

Pre analysis plan for analyses of effects of the Integrated Housing and Development Programme in Ethiopia

Asbjørn Goul Andersen, Tigabu Getahun, Andreas Kotsadam, Vincent Somville and Espen Villanger

November 20, 2018

Introduction

Assessing the causal impact of wealth on attitudes and well-being is a challenging task, since an individual's economic situation may depend on unobserved characteristics which are likely to be correlated with such outcomes. We overcome this challenge by exploiting the exogenous variation in wealth created by a housing lottery in Addis Ababa, Ethiopia, which randomly allocates the right to purchase an apartment among low-income applicants. Since the apartments are sold at highly subsidized rates, winners of the lottery experience a substantial boost to long-term wealth. We focus on the causal effects of winning the lottery on two separate domains: (i) changes in attitudes toward redistribution and meritocratic beliefs and (ii) economic and psychological well-being. This plan is registered before the data collection started. All deviations from the plan will be highlighted in the papers.

The lottery

The Integrated Housing and Development Programme (IHDP) aims at facilitating access to quality housing for low- and middle income groups in Addis Ababa, Ethiopia. The Addis Ababa Housing and Development Project Office (AAHDPO) is responsible for organizing and financing the construction of condominium apartments. The apartments are sold at highly subsidized prices, and home-buyers are given access to finance through the Commercial Bank of Ethiopia (CBE).

The Addis Ababa Housing Development and Administration Agency (AAHDAA) is responsible for allocating the apartments. Given the excess demand for housing at the subsidized prices, condominium apartments are allocated through a lottery among eligible registrants. The lottery is computer-based and carried out by the Information Network Security Agency (INSA). Eligibility is based on three requirements, (i) having resided in Addis Ababa for at least the previous 2 years, (ii) not having any other house or lease land registered (by one's own or the spouse's name), and (iii) having opened a savings account at the CBE and deposited the required monthly savings for at least 29 months (with no breaks longer than 6 months).

When registering for the program, applicants must select the desired apartment type (studio, one-, two-, or three-bedrooms). Since supply and demand vary across unit type, lotteries are held for each type. Within each lottery, quotas exist for women, disabled, and government employees. First, 30 percent of the winners are drawn among female applicants. Then 20 percent of the winners are drawn among government employees. Finally, there is a 5 percent quota for people with physical disabilities. All quotas were decided upon after registration.

As of today, there have been two rounds of registration and 12 lotteries. We focus exclusively on the first round of registration and the 11th lottery.

Sampling and data collection

We sample applicants who registered in the first round, in 2005, for a studio, a one-, or a two-bedroom apartment,¹ and who were eligible for the 11th lottery in 2016.

The Ethiopian Development Research Institute (EDRI) conducts the data collection. We have two types of lists with people: one for winners and one for losers. EDRI obtained the list of winners and losers from the 11th lottery (which took place in 2016) from the AAHDAA.

Starting with the winners, we randomly ranked the individuals on this list that had unique telephone numbers and that had not won a 3-bedroom apartment. All individuals on this list registered in 2005. We sample the 2000 individuals with the lowest numbers and call this list our “winners sample”. For the winners, we have information about the apartment type they won, the gender, and employment in the public sector at the time of the registration. We also have information about the location of the apartment that they won. We do not have information about physical disability so we have to ask them this separately.

EDRI also obtained a list of individuals who registered in 2005, and were qualified for the 11th lottery but who did not win it (and did not win the 12th lottery either). The list includes the type of apartment the individuals applied for. Unfortunately, the list does not contain information about gender and employment status but it includes physical disability status. Gender can, however, be predicted with very high precision from the first name of the applicant. We also ranked all individuals in this list randomly and then we selected a sample of 2000 individuals that matched the 2000 individuals in the winner list with respect to the proportion female within each apartment type. We call this list our “losers sample”.

We then take the winners and losers samples, add them together and randomize the order again. We create a new ID variable and keep only the ID, the names, and phone numbers of the people before sending the list to the data collection team at EDRI. This way the treatment status is blinded for the enumerators and we escape issues with confounding factors due to different timing and different enumerators. EDRI will interview the sampled individuals by phone using the survey questionnaire developed by the research team. They will stop when they have reached 3000 interviews.

