Investigating how administrative burden and search costs affect social inequalities in early childcare access

Draft of a formal model

Arthur Heim

2025-01-31

I Setting

We model the decision to apply for childcare as a function of perceived eligibility, transaction costs, and potential benefits. Households face uncertainty regarding their eligibility and the costs associated with the application process. The intervention aims to reduce these frictions through information provision and administrative support.

We define latent variables with a superscript V^* and a measurement of that varriable as V.

We consider a sequential decision process where at the first period (baseline), a pregnant woman has *ex-ante* unobserved **preference profile** θ_0^* with regard to future childcare options. We observe a set of attributes $\mathbf{X}$ and intention to use I_0. Among $\mathbf{X}$, there are social group metrics (G) such as SES status (based on education), migration background, and rough metrics of temporal preferences ρ and knowledge of childcare $\mathcal{I}$, which capture latent traits ρ^* and $\mathcal{I}^*$ with error. There are also cognitive and structural factor that may hinder decisions. Let $\mathcal{C}$ denote the subjective application costs. Without loss of generality, let $\mathbf{Z}$ denote the set of observable $\mathbf{X}$ supplemented with random shocks such as treatments.

I.1 Intention gap by social groups

Formally, the gap between two sets of values of covariates $\mathbf{x}$ $\mathbf{x}'$ in intention to use at baseline is:

$$\Delta_{x,x'}(I_0) = \mathbb{E}\Big[I_{0i}|\mathbf{X}=\mathbf{x}\Big] - \mathbb{E}\Big[I_{0i}|\mathbf{X}=\mathbf{x}'\Big] \tag{1}$$

With a random sample, these conditional differences in reported intention to use childcare can be estimated using OLS and heteroskedasticity robust standard errors (Linear probability model) or by first estimating a model with parametric restrictions on I_0^* (such as logit or probit) and computing marginal effects.

However, these parameters and sample analogues are only descriptive difference in means between two groups who also differs in observable characteristics and latent U_I.

I.2 Application and access gap in the endline data

The endline survey includes the different sets of outcomes including some metrics of knowledges, application behaviours and access to formal childcare. We leave attrition issues for now and focus on application behaviours. We come back to these points later.

Let $\tilde{Y}_i = \mathbb{1}(\text{applied to any childcare})$ be the endline measure of application behaviours which derive from:

$$\tilde{Y}_i = \underbrace{\mathbb{1}(\text{applied for childminders})}_{Y^c} + \underbrace{\mathbb{1}(\text{applied for daycare})}_{Y^d}$$

In practice, $\tilde{Y}$ also include other childcare arrangement.

The application gap is :

$$\Delta_{x,x'}(\tilde{Y}_i) = \mathbb{E}\left[\tilde{Y}_i|\mathbf{X} = \mathbf{x}\right] - \mathbb{E}\left[\tilde{Y}_i|\mathbf{X} = \mathbf{x}'\right] \tag{2}$$

Denoting $\tilde{Y^*}$ the variable for access to formal childcare, The access gap is :

$$\Delta_{x,x'}(\tilde{Y*}_i) = \mathbb{E}\left[\tilde{Y^*}_i|\mathbf{X} = \mathbf{x}\right] - \mathbb{E}\left[\tilde{Y^*}_i|\mathbf{X} = \mathbf{x}'\right] \tag{3}$$

In our setting, the covariates are discrete. These three parameters can therefore be estimated from the control group by regressing the outcome on group dummies using OLS and heterokedasticity robust standard errors, or by computing marginal effects from a logit or probit model.

