Education, Immigration and HPV Vaccination: an Informational Randomized Trial

Pre-analysis Plan

Alice Dominici[*], Lisen Arnheim Dahlström[†], Gustav Axen[‡]

1 Introduction

This study investigates whether targeting childhood vaccines' information framing to the background of parents, defined by education and immigration, is effective in incentivizing vaccination uptake and in reducing misconceptions on vaccines. Our field is the Stockholm County in Sweden, where vaccines are offered for free directly in schools upon parents' authorization, thus drastically reducing both monetary and non-monetary costs of vaccinating children. The presence of Swedish population and vaccination registers also allows us to sample from the entire population of parents with children due to immunization, to observe socioeconomic indicators from administrative data, and to have an objective measure of whether the HPV vaccination takes place (not self-reported). Namely, within strata defined by immigration status and education level of mothers, we randomize an informational intervention on the HPV vaccine: the information consists of sending to mothers' home address a 650 words leaflet addressing specific concerns highlighted by epidemiological research, where we vary the information framing in either emotionally charged or statistical/scientific terms. We also include a pure placebo intervention arm. While HPV vaccination is observed from vaccination registers, we measure misconceptions on vaccines with two surveys: one is administered right after the informational intervention, the second after vaccination has taken place (5 months after the intervention). The second survey also investigates whether a series of common concerns of parents have been reduced by the intervention, in order to shed light on mechanisms. We will adjust p-values for multiple hypothesis testing as appropriate within classes of primary outcomes.

[*]European University Institute, Department of Economics and Karolinska Institute, Department of Medical Epidemiology and Biostatistics.

[†]Karolinska Institute, Department of Medical Epidemiology and Biostatistics.

[‡]European University Institute, Department of Economics.

2 Empirical strategy

2.1 Design

Our informational intervention is described in other sections of this pre-registration.

Randomization is at the individual level, within strata. Our population of interest consists of mothers of children due to HPV immunization in September 2021, who attend school in the Stockholm County. From population registers, we first randomly sample 7500 children who will be offered the HPV vaccine in September 2021, in the Stockholm County. This sample size already takes into account partial participation rates[1]. We then link children to their mothers, and stratify mothers based on whether they are immigrants and, if Swedish-born, based on their education level (see Table 1). We also observe socioeconomic indicators for the other parent, and use them as baseline controls. Immigrant mothers are defined as those born in one of the following countries: Eritrea, Somalia, Iran, Iraq, Syria or Afghanistan. This is to ensure representation of other immigrant communities in Europe, on which existing epidemiological studies are based, and for which any policy recommendation would be most relevant.

Overall, our design is summarized by the following table[2]:

Table 1: Sampling, stratification and randomization

Stratum	**Stratum definition**	**N (invited)**	**Expected participants**	**Of which C**	**Of which T1**	**Of which T2**
Immigrants	Selected countries	2500	500	120	190	190
educ-level-1	< High school degree	1600	500	120	190	190
educ-level-2	High school degree	1400	500	120	190	190
educ-level-3	<2 years university	1000	500	120	190	190
educ-level-4	≥2 years university	1000	500	120	190	190
	Total	7500	2500	600	950	950

Since baseline socioeconomic variables are observed for all invited parents, we will test for self-selection of participants by running the following probability model within each stratum and looking at its R^2:

$$1\{Participation\}_i = \alpha_0 + \boldsymbol{\alpha_1 X}_i + \varepsilon_i$$

where $\mathbf{X}$ includes all variables described in Appendix A (except mothers' education level, which defines strata).

2.2 Timeline

The phases of the study are:

1. May 2021: sampling, randomization from population registers (carried out by implementation partner, *Statistics Sweden*);

[1] For each stratum, we rely on estimates of participation rates by our implementing partner, *Statistics Sweden*, which are based on previous studies they run in similar sub-populations.

[2] Initially, the cutoff between the educ-level-3 and educ-level-4 strata was the undergraduate degree. However, the scarcity of mothers in the population with postgraduate education induced us to reformulate the cutoff at 2 years of college education.

2. Mid-June 2021: the implementing partner *Statistics Sweden* delivers baseline data and the results of randomization for baseline calculations;

3. 16th June 2021: informational intervention. Information leaflet and first survey are sent to mothers' home address by post[3]. The initial survey contains the consent form;

4. July-August 2021: mothers who did not respond to the first survey receive up to 3 reminders to their home address. Each reminder includes the information sheet, either in paper or digital format, and the initial survey;

5. September-October 2021: HPV vaccination is offered in schools;

6. First week of November 2021: endline survey is sent to participant mothers' home address.

7. January 14th 2022: the implementing partner *Statistics Sweden* delivers the final dataset containing baseline variables, answers to both surveys, and information on the HPV vaccination from vaccination registers.