¹ We have excluded applicants to three-bedroom apartments, since almost everyone in this group has received an apartment at the time of writing.

Coding of variables

In this section we present the variables that we will use in the analysis. We start with the main outcome variables and continue with the covariates and the variables used to study heterogeneous effects.

Main outcomes

We will measure the effect of winning the housing lottery on two separate domains and will most likely write about them in two separate papers. We want to know how winning the lottery affects people's *preference for redistribution*, their *beliefs about the causes of poverty*, as well as their *acceptance of inequality*. We also investigate how it impacts *life satisfaction and psychological well-being*. Despite probably being part of two different papers we write about them in the same plan and use all of them in correcting for multiple hypotheses testing as they are based on the same treatment.

I. Preference for redistribution and causes of poverty

Preference for redistribution

We measure the respondents' preferences for redistribution at three levels. At the more general level, we ask whether they agree that the government should intervene to reduce economic inequality. We then ask more specifically if the redistribution should be based on property taxation. Finally, to capture their preference toward privately funded redistribution, we give them the opportunity to share a monetary amount between themselves and an NGO (which we randomly assign to be either one supporting poor people or one supporting the elderly and disabled people).

The specific questions that we use are:

PR1. Please say whether you strongly disagree, disagree, agree or strongly agree with these statements:

In Ethiopia, the national government should aim to reduce the economic differences between the rich and the poor.

(1 = Strongly agree 2 = Agree 3 = Disagree 4 = Strongly disagree)

PR2. *In Ethiopia, the national government should have taxes on people owning houses to reduce the economic differences between the rich and the poor.*

(1 = Strongly agree 2 = Agree 3 = Disagree 4 = Strongly disagree)

PR3. As we stated in the start, you will be given 50 birr in airtime that we send to your phone. You are given the possibility to donate a share of this money to **X**. If you want to donate, we will send the money to the organization. Do you want to donate any of the airtime? If yes, how much? (5, 10, 15, 20, 25, 30, 35, 40, 45, 50).

X is randomly assigned at the individual level to be either “Mekodonia (a disabled and elderly association)” or “Mary Joy (an organization supporting poor people)”.

The first two variables will be coded as binary variables. The binary variable will be set in order to create 2 groups, as similar in size as possible, while respecting the order of the answers (from 1 to 4). For example, if 25% answer 1, 25% answer 2, 25% answer 3 and 25% answer 4, then the binary variable will be equal to 1 if the respondent answers 1 or 2 and zero otherwise. On the other hand, if for example 20% answer 1, 10% answer 2, 5% answer 3 and 65% answer 4, then the binary variable will be equal to one of the respondent answers 1, 2 or 3. We will only consider the answers of the respondents in the control group (the lottery losers) when creating the coding rules for the binary variables.

For the third variable, PR3, we will use the amount donated for those answering yes and replace the value by zero for those answering no. The variable will be treated as a continuous variable.

Beliefs about the causes of poverty and inequality acceptance

We want to know if the respondents believe that poverty is rather due to a poor character, or to an unfair environment. We use the following question:

CP1. Why, in your opinion, are there people in this country who live in need? Here are two opinions: Which comes closest to your view?

1. People are poor because of laziness and lack of will power
2. People are poor because of an unfair society

We create a dummy variable, CP1, that equals one if people answer 1 and zero if they answer 2.

We also create a measure of inequality acceptance based on the question:

IA1: Which opinion about inequality comes closest to your view?

1. Large differences in people's incomes are acceptable to properly reward differences in talents and efforts.
2. For a society to be fair, differences in people's standard of living should be small.

We create a dummy variable, IA1, that equals one if people answer 1 and zero if they answer 2.

II. Well-being

We will evaluate the effects of winning the lottery on two aspects of the people's well-being: their psychological well-being, and their general satisfaction with life.

