Pre-Analysis Plan for
"(Not) Thinking About the Future: Maternal Labor Force Participation"

[This is an exact copy of the PAP published on November 18, 2022, with all identifying information that might reveal the identity of the participating regions removed.]

1. Introduction

This document pre-specifies the analysis we will conduct with data from our baseline survey and the administrative data. We will submit separate preanalysis plans for our two follow up surveys before we start the data collection for these waves.

The research goal of this study is to test whether learning about the future long-term financial consequences of having a reduced workload affects women's financial awareness and behaviors and, ultimately, increases their labor supply. To test this hypothesis, we designed a field experiment where the treatment consists of an informational video discussing the long-term financial consequences for an example case and a tool that allows users to visualize the long-term financial consequences of different workload scenarios for their individual case. We compare this treatment to a control group that receives videos with unrelated information.

The randomized control trial will be implemented among mothers who work as teachers in the cantons of [name canton A, canton B, and canton C] in Switzerland, and potential additional cantons.

2. Description of the intervention

To test whether learning about the long-term financial consequences of having a reduced workload impacts mothers' behaviors, we designed the following intervention material:

1) Informational video containing three main dimensions of objective information about the long-term financial consequences of having a reduced workload
    a. Information on the total earnings lost in the long-term
    b. Information on the financial consequences for pensions
    c. Information on the financial risk in case of adverse events (such as divorce)

The video also puts the life-time financial losses in perspective to the short-term childcare costs.

2) Online tool: together with a Swiss bank, we designed an online calculator that allows users to simulate the long-term financial consequences of workload reductions for their individual case. We again provide information about the total earnings loss, the pensions

loss, and the comparison with the short-term childcare costs. The tool is available here: https://family-calculator-staging.econ.uzh.ch/login. It is secured with an individual password. The tool automatically gathers the data about the individual calculations, such that we can link them to the individual.

The control group receives a video of similar duration but on an unrelated topic. We use videos created by the national public television as part of their regular programming (https://www.srf.ch/). We randomize the control group into three videos:

- a video containing information about the gender pay gap in Switzerland
- a video with information on proposed tax breaks for families in Switzerland
- a video discussing the current structure of costs of renting or buying housing in Switzerland

In the online baseline survey, participants watch their assigned video. The treatment group receives the online calculator by email directly after completing the baseline. For outcomes in the baseline survey, participants will have received the information in the treatment video only.

3. Recruitment of participants and timeline

The field experiment will be implemented by inviting female teachers via email and physical letters.

3.1 Sample

At the time of writing the PAP, we have access to the contact information of female staff aged between 25 and 50 in cantonal employment contracts at schools, including kindergarten, primary, and secondary schoolteachers, for the school year 2022/23 for the cantons of [canton A, canton B, and canton C].

We do in all three cantons observe their contact information, whether they are on a permanent or temporary contract, and the school of employment. We do not have consistent additional information about the teachers in all three cantons.

Below we describe the sample for each of the cantons.

Canton of [canton A]

Our baseline sample is obtained from administrative data from the Department of Education of the Canton of [A] (DoE [A]). By definition, our sample consist of teachers who all had at least one active contract in the current academic year. The contact data contains [N in canton A] unique individuals, of which [X] have active contracts at more than one school. We restrict our sample to teachers who work at only one school, such that our final sample includes [N-X in canton A]

observations. We include teachers who have a temporary contract that ends before the new school year.
In addition, based on our pre-treatment baseline questions, we restrict the analysis sample to women with children and who are not working full-time (>= 90% workload), as we do not expect the intervention to have an effect on women without children and/or working full-time. We only include full-time working women for descriptive purposes.[1]

Canton of [canton B]
Our baseline sample is obtained from administrative data from the Department of Education of the Canton of [B] (DoE [B]). The contact data contains [N in canton B] unique individuals, of which [Y] have active contracts at more than one school.
We impose the same sample inclusion criteria as in [canton A]: we restrict our sample for the first round of contact to teachers who work at only one school (excluding N=[Y]). We will restrict the analysis sample to women with children and who are not working full-time (>= 90% workload).

Canton of [canton C]
Our baseline sample is obtained from administrative data from the Department of Education of the Canton of [C] (DoE [C]). The DoE [C] provided us with the contact information of female schoolteachers in cantonal employment contracts for the school year 2022/23, aged between 25 and 50. The contact data contains [N in canton C] unique individuals, of which [V] have active contracts at more than one school.
We impose the same sample inclusion criteria as in [canton A]: we restrict our sample for the first round of contact to teachers who work at only one school (excluding N=[V]). We will restrict the analysis sample to women with children and who are not working full-time (>= 90% workload).

3.2 Timeline

We will invite all teachers from the Canton of [A] (end of November 2022) and the Canton of [C] (mid of November 2022) via email. In the Canton of [A], the contact data contains an e-mail address for 92% of the observations. However, we were informed by the DoE [A] that these e-mails are not updated on a regular basis. In the Canton of [C], all contact data contains an e-mail address. We do not have email addresses for the teachers from the Canton of [B].

