round 2	Vote share Royal, average
Observations	3,390	2,660	2,660	2,660	2,660	2,660	2,660
R <sup>2</sup>	0.258	0.424	0.423	0.424	0.423	0.424	0.424
Mean in control group	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: The table shows first stage results from equation (3). The unit of observation is the unit of randomization (precinct, or municipality). Robust standard errors are in parentheses. All regressions include strata fixed effects. Regressions in columns 2 through 7 control for PO (proxy for the potential to win votes) and for past outcomes, measured at the level of randomization: voter turnout or vote share obtained by Ségolène Royal in the first round, in the second round, or averaged over both rounds of the 2007 presidential election. Additional controls include the number of registered citizens in the precinct or municipality as well as the level and the five-year change of the following census variables: the municipality’s population, the share of men, the share of different age groups (from 0 to 14; from 15 to 29; from 30 to 44; from 45 to 59; from 60 to 74; above 75), the share of working population, and the share of unemployed population among the working population. Regressions controlling for past outcomes need to exclude precincts whose boundaries had changed after 2007, which explains the lower number of observations.

Column (1) of Table 3 shows results for equation (3). Complete the following sentence:

*The probability of having canvassers do a door-to-door campaign is \_\_\_\_\_ higher in precincts selected for treatment compared to control precincts. The effect is significant at the \_\_\_\_\_ level / insignificant. This corresponds to a situation with / without partial compliance. The identification of the causal effect of the door-to-door campaign would have been impossible if this effect had been \_\_\_\_\_ .*

9) The author does not actually estimate equation (2) with the  $A_i$  dummy. Instead,  $A_i$  is first estimated using equation (3), and the estimated  $\hat{A}_i$  is used in equation (2). What would be an issue with running regression (2) as such ?

10) Please refer to Table 5 below. The outcome variable is a dummy 0/1 for “Voting for Hollande”. Fill in the following sentences :

*In treatment areas, the door-to-door visits increased Hollande’s vote share by \_\_\_\_\_ in the First Round, which represents a \_\_\_\_\_% increase, and by \_\_\_\_\_ in the Second Round (when controlling for strata fixed effects, past outcome and additional controls). These effects are/are not significant at the 5% level. In areas that were actually allocated to canvasser (i.e., actually treated) the door-to-door visits increased the vote share by \_\_\_\_\_ in the First Round (when controlling for strata fixed effects, past outcome and additional controls).*

11) *Food for thought (Extra)*: As a researcher, what arguments would you have given the Socialist Party to convince them to work on this randomized evaluation with you? What could have been some of their concerns?

TABLE 5—IMPACT ON HOLLANDE’S VOTE SHARE

	First round			Second round			Average of first and second rounds		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Panel A. ITT estimation</i>									
Treatment	0.0063 (0.0023)	0.0050 (0.0019)	0.0044 (0.0018)	0.0048 (0.0028)	0.0053 (0.0019)	0.0046 (0.0018)	0.0056 (0.0024)	0.0049 (0.0017)	0.0043 (0.0016)
Strata fixed effects	X	X	X	X	X	X	X	X	X
Control for past outcome and PO		X	X		X	X		X	X
Additional controls			X			X			X
Observations	3,390	2,660	2,660	3,390	2,660	2,660	3,390	2,660	2,660
R <sup>2</sup>	0.003	0.516	0.528	0.001	0.632	0.645	0.002	0.645	0.655
Mean in control group	0.3157	0.2994	0.2994	0.5757	0.5597	0.5597	0.4457	0.4295	0.4295
<i>Panel B. Instrumental variable estimation: “allocated to canvassers” instrumented with “treatment”</i>									
Allocated to canvassers	0.0112 (0.0041)	0.0094 (0.0036)	0.0084 (0.0035)	0.0084 (0.0050)	0.0099 (0.0036)	0.0087 (0.0035)	0.0098 (0.0042)	0.0092 (0.0031)	0.0081 (0.0030)
Strata fixed effects	X	X	X	X	X	X	X	X	X
Control for past outcome and PO		X	X		X	X		X	X
Additional controls			X			X			X
Observations	3,390	2,660	2,660	3,390	2,660	2,660	3,390	2,660	2,660

*Notes:* Panel A shows the effect of a precinct being assigned to the treatment group (ITT results from equation (1)). Panel B shows the effect of a precinct being allocated to canvassers (2SLS results from equation (2)). The unit of observation is the unit of randomization (precinct, or municipality). Robust standard errors are in parentheses. All regressions include strata fixed effects. Regressions in columns 2, 5, and 8 also control for PO (proxy for the potential to win votes) and for past outcomes, measured at the level of randomization. Additional controls in columns 3, 6, and 9 include the number of registered citizens in the precinct or municipality as well as the level and the five-year change of the following census variables: the municipality’s population, the share of men, the share of different age groups (from 0 to 14; from 15 to 29; from 30 to 44; from 45 to 59; from 60 to 74; above 75), the share of working population, and the share of unemployed population among the working population. Regressions controlling for past outcomes need to exclude precincts whose boundaries had changed after 2007, which explains the lower number of observations.