Psychological well-being

To measure psychological well-being, we use the first ten items from the Kessler scale and aggregate them in one index (Kessler et al. 2002):

PWB1. *These questions concern how you have been feeling over the past 30 days. During the last 30 days, about how often did :*

1. *you feel tired out for no good reason?* | ____ |
2. *you feel nervous?* | ____ |
3. *you feel so nervous that nothing could calm you down?* | ____ |
4. *you feel hopeless?* | ____ |
5. *you feel restless or fidgety?* | ____ |
6. *you feel so restless you could not sit still?* | ____ |
7. *you feel depressed?* | ____ |
8. *you feel that everything was an effort?* | ____ |
9. *you feel so sad that nothing could cheer you up?* | ____ |
10. *you feel worthless?* | ____ |

(1 = None of the time 2 = A little of the time 3 = Some of the time 4 = Most of the time 5=All of the time)

The numbers attached to the 10 responses are added up and the total score is the score on the Kessler Psychological Distress Scale (K10). Scores will range from 10 to 50.

Life satisfaction

To measure life satisfaction, we use the general satisfaction measure from the World Value Survey:

LS1. *Using a scale from 0 to 10, where 0 means not at all satisfied, and 10 is completely satisfied, how satisfied are you with your life as a whole these days?*

We will standardize LS1 by subtracting the mean of the control group and dividing by the standard deviation of the control group.

Mechanisms

We will estimate the treatment effects on other variables in order to enrich our discussion of the plausible mechanisms behind the impacts that we will find on the main outcomes.

Winning the lottery implies a substantial increase in wealth. This change in wealth may have direct consequences on people's preferences for redistribution and well-being. We will start by checking that lottery winners are indeed wealthier than the losers. In addition, we estimate changes in their incomes and in the main expenditures related to housing.

Even if we observe increases in wealth, the increase may not be sufficient to change the people's perception about their relative economic position in society. That perception may, however, play an important role in shaping their preferences. We will therefore test whether winning the lottery also affects perceived relative position in society.

Moreover, even if the winners experience a substantial wealth increase, they may also face economic stress and liquidity constraints linked for example to the obligation to pay the mortgage. This may also affect our main outcome measures and we will use a set of financial distress indicators to capture this.

Another plausibly important factor behind people's preferences for redistribution is their beliefs about the causes of poverty. In particular, whether people believe that poverty and wealth are due to luck or to merit may influence their preferences. Since lottery winners experience a large increase in wealth, only because of luck, it is plausible that the lottery changes their beliefs about the causes of poverty. We therefore further measure and test the lottery's impact on the beliefs about the causes of poverty.

Wealth

Our first measure of wealth is the value of their properties. We ask everyone:

Do you own a house or an apartment? (0 = No, 1 = Yes)

If yes, how many houses/apartments do you own? _____

If yes, if you were to sell it, how much do you think you would get for it?

	Selling price of the house/apartment
House1	
House2	
House3	
House4	
House5	

In addition, to check the status of the winners and losers samples, we ask at the end of the interview:

We received your number from a list of participants in the low cost condominium housing lottery. Did you win the lottery?

If yes, we then asked if they have sold it or not. If they have sold it, we ask: *For how much did you sell it?*

If they didn't sell it, we ask: *If you have not sold the house/flat/condo, how much do you think you would get for it if you were to sell it?* And *Did you include this value when we asked you about all your houses before?*

We added those questions to both winners and losers to check that there were no mistakes in the lists, and to minimize reporting errors on the property value question.

The variable *housing related wealth* will then be coded as follows. The value is given by the answer to the first question about expected selling price of their houses (if they own any). If they did not include the value of the condominium in the first question, and we learn about it in the second question, then the value of the condominium is added. For the ones winning the lottery, we will also add the selling amount if they sold the apartment.

As we expect that most individuals will not have sold the flats that they won, these questions give us the respondents' evaluation of the current market value of their property and are intended to capture the lottery's impact on wealth. We believe that this is the relevant measure for our purposes as it answers how much more wealthy the winners think that they are. To avoid the potential heavy influence of extreme values, we will apply the hyperbolic inverse sine transformation to the values in *birr*.

