Pre-Analysis Plan for Perceived Discrimination and Integration

Bernt Bratsberg (Frisch Centre for Economic Research)

Henning Finseraas (Norwegian University of Science and Technology)

Jeremy Ferwerda (Dartmouth College)

Karen Hauge (Frisch Centre for Economic Research)

Andreas Kotsadam (Frisch Centre for Economic Research)

Abstract

We estimate the causal relationship between perceptions of discrimination and sense of inclusion for refugees in Norway by experimentally providing information about the extent of discrimination. This pre-plan describes how we will analyze the data from the information treatments and all deviations will be noted in the paper.

Background and motivation

Refugees' sense of social integration will be low if discrimination and/or perceived discrimination is widespread (see e.g. Oskooi 2020). In this study we use unique Norwegian survey data combined with administrative data to examine determinants of perceptions as well as determinants of integration. We will further estimate the causal relationship between perceptions of discrimination and sense of inclusion. In the study, which targets the full population of UN refugees in Norway, we map personal experiences with discrimination and experimentally provide information treatments on the extent of discrimination against minorities, and study how these variables shape sense of belonging and exclusion. This pre-plan describes how we will analyze the data from the information treatments.

The survey experiment

Respondents are randomized into two groups:

Group A receives the question preamble "Recent Norwegian studies show that *employers and landlords discriminate against applicants with foreign-sounding names*". The treatment is indicated in italics.

Group B receives the question preamble "Recent Norwegian studies show that *most employers and landlords do not discriminate against applicants with foreign-sounding names.* The treatment is again indicated in italics.

Outcomes measured after the treatment

To measure whether the treatment have an effect on perceptions of discrimination, respondents are asked:

- How widespread do you think discrimination is in Norway?

The respondents are asked to choose between 4-categories: Very widespread, somewhat widespread, not very widespread, not widespread at all. We standardize the variable to have mean 0 and standard deviation 1.

Our second main outcome is an index of belonging. Respondents are asked three questions on their sense of belonging:

- To what degree would you like to have more social contact with native Norwegians?
- If you think about your social life, work situation, and access to public services, to what degree do you feel excluded in Norway?
- How connected do you feel with Norway?

For the first two statements, the respondents are asked to choose between 5-categories, ranging from 1-5: To a very large degree, To a large degree, Somewhat, To a small degree, Not at all. For the third statement the categories also range from 1-5 and are: I do not feel a connection at all, I feel a weak connection, I feel a moderately close connection, I feel a very close connection, I feel an extremely close connection. We reverse code the answers to the first statement and then create an additive index, which we standardize to have mean 0 and standard deviation 1.

We measure behavior by asking whether respondents would like to donate to The Norwegian Center against Racism. The survey question reads:

After the survey, the research project organizing this survey will donate NOK 10,000 to a charitable cause. All respondents, including you, can vote for which organization will receive the donation. The charity that gets the most votes will receive a donation of NOK 10,000. Which organization do you give your vote to?

Alternative 1 Norwegian Centre Against Racism (Antirasistisk senter)

Alternative 2 Norwegian Refugee Council (Flyktninghjelpen)

Alternative 3 Norwegian Cancer Society (Kreftforeningen)

Alternative 4 Norwegian Rescue Dogs (Norske redningshunder)

We code the variable: *Donation to the Norwegian Center against Racism* to equal one if they vote for alternative 1 and zero otherwise.

We measure preferences for segregated living by answering 2 on the question:

Suppose you were choosing where to live. Which of the three types of areas would you ideally wish to live in?
o An area where almost everyone were born in Norway 1

o An area where almost everyone were born outside of Norway 2
o An area where about half were born in Norway and half were born outside Norway 3
o It would make no difference where people were born 4

We measure trust in Norwegians based on the following survey question:

On a score of 0-10 how much you personally trust each of these types of individuals. 0 means you do not trust an individual at all, and 10 means you have complete trust.
… People born in Norway?

Hypotheses

We propose the following primary hypotheses:

H1: Respondents assigned to group A will report that discrimination is more widespread than respondents assigned to group B.

H2: Respondents assigned to group A will report a lower sense of belonging than respondents assigned to group B.