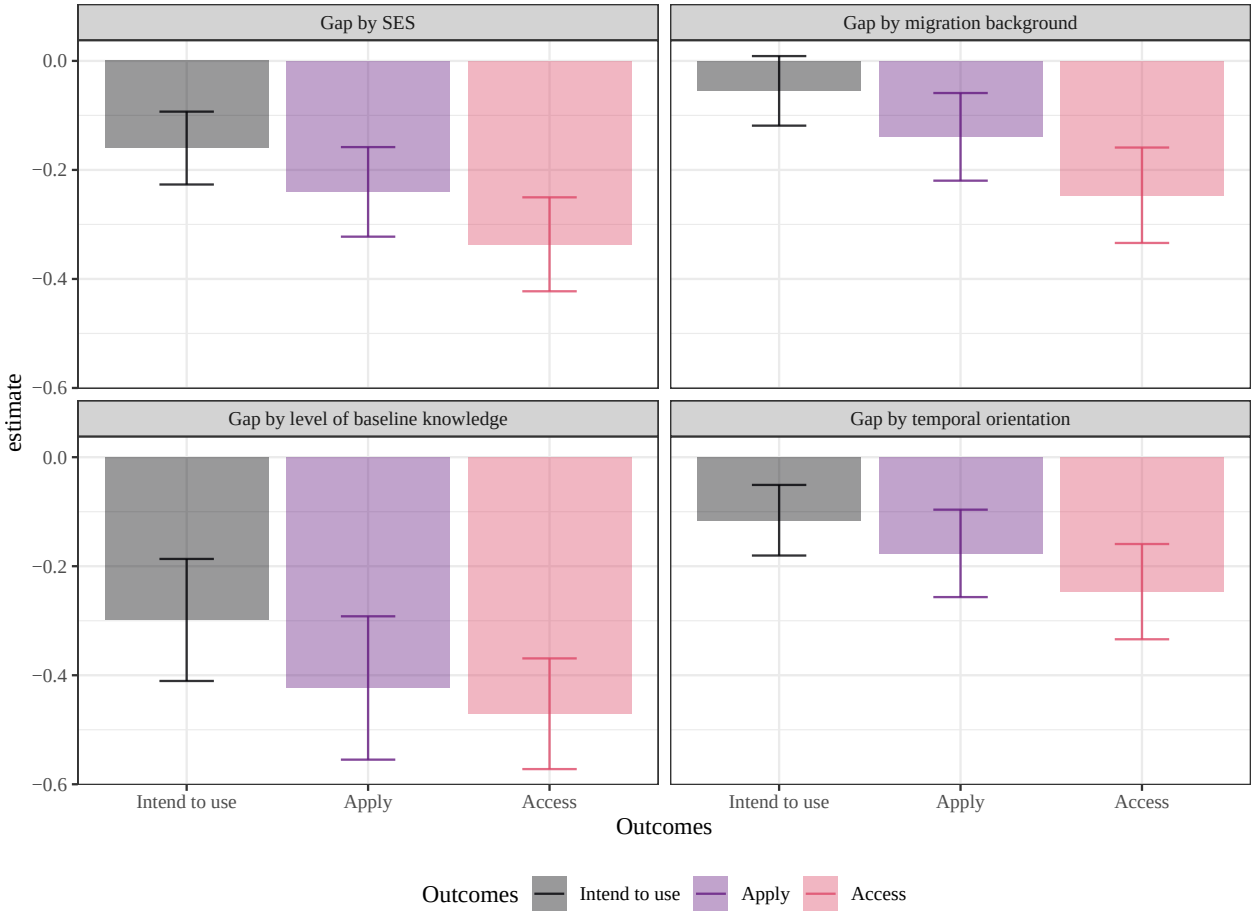
In practice, we simply estimate:

$$W_i = \alpha_w + \beta_w X + \upsilon_{iw} \tag{4}$$

Where W_i denotes any outcome and X one of the social group or subjective attribute variable we test.