The housing wealth variable is also interesting as winners are randomly assigned to houses in different areas, implying that the values of the apartments might differ. As a supplement to the binary treatment variable of winning the lottery, we will also rank winners according to the size of the wealth shock, they experienced. In that way, we can see if the size of the wealth shock is correlated with our main outcome variables.

This ranking can be based on different measures. The easiest is to use the selling prices or the expected selling prices that the respondents reported in the survey based on the questions above. This has the disadvantage of being subjective and subjective evaluations at the individual level of prices are likely correlated with many other things we cannot observe. A better measure is to take the jackknifed values (i.e., the values of all other except the respondent) in an area for the same type of apartment. This will work well in larger areas but not in areas with only one respondent. We will also try to get access to external measures of apartment values in different areas.

House ownership is an important but not exclusive source of wealth. It may be that those winning the lottery invest in other real estate or in land. We ask all respondents if they own land and real estate, and we ask them to estimate the value of these assets. We will also evaluate their current wealth by asking:

How much debt does your household have in total?

How much savings does your household have in cash?

How much savings does your household have in the bank?

A total measure of *net wealth* will be created by summing up housing wealth (not including the selling price if they sold it), other real estate or land, savings in cash and in the bank minus debt. We will apply the hyperbolic inverse sine transformation to the values of wealth, debt and savings in birr.

We will also complement these wealth measures with more qualitative variables.

We will first ask them to evaluate their economic trajectory:

Richer_plus_5: Do you think you will be richer or poorer in 5 years from now? We will code them as *Richer_plus_5* =1 if they answer richer and zero if the answer poorer.

Richer_minus_5: Are you richer than 5 years ago? (Yes=1, No=0)

We will also construct an asset index based on the principal component analysis of the following items:

Does your household have the following... (Yes/No)?

1. *Radio*
2. *Television*
3. *Refrigerator*
4. *Car*
5. *Computer*
6. *Tablet*
7. *Satellite dish*
8. *Smartphone*
9. *Electric mitad*

The asset index will be standardized by subtracting the mean of the control group and dividing by the standard deviation of the control group.

Income and expenses

We will sum the self-reported incomes of the respondent and of his or her partner from the following sources:

- Wage employment
- Rental income
- Self-employment
- Remittances
- Government or NGO transfers
- Pension
- Other income sources

In our estimations, we will use the inverse hyperbolic sine transformation of this sum.

We will also estimate the treatment impacts on the expenditures related to housing:

- Rents paid
- Mortgage repayments
- Repayments of other debts

Again, we will use the inverse hyperbolic sine transformation of the sum of these categories.

Perceived economic position in society

We use the following question to measure perceived economic position:

Think about the people Ethiopia in general. Do you think you are richer, equally rich, or poorer than the majority of them? (1 = Richer, 2 = Equally rich, 3 = Poorer)

We will transform the answers into a binary variable equal to one if they answer “Richer”. The answer “Equally rich” will be pooled with “richer” if, in the control group, fewer people answer “richer” than “poorer”, otherwise it will be pooled with “poorer”.

Financial distress

To measure financial distress we will use the following questions:

- 1) *If you suddenly ended up in an unforeseen situation, where you have to raise 20,000 Birr, would you be able to? (Yes/No)*

- 2) *In the past six months, has your family had inadequate money to cope with the family expenses? (Never, Rarely, Sometimes, Always)*
- 3) *In the past six months, has your family delayed the payment of bills because of financial difficulty? (Never, Rarely, Sometimes, Always)*
- 4) *What has been the economic condition of your family in the past six months? (No financial difficulty, some financial difficulty, Considerable financial difficulty, much difficulty)*

The first variable will be coded as a binary indicator equal to one if the answer is No. We will create binary variables for each of the potential answers to the second, third and fourth questions by splitting them in a way that retains the ordering while it minimizes the difference in number of observations between the two categories based on the answers in the control group.

We will also measure their economic satisfaction with the question:

How satisfied or dissatisfied are you with your personal economy?