2.3 Benchmark model

Within strata and in the entire sample, we test the following null hypotheses:

- None of the informational interventions affected the primary outcome variables ($H_0 : E(Y|C, X) = E(Y|T1 \text{ or } T2, X)$), denoted as C *vs* T;
- Emotional framing is as effective as statistical/scientific framing in raising uptake and reducing misconceptions ($H_0 : E(Y|T1, X) = E(Y|T2, X)$), denoted as T1 *vs* T2;
- Individual treatments are more effective than the placebo in raising uptake and reducing misconceptions ($H_0 : E(Y|T1, X) = E(Y|C, X)$ and $H_0 : E(Y|T2, X) = E(Y|C, X)$), denoted as T_i *vs* C, where $i \in \{1, 2\}$.

$\boldsymbol{X}$ is a vector of baseline characteristics measured at the individual level: it includes the relevant variables from population registers (Appendix A) and self-reported baseline variables such as pre-intervention knowledge of the HPV, preferred source of information on vaccines and presence of a health professional among close friends and relatives.

Tests are performed by estimating average intention to treat effects with ANCOVA regressions, which control for a proxy the primary outcomes at baseline[4], and individual-specific variables (of the mother, of her partner, and the child) which are likely to predict our primary outcomes. For all regressions, we include school fixed effects, since school nurses are an important potential source of extra information on the HPV vaccine:

$$Y_i = \alpha + \beta_{ITT} T_i + \boldsymbol{X_i'}\boldsymbol{\gamma} + school_i + stratum_i + Baseline\ Y_i + \varepsilon_i$$

where T_i is a dummy equal to 1 for treated individuals, where treatment is defined based on the null hypothesis, and stratum fixed effects ($stratum_i$) are only included in the full-sample estimation. We will also present results without controlling for $\boldsymbol{X}$ and school fixed effects. $\hat{\beta}_{ITT}$ is the estimated intention to treat effect.

[3] This is typical in Sweden.

[4] Whether the child has received the second dose of the MMR vaccine, which is the most recent in the Swedish national program schedule, is the vaccine on which most misinformation is focused, and originated most of common misconceptions on vaccines.

2.4 Estimating LATE

Because our behavioral intervention is delivered as a 650 words information leaflet, we expect heterogeneity in implementation intensity. To take this into account, in the survey administered immediately after the the informational intervention, we ask mothers to report the percentage of text they have read. This information will be used to obtain LATE estimates (Local Average Treatment Effects) in the comparisons of treatments with placebo (where the intensity of treatment for group C can be assumed to be equal to 0).

3 Secondary outcomes

We collect three types of secondary outcomes, for three different reasons:

1. **Mechanisms**

   At endline, we ask mothers to report how much they think the vaccines affects the probability of some events occurring. These events are chosen to mirror the specific concerns of specific categories of parents highlighted by the epidemiological literature. These are: (i) The child becomes sterile; (ii) The child develops serious health issues - we ask this twice: before turning 35 and after turning 35; (iii) The child develops cancer - we ask this twice: before turning 35 and after turning 35 (iv) The child needs to undergo invasive and distressing medical procedures in the future; (v) The child has sex before marriage.

   Note that while events (i)-(iv) are actually made less likely by the HPV vaccine, misinformation states the opposite. These variables will help us assess if the variation of these beliefs across treatment arms are a channel through which our intervention affects primary outcomes.

2. **Heterogeneity analysis**

   We will assess whether the intervention has a significant interaction with the baseline variables indicated in Appendix A, and with the following self-reported indicators from surveys: (i) Presence of a medical doctor among mothers' close friends and family (ii) Self-declared percentage of the information sheet they actually read; (iii) Preferred source of information on vaccines (categorical indicator of reliability); (iv) Previous knowledge of HPV and HPV vaccine; (v) Perception of being sufficiently informed; (vi) Importance attributed to children's fertility and virginity before marriage.

3. **Objective vs Subjective indicator of vaccination**

   Immediately after the intervention, we ask mothers to report their intention to vaccinate their child against HPV. We will test if this subjective measure is statistically different from the objective measure from vaccination registers observed at endline.

A Variables from population registers

Relative to:	Indicator
Immigrant mothers	Year of immigration
	Years of education completed in Sweden
All mothers (immigrant and Swedish-born)	Final high school GPA
	Age
	Civil status
	Number of children
Both parents	Highest education level
	Field of study
	Years of education
	Sector of occupation
	Occupation
	Income from the main occupation
	Income from assets (financial and housing)
Child due for HPV immunization	Gender
	Anonymized school code
	Order of birth (i.e. first child, second child, etc.)
	Binary indicator for adopted children
	Vaccinated with the second dose of MMR vaccine (most recent vaccination according to the Swedish national program)