H3: Respondents assigned to group A will be more likely to vote for a donation against racism than respondents assigned to group B.

H4: Respondents assigned to group A will be more likely to have a preference for segregated living than respondents assigned to group B.

H5: Respondents assigned to group A will report lower trust in Norwegians than respondents assigned to group B.

Sample and data collection

The survey experiment is included in a survey questionnaire that is distributed to 31,600 UN refugees settled in Norway after 1990 and that are at least 18 years old in 2023. The survey is administered by Statistics Norway. The expected number of respondents is 5000-8000.

Randomization

Respondents will be randomized to Groups A or B using simple randomization (equal probabilities of assignment), as build into the software used by Statistics Norway.

Identification of treatment effects

We will test the hypotheses in a regression framework using OLS. We use the following main specification:

$$Y_{i2} = \beta Treated\_A_i + \chi X_{i1} + \varepsilon_i \qquad (1)$$

where i indexes individuals, and t is time (t_2 then implies after the treatment). We create indicators for assignment to Group A and use Group B as the reference group. β gives the effects of assignment to treatment A versus B. X_{i1} is a vector of controls (further discussed below). To make the models fully saturated, we partition the covariate space and add control variables as indicator variables rather than using their multi-valued codings (Athey and Imbens, 2017). If cells are too small, with less than 5 percent of the observations, adjacent cells are combined. When using interaction terms and in tests of balance we will retain the continuous coding of the variables (they will be standardized for interaction terms). If we have missing values on explanatory variables we will code the variables as zero and include dummy variables controlling for missing status so that we do not lose observations. We use robust standard errors.

We will present results from estimations without control variables but our main estimation will be one with optimal controls being chosen from the total list of controls using a post-double LASSO selection approach of Belloni et al. (2014). The LASSO selection approach selects those variables that are correlated with both treatment and the outcomes, which may improve precision in the estimates, and it helps to correct for observed imbalances across groups.

We will not apply survey weights as Miratrix et al. (2018) show that weighting is typically not necessary for unbiased estimates when one analyzes survey experiments embedded in large, broadly representative samples, while applying them might result in substantial loss of statistical power.

Balance checks

We will examine balance across gender (indicator for male, from register data), age (continuous variable, from register data), education (binary indicator of whether the respondent answers that they have completed education at the university/university college

level before arriving to Norway), region (one dummy indicator for each region from the register data: Oslo/Østlandet [Eastern Norway]; Sørlandet/Vestlandet [Southern and Western Norway]; Trøndelag/Nord-Norge [Mid- and Northern Norway]), perceptions about discrimination in Norway before the treatment (continuous variable)[1], self-reported experience with discrimination (binary indicator for answering yes to the question "Have you ever been treated worse than others in Norway because of your ethnic background, religion, or skin color?"), Trust in people (continuous variable based on the question "To what extent do you trust people that you do not know?") and a continuous scale index of societal integration (see next section for coding).

We will present means for all variables by treatment group and use an F-test of joint significance to test balance.

Coding of control variables

As our main specification is the post-double LASSO selection model, there is no limit to how many variables we can include as controls and we let the final set of control variables be chosen by the procedure. In addition to the variables described in the balance section, we will also include:

The subcomponents of the societal integration index. The societal integration index is an index of questions that tap into different dimensions of societal integration. Our index is derived from a subset of questions from Harder et al. (2018).

The first dimension is psychological integration, which consists of the questions *connected* ("How *connected* do you feel to Norway?"), *excluded* ("Thinking about your social life, work situation, and access to public services, to what degree do you feel excluded in Norway?"), and *future* ("In the future, do you want to live in Norway?").

The second dimension is linguistic integration, which consists of the statements *reading* ("Thinking about your reading skills in Norwegian, how well can you read and understand the

[1] The question about how widespread discrimination is is also asked early on in the survey, before the treatment, so that we can see how perceptions are updated and measure effects based on the degree and direction of updating. Asking the same question twice in one survey runs the risk of anchoring effects but experimental evidence suggest that repeated measures designs tend to yield the same results as a design that split responses across survey waves (Clifford et al. 2021).

main points in newspaper articles on familiar topics?") and *speaking* ("Thinking about your Norwegian language skills, how well can you speak about familiar topics and express personal opinions in a conversation?").