(0 = Very dissatisfied, 1 = Rather dissatisfied, 2 = Somewhat dissatisfied, 3 = Somewhat satisfied, 4 = Rather satisfied, 5 = Very satisfied 6 = Not applicable)

We will transform this into a binary variable by splitting it in a way that retains the ordering while it minimizes the difference in number of observations between the two categories based on the answers in the control group, as we do for PR1. The answer “6=Not applicable” will be coded as missing

Beliefs about the causes of poverty

The main outcome CP1 gives us a first insight into the respondents’ beliefs about the causes of poverty. We will use additional questions to enrich this discussion and evaluate whether the impacts on preferences for redistribution can be due to changes in the beliefs about the causes of poverty.

We elicit their beliefs about the importance of different factors acting at four levels: individual, structural at the family level, structural at the society level, or simply fatalistic.

The initial question is in all cases:

In your opinion, to what degree do each of the following factors currently cause people to become poor? (0 = To a small degree 1 = To a large degree)

Individual factors:

- Lack of ability of competence
- Poor character

- Lack of individual effort
- Lack of ambition

Institutional factors at the family level:

- Growing up in a poor family
- Having poorly educated parents

Institutional factors at the society level:

- Biases or discrimination in society
- Lack of equal opportunity in society

Fatalism:

- Bad luck

Even though these variables can be placed in different families, they are also quite different. We will therefore present a table with regression results where we run a regression of each variable separately

Strata and covariates

When estimating the impacts of winning the lottery we will control for the strata that are used in the lottery:

S1. A binary variable equal to one for female applicants. This is from the administrative register for winners and coded from names for the losers. We update the information for the losers with the enumerator coding of the respondent gender during the interview and they ask at the end of the interview if they are unsure.

S2. A binary variable equal to one for public sector employees. This is from the administrative register for winners and based on the following question for the losers: *What was your occupation in 2005 (at time of housing registration)*. We code this as one if they answer “Government” and zero otherwise.

S3. A set of binary variables indicating which type of housing the applicant applied for (i.e., a studio, a one-, or a two-bedroom apartment). This is from the administrative registers for both winners and losers.

S4. A binary variable equal to one for people with physical disabilities. This is from the administrative register for the losers but for the winners it is based on the following question: *Did you have any physical disability at time of registration(2005)?*

Ex-ante, we believe that including some basic demographic characteristics may improve the precision of our estimates. We will present the estimated coefficients with and without the following covariates:

C1. Age in years.

C2. A set of binary variables indicating the applicant's ethnicity. We will have separate dummies for the ethnicities that comprise at least 5 percent of the observations and put all other in a separate variable ("other ethnicity") which will be the excluded category.

C3. A set of binary variables indicating the applicant's religion. We will have separate dummies for the religions that comprise at least 5 percent of the observations and put all other in a separate and put all other in a separate variable ("other religion") which will be the excluded category.

C4. A set of binary variables indicating the applicant's place of birth. We will have separate dummies for all places that comprise at least 5 percent of the observations and put all other in a separate variable ("other birth place") which will be the excluded category.

C5. Questions about earnings: We ask both winners and losers the following questions:

What was your earnings per month at the time of the registration in 2005. That is, the earnings you reported on the registration form in 2005?

What was the earnings of your partner, if you had one at the time, per month at the time of the registration in 2005?

What was your earnings per month in 2015 (i.e. three years ago).

What was the earnings of your partner per month in 2015 (i.e. 3 years ago), if you had one at the time?

We will create a measure of own earnings and of couple earnings and in any estimation we will use the inverse hyperbolic sine transformation.

C6. : *Did you have a spouse or partner that you lived with at the time of registration?*

We will also see if we can improve precision in the estimates by picking optimal controls from the total list of controls using LASSO (Belloni et al. 2014; Ahrens et al. 2018).

Heterogeneity

We will use machine learning techniques to automate the search for heterogeneous treatment effects. There are many different types of machine learning algorithms and we have decided to use classification and regression trees (R package `causalTree`, Athey and Imbens 2016) and random forests (R package `grf`, Wager and Athey 2017). As this field is moving rapidly, however, it is possible that there will be other techniques that are relevant for us once we start analyzing the data.

