Pre-Analysis Plan:

# Self-Help Online program for Ukrainian refugees

ISRCTN96282810: https://doi.org/10.1186/ISRCTN96282810

Date of the latest version: 18 July 2022

## 1. General Overview

The study wants to test if the self-help online program helps adult people to cope with high-stress situations. The program focused on Ukrainian citizens, but do not exclude the people with other citizenship who were living in Ukraine. The program is not open to people living outside Ukraine and European Union. This program aims to help people to improve their psychological well-being and reduce stress. The results of the study will help to understand if a self-help online program can be a solution to help many people across borders cope with high-stress situations. Here we provide a pre-analysis plan that is formulated right **after** the baseline data collection but **before** any endline data collection.

## 2. Status of the program

The program was advertised for two weeks among Ukrainian in Ukraine and Europe from the 22nd of June to the 6th of July 2022. During this recruitment period, access to the materials was closed. 1201 people show interest in the program by registering for the program, actively consenting to participate in the study, and filling out the baseline survey before the deadline.[1] We observe the balance on observables across treatment groups:

- Treatment group: 90-minute self-help course and extensive information material about the support offered to Ukrainians in Ukraine and other countries.
- Control Group: only have access to the information material.

After the program has opened registration was closed. The participants can learn their treatment status only when they log in on the platform after the opening of the program (7th of July). On 15th of July, we observe ~480 people started to participate in the program by leaving any identifiable response on the platform after the start of the program on 7th of July. We consider among those who left or will leave any identifiable response on the platform after the start of the program on 7th of July and eligible for the program as participants of the program – a sample of the main analysis for which this pre-analysis plan is provided.

[1] We observe 1 201 registrations with a complete baseline survey out of which 35 were located neither in Ukraine nor European Union, and 1 double registration. That is, our baseline survey sample consists of 1 164 eligible participants.

## 3. HYPOTHESES TO BE TESTED

### A. Primary outcome

*Hypothesis 1:* Participants who receive the self-help online course show a lower level of stress compared to the control group participants.

Outcome variable: Kessler 6 (Kessler, 2010) on the scale from 0 to 24 at the endline survey.

### B. Secondary outcome

*Hypothesis 2:* Participants who receive the online self-help course show a higher level of psychological well-being compared to the control group participants.

Outcome variables:

- WHO-5 (Topp et al., 2015) on a scale from 0 to 25 at the endline survey.
- MHI-5 (Veit et al., 1983) on a scale from 0 to 100, summing the scores of each question item and then transforming the raw scores as suggested in Ware (1993).

## 4. General Estimation Strategy

We will use the next AN**C**OVA specification for participant *i* to estimate the effect of being offered the Self-Help Online Course or not:

$$Y_{i,endline} = \beta_0 + \beta_T T_i + Y_{i,\text{baseline}} + controls_i + \varepsilon_i, \qquad (1)$$

In this specification $Y_{i,endline}$ is the variable measures the outcome of interest for participant *i* at the endline. The variable $T_i$ is the dummy variable that equals 1 if the participant was assigned to receive the Self-Help Online Course and 0 if this participant was assigned not to receive the course (information only). $Y_{i,\text{j,baseline}}$ - the variable measures the outcome of interest for the participant at the baseline. We will apply the post-double-selection Lasso approach of Belloni et al. (2014) to select control variables from baseline information (see Appendix A). $\varepsilon_i$ is the error term.

## 5. Exploratory analysis.

### 5.1 Life-related outcome.

Given the short period between baseline and endline survey, we state the next exploratory hypothesis:

*Hypothesis 3:* Participants who receive the online self-help course show better results on the life-related outcome, variables compared to the control group participants.

Outcome variable: Average of standardized z-scores of answers on the next questions:

1. "How many hours in the past week did you spend searching for jobs?" (Winsorized)
2. "How many hours in the past week did you spend working (for money)?" (Winsorized)
3. "How many hours in the past week did you spend volunteering (doing any work for free)?" (Winsorized)
4. "How many hours in the past week did you spend searching and applying for educational options (university, courses, ect.)?" (Winsorized)
5. "Are you currently learning any language?"

### 5.2 Heterogeneous Treatment Effects

We will use three strategies to understand heterogeneous treatment effects.

A. We want to assess heterogeneous treatment effects for the Participant Status at the baseline:
    a. Not Displaced, Located in Ukraine
    b. Not Displaced, Located Abroad (Migrants) (in case more than 20 participants reply at the endline)
    c. Displaced, Located in Ukraine (Internally Displaced)
    d. Displaced, Located Abroad (Refugee)

   *Hypothesis:* The online self-help course has a different effect conditional on participant status.

B. We will use the Policy Tree algorithm proposed by Athey and Wager (2021) to see for which groups the online self-help course is particularly effective.
C. We will use Generic Machine Learning Inference on Heterogeneous Treatment Effects (Chernozhukov et al., 2020) to understand if there is any relevant heterogeneity in treatment effect by covariates.