The third dimension is political integration, which consists of the questions regarding political knowledge: *false* (Answering false on: "Rødt is considered to be right of Arbeiderpartiet"), *Prime minister* (Answering Arbeiderpartiet on: "To which of the following parties does the current *Prime minister* of Norway belong?)" , and *participate* (Answering yes to: "Did you vote in the last municipal election?").

The fourth dimension is social integration which consists of the questions *dinner* ("In the last 12 months, how often have you eaten a meal with Norwegian-born people who are not part of your family?"), *friends* (the share of Norwegian born friends, "How many of your close friends were born in Norway?" divided by "How many individuals do you consider close friends?", *voluntary work* (sum of the items local sports club, organization for cultural activity, organization for the protection of the environment, organization for human rights/humanitarian work). All questions use the original scale, recoded to the 0-1 range. If a respondent has a valid response to at least one component variable of an index, then missing values for other component measures are imputed as the mean of that variable.

For each of the four dimensions, we run a principal factor analysis restricted to produce one factor, using the survey items described above as the input. After deriving the four factor scores we run a new principal factor analysis, again restricted to produce one factor, using the four factor scores as the input. The predicted factor scores for the final factor analysis constitute the Societal integration index.

From the survey data we will also include the following control variables: *Employed in the home country* ("Before you arrived in Norway, were you ever employed by an employer in a paid job?"). *Minority in the home country* ("Were you part of a minority group in your home country?"), *Politically active in the home country* ("Were you politically active (such as voting or being a member of a political party or organization) before you fled your home country?"), and *Done well* ("If you compare yourself to other people, how well would you say that you have done in life?").

From the register data we additionally have the following variables we control for: Employment last year, occupation last year, number of years in employment, earnings, years since arrival, completed education (including also in Norway), and share of immigrants in the neighborhood.

Treatment heterogeneity and other exploratory analyses

The baseline beliefs about discrimination may play an important role in the information treatment analysis since there may be heterogeneous effects depending on the direction of updating (Haaland et al. 2023). This is important as we would expect different effects based on whether the prior overestimated or underestimated the degree of discrimination as compared to the posterior signal received. We will therefore also interact the treatments with baseline beliefs about the extent of discrimination. We will conduct heterogeneity analyses based on social integration before the experiment as well.

We will also explore treatment heterogeneity across all baseline variables described in the "Balance checks" and "Control variables" sections. The treatment heterogeneity analysis will be conducted using the honest causal forest method (Wager and Athey 2018). It is likely that we will add other variables from both the survey and the register data to test different outcomes as well as other heterogeneity. This will be clearly labelled as exploratory.

We will further explore determinants of social integration, economic integration, and perceptions of discrimination all as measured before the treatment. In doing this we will use random forest estimation with all variables. We are particularly interested in whether the same variables are important for all of these outcomes and in comparing the variable importance in the descriptive analyses with the heterogeneity analysis in the experimental analysis.

Multiple hypothesis adjustments

We will present conventional as well as adjusted critical p-values using the false discovery rate (FDR) method of Benjamini and Hochberg (1995). The number of main hypotheses is restricted to 5 (see above).

Power

Conservatively assuming a response rate of only 15 percent, the survey data will include 4740 individuals, which would make our analyses well powered to detect relatively small effects. With 5 hypotheses and using the adjusted FDR p-values we will be able to detect relatively small effects of less than 0.1 of a standard deviation of the outcome. The minimum detectable effect if we only manage to get 3000 responses is still 0.13 with five hypotheses. These are small effects, and more well powered than the power suggested by Haaland et al. (2023), whom propose to have power to identify an effect of 0.15 of a standard deviation in information provision experiments.

Some of our findings are likely to be null results. It is often difficult to judge whether such results are showing a meaningful lack of effect or whether they arise due to low power. To investigate if the effects are meaningful null findings we will conduct equivalence tests with two one-sided t-tests (TOST) and show how large positive and negative effects we can reject. The tests are one sided in equivalence testing as one tests whether effects are larger than a highest value and lower than a lowest value. In practice, the procedure is equivalent to presenting the bounds of a 90 percent confidence interval.