Of particular interest in the heterogeneity analysis is the previous earnings. We would expect that winning the lottery has different effects for high and low baseline earners. We will therefore also add a specification where we interact high (at or above the median) pre-earnings (in 2015) with treatment, controlling for all the other covariates (except earnings in 2005).

Another interesting variable for the heterogeneity analysis is the randomly assigned NGO. We will test whether treatment effects differ for people assigned the two different NGOs.

Other exploratory outcomes

Our survey measures many other outcomes that we will use in the exploratory analysis, to complement our main findings. In particular, we will measure the lottery's impacts on the family composition and the children education.

We will test if the lottery changes:

- A binary variable equal to one if the respondent is single
- A binary variable equal to one if the respondent has been engaged in any income generating activities during the last 12 months as well as binary variables for what type of activity (Farming, own business, manufacturing, construction, service, public sector, NGO).
- A binary variable equal to one if the respondent was the primary breadwinner of the household during the last six months
- A binary variable equal to one if the respondent answers "Most people can be trusted" on the trust question.
- Years of completed education of their partner
- Total income of their partner
- Number of children born after the lottery
- The years of education that they expect their children to complete. For this outcome we will have one observation per child and we will restrict ourselves to children born before the lottery.

We will also include measures for the other domain specific life satisfaction questions. In particular we use the question: *Now follow some questions on how satisfied or dissatisfied you are with some different areas of your life. How satisfied or dissatisfied are you with. . .*

your health?

your leisure time?

your friends?

your relatives?

the home that you live in?

the neighborhood that you live in?

Ethiopian society?

your work? (Not working=6)

These other questions will be coded in the same way as we code satisfaction with income (see above). The question about their work will only be estimated for people working.

We will also measure gender equality and domestic violence with the following questions: *Please say whether you strongly disagree, disagree, agree or strongly agree with these statements...*

1. *It is okay for women to work outside of the home*
2. *It is okay for women to earn more money than their partners*
3. *A husband justified in beating his wife if she neglects the children*

We will transform the three answers into binary variable following the same procedure as for PR1.

Other descriptive measures

The survey also includes a series of questions that will be important to describe the general behavior of winners and losers and their views about the lottery. We will use these questions for descriptive purposes.

In particular we will use the question asked to all about whether they perceive the lottery to be fair and transparent. We will use questions asked to the losers about their perceived chances of winning in the future and whether they know someone who won. Interesting questions to the winners are if they have moved in to the flat and whether/when they plan to do so (or why not).

Estimation strategy

The following equations will be estimated using ordinary least squares estimators.

We first regress the outcome of interest on treatment status; i.e. a dummy variable equal to one if the individual has won the lottery, and a set of fixed effects, S_i :

$$(1) \quad Y_i = \alpha + \beta_1 T_i + g S_i + \varepsilon_i$$

The variables in the S vector are the covariates in $S1$ to $S4$. Furthermore, we will also estimate equation (1) with a set of covariates X :

$$(2) \quad Y_i = \alpha + \beta_1 T_i + \beta_2 X_i + g S_i + \varepsilon_i$$

The vector X comprises the covariates in $C1$ to $C6$.

We will also estimate the same type of regressions but instead of T have the continuous measures of housing wealth and focus only on winners. Our main specification will, however, be with the binary treatment variable and with the strata variables but without other covariates, unless there is imbalance across winners and losers. To test for balance we will regress our main treatment variable on the control variables described above both individually and together, while controlling for the Strata fixed effects S (gender, public sector, disabled, and type of apartment). We will judge whether the randomization worked by conducting an F-test of whether the control variables jointly predict treatment status.

Given that the lottery (randomization) is made at the individual level, we will not cluster the standard errors.

Power

We will interview 3000 individuals, with around 1500 winners and 1500 losers. At the conventional level of significance of 0.05 and a power of 0.8, our sample size would allow for a minimum detectable effect of 0.1024 standard deviations. These calculations do not take into account the potential gains in precision from including the covariates in the estimation.