## 6. Attrition.

To deal with the non-existence of the data in the endline survey among those who participated in the program. First, we will identify the overall attrition rate among those who participated in the program (those eligible participants who left any identifiable response on the platform after the

start of the program on 7th of July). Second, we will assess whether the attrition is independent of the treatment. If the response rate is below 80% and attrition is not independent of the treatment, we will (a) test the robustness of our results using Lee bounds (Lee, 2009) and (b) reweight the sample.

**7. Potential non-compliers.**

Though there is a restriction on registering multiple accounts, if the participant manages to register more than one account and is assigned to different treatment groups, we exclude this participant from the analysis.

**8. Multiple Hypothesis Testing in Case of Secondary Outcome**

If there is a disagreement in treatment effect between WHO-5 and MHI-5 measures, we will also calculate Romano-Wolf stepdown p-values for these outcomes.

## References


Athey, S. and Wager, S., 2021. Policy learning with observational data. *Econometrica*, 89(1), pp.133-161.

Chernozhukov, V., Demirer, M. and Duflo, E., 2020. Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments. *arXiv. org.*

Topp, C.W., Østergaard, S.D., Søndergaard, S. and Bech, P., 2015. The WHO-5 Well-Being Index: a systematic review of the literature. *Psychotherapy and psychosomatics*, *84*(3), pp.167-176.

Kessler, R.C., Green, J.G., Gruber, M.J., Sampson, N.A., Bromet, E., Cuitan, M., Furukawa, T.A., Gureje, O., Hinkov, H., Hu, C.Y. and Lara, C., 2010. Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. *International journal of methods in psychiatric research*, *19*(S1), pp.4-22.

Veit, Clairice & Ware, John. (1983). The Structure of Psychological Distress and Well-Being in General Populations. Journal of consulting and clinical psychology. 51. 730-42. 10.1037/0022-006X.51.5.730.

Ware, J.E., 1993. SF-36 health survey: manual and interpretation guide. Health Institute.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014) "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28 (2): 29-50.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. Review of Economic Studies, 76(3), 1071–1102. https://doi.org/10.1111/j.1467-937X.2009.00536.x

Appendix A: List of baseline variables for the post-double-selection Lasso.

1. Date of the Registration on the educational platform
2. Date of the registration for the Program
3. Female - a binary variable equal to 1 if the participant is female
4. Ukrainian Language - a binary variable equal to 1 if the speaks Ukrainian in everyday life
5. Birth Region - a factor variable with the following options:
   - West
   - East
   - Center
   - South
6. Residence Region - a factor variable with the following options:
   - West
   - East
   - Center
   - South
7. Education - a factor variable with the following options:
   - "Basic Secondary School Education (9 years)",
   - "Full Secondary School Education (11 years)",
   - "Bachelor's degree",
   - "Master Degree",
   - "PhD"
8. General Life-Satisfaction:
   Overall, how satisfied are you with your life as a whole these days?
   "The following question asks how satisfied you feel about your life, on a scale from "0" to "10" where zero means you feel "not at all satisfied" and "10" means "completely satisfied".
   (Please select one response.)"
9. WHO-5 on a scale from 0 to 25
10. K6 -Kessler 6 on the scale from 0 to 24
11. MHI-5 on a scale from 0 to 100
12. General Health. Answer on the next question: In general, how would you describe your health? (on the scale from 0-10)
13. Time Preferences (on the scale from 0-10)
14. Status - a factor variable with options:
    - "Not Displaced, Located in Ukraine"
    - "Not Displaced, Located Abroad (Migrants)",
    - "Displaced, Located in Ukraine (Internally Displaced)",
    - "Displaced, Located Abroad (Refugee)",
15. Employment Status, a factor variable with options:
    - Unemployed
    - Partially employed
    - Fully employed
    - Business owner / self-employed
    - Student

- Not employed and not looking for work
- Does not know

16. Working Hours: How many hours in the past week did you spend working (for money)? (Winsorized)
17. Volunteering: How many hours in the past week did you spend volunteering (doing any work for free)? (Winsorized)
18. Searching for Job: How many hours in the past week did you spend searching for jobs? (Winsorized)
19. Educational Status, a factor variable with options:
    - Not enrolled in any education
    - The preparatory school (to enter university)
    - Technical academy careers of 1 - 2 years of training
    - College/university program
    - Vocational training program
    - Online course
    - Do not know
    - Other (specify) (Mainly Language courses, Self-Study, PhD and Integrational courses, need to be classified)
20. Looking for Education - How many hours in the past week did you spend searching and applying for educational options (university, courses, ect.)? (Winsorized)
21. Learning a foreign language - a binary variable equal to 1 if the participant is learning a foreign language currently.
22. Probability of take-up: How likely are you to participate in online psychological courses on Prometheus or other platforms? (on the scale from 1-5)