Missing data

We will examine whether missing outcomes are correlated with treatment group. We will use the treatment effect equation to conduct the test. If treatment status correlates with missing outcomes, we will calculate extreme bounds and trimming bounds for the treatment effect for the always-reporters (see Gerber and Green 2012: 226ff).

If a respondent has a valid response to at least one component variable of the sense of belonging index, then missing values for other component measures are imputed as the mean of the random assignment group (see Kling et al. 2007).

Tests of differential selection into the survey

Differential selection into the survey may reduce the external validity of our experimental results. Since the response rate is assumed to be low, it is very unclear how the results would be if we had conducted the experiment with the full population of quota refugees in Norway. The issue is further aggravated for the descriptive analyses of the determinants of social integration and perceived discrimination, as well as for the analyses of moderating factors in the experiment. The reason is that we expect factors that correlate with e.g. social integration also correlate with unobserved factors that in turn correlate with the probability of answering the survey. As such, it is even difficult to interpret a simple correlation such as one between education and integration. Say that we were to find that education level is positively correlated with our index of social integration. Since education level and social integration are likely correlated with unmeasured factors that make respondents more likely to answer the survey, our survey responders will be a biased representation of the full population. This goes beyond a classic omitted variables bias as it may also introduce a mediation bias. The individuals with low education that answer the survey are likely different in other ways from other low educated individuals, since they answer the survey they probably have some other trait that is correlated with propensity to answer *and* with social integration. We stress that the internal validity of the main experimental results are not in danger, however, since treatment is randomized within the pool of respondents. We here also outline some ways in which we will investigate and try to correct for biases in the other analyses induced by selective participation.

Access to the whole population in the register data enables us to closely investigate non-representativeness induced by non-response. We will first compare the survey respondents to the quota refugee population in total.

We will additionally vary incentives and reminders to experimentally induce random variation in response rates in order to investigate the importance of response rates for the different outcomes. We also randomly vary the mode of contact and we will test how response rates affect responses by testing for differential results along the chain of reminders. We will also keep track of how easy it is to get the respondents to answer among those that answered. Assuming that there is a correlation between hard to reach respondents and non-respondents, investigating hard to reach respondents may teach us something about the selection problem induced by incomplete survey data.

Ethical considerations

The study is approved by the data protection officer at the Frisch Centre as well as by the data protection officer at statistics Norway.

References

Athey, S, and G. Imbens (2017). "The State of Applied Econometrics: Causality and Policy Evaluation" Journal of Economic Perspectives 31(2): 3-32.

Belloni, A., V. Chernozhukov, and C. Hansen (2014): "Inference on Treatment Effects After Selection Amongst High-Dimensional Controls," Review of Economic Studies, 81, 608–650.

Benjamini Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)* 57(1):289–300.

Clifford, S., Sheagley, G., & Piston, S. (2021). Increasing precision without altering treatment effects: Repeated measures designs in survey experiments. *American Political Science Review*, *115*(3), 1048-1065.

Gerber, A. S. and D. P. Green. 2012. *Field Experiments*. New York: Norton.

Haaland, I., Roth, C., & Wohlfart, J. (2023). Designing information provision experiments. *Journal of economic literature*, *61*(1), 3-40.

Haaland, I., & Roth, C. (2023). Beliefs about racial discrimination and support for pro-black policies. *Review of Economics and Statistics*, 1-15.

Harder, N., Figueroa, L., Gillum, R. M., Hangartner, D., Laitin, D. D., & Hainmueller, J. (2018). Multidimensional measure of immigrant integration. Proceedings of the National Academy of Sciences, 115(45), 11483-11488

Kling, J.R., J.B. Liebman, and L.F. Katz 2007. Experimental Analysis of Neighborhood Effects. *Econometrica* 75(1): 83-119.

Miratrix, L., J. S. Sekhoon, A. G. Theodoridis, and L.F. Campos 2018. Worth Weighting? How to Think About and Use Weights in Survey Experiments. *Political Analysis* 26(3): 275-291.

Oskooii, Kassra. 2020. "Perceived Discrimination and Political Behavior." British Journal of Political Science 50(3): 867–892.

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228-1242.