[1] We randomize full-time working mothers and pregnant mothers separately.

We will, in addition, send physical invitations for our baseline survey by mail to all teachers in our sample in the three cantons during the last weeks of November 2022. The letters will arrive during the first week of December.
The link we provide is only accessible once, so participants can only enter the survey once (via mail or the physical letter). We will start a soft launch on November 16 by inviting teachers from a few schools in the Canton of [A], as well as all teachers in [C] by email (November 18).

We time receipt of the letter and E-mail before teachers start conversations with school principals about their desired workload for the upcoming academic year. We plan to send a reminder of the treatment information to teachers in the treatment group in January 2023 before workload decisions are finalized.

4. Randomization

Our main focus for the analysis is an individual level randomization of teachers into treatment and control group. Due to the potential presence of spillovers between teachers working in the same school, we add a hold-out control group to this design. For the Canton of [A], we implement a two-stage randomization in the following way:

1. In the first stage, we randomize 2/3 of the schools into treatment schools and 1/3 of schools into control schools (from now on referred as the "pure control group"). We stratify the sample by school size terciles (proxied by the number of female teachers aged 25-50 years working in each school based on the DoE contact list), school type (primary or secondary), and type of municipality (rural, semi-urban, city).
2. In the second stage, we randomize teachers in treatment schools at the individual level to treatment or control when participating in the survey. The individual level randomization takes place during the survey, just before the intervention video starts to play. We assign ½ of teachers to treatment and the rest to control.

For the Cantons of [B] and [C], we only randomize at the individual level (i.e., only the second step). We stratify the individual level randomization by canton.
Within the control group (both for the pure control and for the control), we randomize the three control videos described in section 2 with equal probability.
In the Canton of [A], there are [N treated schools in canton A] schools in the treatment group, and [N pure control schools in canton A] schools in the pure control group. This corresponds to sending invitations to [N teachers in treated schools in canton A] teachers in the treatment group, and [N teachers in pure control schools in canton A] teachers in the pure control group. The final number of schools in our sample will depend on the response rate. The same applies for the final number of teachers who will be randomized to treatment and control within treatment schools.

In the Canton of [B], we invite [N teachers invited in canton B] teachers from [N schools in canton B] schools. In the Canton of [C], [N teachers invited in canton C] teachers in [N schools in canton C] schools. Again, the final number of teachers who will be in our sample will depend on the response rate.

5. Primary outcomes

Hypotheses

The intervention is hypothesized to increase:

a) Financial empowerment: women's financial awareness and likelihood of considering long-term financial factors.
b) Workload aspirations: teachers' desired future workload
c) Workload: teachers' actual workload in the next school year

Outcomes

Following the hypotheses stated above, we will have three families of outcomes where we aggregate groups of questions and construct an index for each family. This will reduce the number of hypotheses tested. The full questions can be found in the appendix. We summarize here the three dimensions:

1) Financial awareness and behavior. We have the following questions in this domain:
    a. Important factors question. "*Thinking about Lara's long-term financial situation, which factors do you think have the largest financial impact if Lara increases her workload to 100%?*". We code a variable that considers the ranking of long vs short term factors.
    b. Voucher for a personalized financial consultation with a recommended financial specialist. We code a dummy = 1 if they choose the voucher over a voucher to a webshop.
    c. Information materials. We code a variable that reflects their propensity to choose material relating to financial information of workload decisions (6 questions in total).

Note that due to institutional requirements, we could not pilot these questions. Following the approach in Dahr et al. 2022, using the pure control group, we will for each variable intended for this index estimate the correlation between the variable and the index constructed excluding it. If the correlation is negative, we will exclude the variable from the index. A negative correlation would be indicative that the question was, for example, misunderstood by the respondents.

For this dimension, we will also separately test the impact for the incentivized question on the financial consultation (1b).

2) Desired LFP. We have three questions in this domain:
    a. Desired workload next school year
    b. Desired workload 10 years from now
    c. Workload under different scenarios (5 questions, two of them are conditional on having a partner)
3) Actual LFP. We will receive administrative data from each of the cantons on each participating teacher's workload in the next school year (the school year 2023/2024).

We consider 2 and 3 as "second stage" outcomes that we hypothesize are affected by financial awareness.

In addition, we consider 3) as an outcome that might be difficult to adjust for women due to other constraints. We will use question XX about desired workload under different scenarios, to provide suggestive evidence of these constraints.