We will also adjust the p-values for the fact that we are testing the impact on seven outcomes. We follow the recommendations of Fink, McConnell, and Vollmer, (2014) and use a method developed by Benjamini and Hochberg, (1995) and Benjamini and Yekutieli, (2001) to minimize the false non-discovery rate. The main advantage of the method is that it is limiting the risk of false discoveries while only adjusting the critical values based on other true hypotheses. The false discovery rate method developed by Benjamini and Hochberg (1995) implies that the m p-values of the i hypotheses are ordered from low to high and that the critical value of the p-value is then $p(i) = \alpha \cdot i/m$. To illustrate, with 7 hypotheses and a significance level (α) of 0.05, the critical p-value would be 0.0071 for the one with the lowest p-value ($0.05 \cdot 1/7$, which is the same as a Bonferroni

correction). For the second hypothesis, the critical p-value is 0.014 ($0.05 \times 2/7$) and for the seventh it is 0.05 ($0.05 \times 7/7$). The minimum detectable effect after accounting for multiple hypothesis testing is 0.13 standard deviations.

Addressing survey attrition and non-response

We will probably not manage to reach all the respondents initially sampled. We will check whether non-response is correlated with winning the lottery. If there is a statistically significant difference in non-response between winners and losers (controlling for the $\mathbf{S}$ vector), we will follow Kling, Liebman and Katz (2007)'s correction. We will obtain lower bounds of the treatment effect by replacing missing observations in the treatment (control) arms by the corresponding arm's mean value minus (plus) 0.05, 0.10 and 0.20 standard deviations of the control group. Upper bounds of the treatment effects will be constructed in a symmetrical way.

No imputation for missing data from item non-response will be performed. We will check whether item non-response is correlated with treatment status following the same procedures as for survey attrition, and if it is, construct bounds for our treatment estimates in the same manner.

Variables with limited variation

To limit noise caused by variables with minimal variation, questions for which 95 percent of observations have the same value within the relevant sample will be omitted from the analysis and will not be included in any indicators or hypothesis tests. In the event that omission decisions result in the exclusion of all constituent variables for an indicator, the indicator will not be calculated. If this happens for one of our main outcomes we will not use it as a main outcome and we will not adjust for multiple testing for that variable.

Contingency plan if a new lottery is held

There is a 13th round of the lottery that will take place sometime soon. It is unclear when this round will happen and how large it will be. If the lottery happens in a time and fashion that transforms a substantial share (>10 percent) of the non-interviewed losers to be winners we will pause the data collection and draw new losers.

Archive

The pre-analysis plan is archived before any data is collected. We archive it at the registry for randomized controlled trials in economics held by The American Economic Association: <https://www.socialscienceregistry.org/> on November 20 2018. We will start data collection on November 21 2018. We expect to get the data in January.

References

- Ahrens, A., Hansen, C. B., & Schaffer, M. (2018). PDSLASSO: Stata module for or post-selection and post-regularization OLS or IV estimation and inference.
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2), 608-650.
- Benjamin, Daniel J., James O. Berger, Magnus Johannesson, Brian A. Nosek, E.-J. Wagenmakers, Richard Berk, Kenneth A. Bollen, et al. 2018. "Redefine Statistical Significance." *Nature Human Behaviour* 2 (1). Nature Publishing Group: 6–10.
- Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." In: *Journal of the royal statistical society. Series B (Methodological)*, pp. 289–300.
- Benjamini, Yoav and Daniel Yekutieli (2001). "The control of the false discovery rate in multiple testing under dependency." In: *Annals of statistics*, pp. 1165–1188
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1), C1-C68.
- Fink, Günther, Margaret McConnell, and Sebastian Vollmer (2014). "Testing for heterogeneous treatment effects in experimental data: false discovery risks and correction procedures." In: *Journal of Development Effectiveness* 6.1, pp. 44–57.
- Kessler, R.C., Andrews, G., Colpe, .et al. 2002. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, 32, 959-956.
- Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1).