6. Empirical strategy

Main analysis

Our objective is to measure the effects of the intervention on financial awareness, workload aspirations, and labor supply. The main analysis includes teachers from treatment schools only (i.e., we exclude teachers from pure control schools) and excludes teachers who were pregnant or worked full-time at baseline.[2] For every primary outcome, we estimate the following specification using the randomization at the individual level:

$$Y_i = \beta_0 + \beta_1 Treat_i + \beta_2 X_i + \beta_3 X_s + \varepsilon_i ,$$

where Y_i is the outcome of interest for teacher *i*. $Treat_i$ is an indicator for treatment status of teacher *i,* X_i is the vector of individual level baseline (pre-determined) covariates, X_s are school-level controls, and ε_i is an error term. We use heteroskedasticity robust standard errors and include canton fixed effects (stratification level). As we expect workload decisions to be fairly persistent, our main specification will include desired workload for the next academic year measured at baseline, or workload in the previous year (from the admin data), as a control to increase precision. For workload outcomes (desired and actual workload), we will also test the

[2] We will separately examine the impact of the intervention for pregnant women. If the sample size is too small, we will include them in the main specification, adding pregnancy as a control variable. Note that within treated schools, pregnant women were randomized separately.

impact by defining the outcome as the change in the desired workload after versus before watching the intervention video (indexes 2) and 3)).

X_i and X_s are baseline teacher and school characteristics that are predictive of the outcome in the control group or are unbalanced at baseline between treatment and control so either improve precision or could create bias. Due to lacking information on the specific list of variables we will obtain from administrative data at the moment of implementing the field experiment, we also use LASSO to determine these, following Belloni, Chernozkuhov, and Hansen (ReStud 2013).

Spillover effects: secondary analysis

To examine if there are spillover effects between teachers in treatment schools, we will use comparisons between teachers in treatment schools and the pure control group. This will be done with the data from the canton of [A]. In this analysis, we cluster standard errors at the school level. Note that spillovers are not a main focus of our research paper, and - in line with our power calculations - we are less powered, such that results should be considered as more suggestive.

In addition, we plan to analyze ``snowball patterns'' in the data to identify schools and individuals for which we suspect spillover effects. This will be based on the within-school time patterns in taking the survey, as well as the share of teachers that were treated. In addition, we collect the names of individuals using the online tool (voluntary input) and will merge any teachers who answered the survey by name to this data.

Exploratory analysis

A section with exploratory analysis will complement our analysis. In particular, we collect data about teachers' usage of the tool. For example, we will analyze the time they spent using the tool, the workloads they used for making calculations, etc.

7. Heterogeneity analysis

Financial awareness & behavior index

a) By financial awareness at baseline

We expect the impact to be stronger for women who are less aware of the long-term financial consequences of reducing labor supply at baseline. To measure this, we will use three questions:

i. Motivation factor: This question asks participants about the factors that a mother should consider when reducing their labor supply. We will first construct a dataset that only contains their identifier and response to this question. Then, three

independent research assistants who will be blind to the purpose of our study, will code a dummy equal to one if the mother mentions long-term financial factors in their response: pensions, career trajectory, forgone earnings in the long-term, or divorce risk.

ii. Childcare costs: We will construct a variable that captures the propensity to agree with the following statement "When considering what workload to choose after having a child the costs for external childcare are usually the key factor considered by families (in Switzerland)".

iii. Pension estimate: we will construct a variable that captures the likelihood of overestimating or underestimating the monthly pension savings in this question "Now, imagine a teacher aged 32 that works 40% and wants to keep this workload until retirement. She earns 4200 CHF per month. How much do you think she would have available each month in her BVG retirement savings?"

b) By workload at baseline

We do not have a directional hypothesis about how the information will affect women who are more attached to the labor force vs women who are less attached to the labor force. For this heterogeneity dimension, we use the workload they report at baseline.

c) By marital status

We expect the information to increase awareness more for women who are not married, as the financial consequences of separation would be bigger. We elicit marital status at baseline.

<u>Desired and actual labor supply</u>

a) By financial awareness at baseline (see definition in a) above).
b) By gender norms

We hypothesize that the impact of the intervention will be smaller for women who hold more conservative gender norms views at baseline.

To measure this, we will construct an index using their views for the following three questions in our survey:

- "A young child (under 3 years old) is likely to suffer if his or her mother works".
- "Family life suffers when the woman has a full-time job".
- "Women are better able to take care of young children (under 3 years old) compared to men"

c) By workload at baseline (see definition in b) above).
d) By difficulty of organizing family life

We expect larger effects for women who would be more able to adjust their family life if they decide to work full-time. We will use the following question: "How easy or difficult would it be for you to organize your family life?"

e) By flexibility in adjusting their workload

We expect larger effects for women who would be more able to adjust workload in their school or at another institution. We will proxy this with the following question:
"How easy or difficult would it be for you to increase your workload (either in your current school or somewhere else)?" and by a measure of how much in need for additional teachers schools are based on proxies in the administrative data.

f) By beliefs about quality of external care.

We expect smaller effects for women with stronger beliefs about the worse quality of external care. The question we use is the following: "How much better or worse do you think external care (in a daycare, in after-school care, youth club or with a nanny) is for your child compared to in-home care by one of the parents or a family member?"

g) By mother's age and age of the youngest child.

We expect the impact to be larger for mothers who are more likely to have completed their fertility and whose youngest child is relatively old.