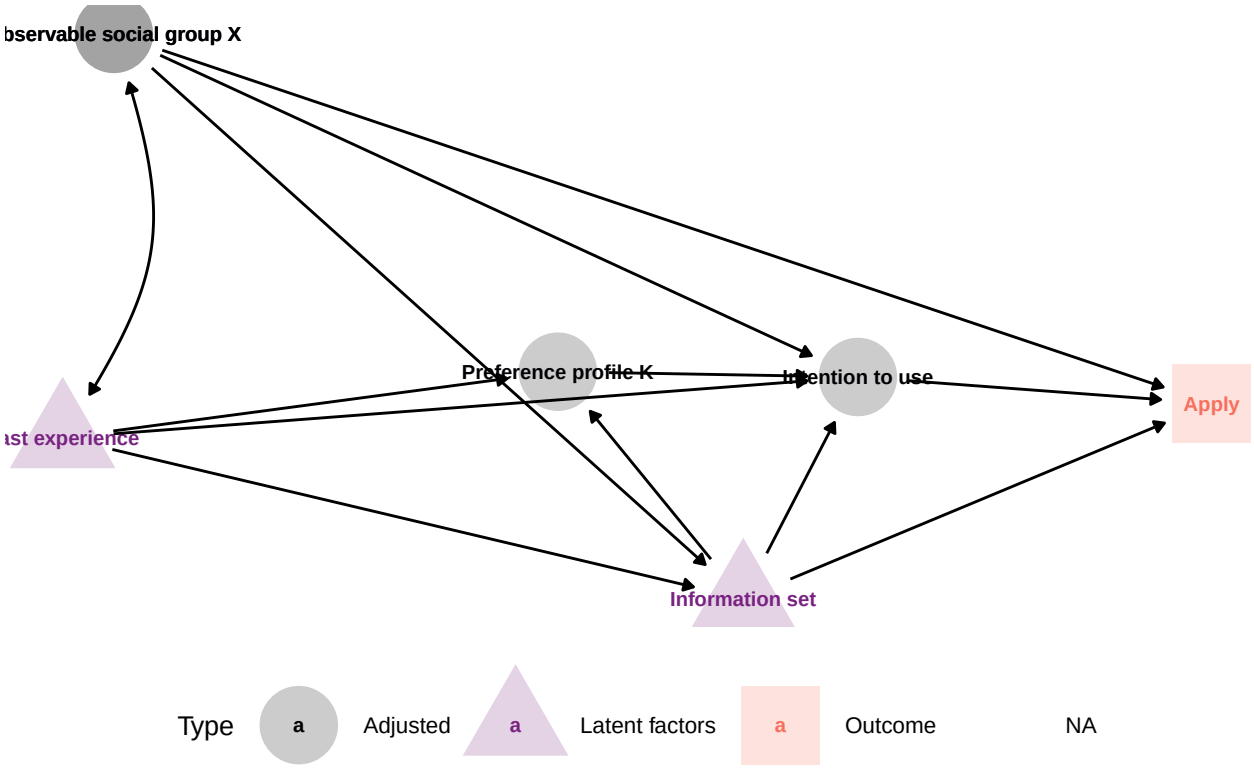
Results are are reported in Figure 1.

Figure 1: The intention-to-action gap in early childcare application and access in the control group across four dimensions of heterogeneity. Intention to use early childcare is measured at the baseline survey during pregnancy. Early childcare application and early childcare access are measured at the endline survey one year after.



Sources: Control group only. Intention are measured at the baseline survey during pregnancy (Q4 2022); application and access are measured at the endline survey one year after (Q4 2023).
Notes: Coefficients of OLS regressions of the outcomes on dummies for the group variable of interest. Error bars indicate pointwise 95% CI based on heteroskedasticity−robust standard errors (HC1). Intention−to−action gap in early childcare application and access gap in the control group across four dimensions of heterogeneity:
− Gap by SES compares households (HH) in which the mother did not attend any kind of post−secondary education with those in which the mother did.
− Gap by migration background compares HH in which the mother was born abroad with HH in which the mother was born in France.
− Gap by level of baseline knowledge compares parents with very low knowledge with those wih higher knowledge of early childcare.
− Gap by temporal orientation compares HH in which the mother prefers a reward of 50. in three days (defined as present−oriented) to HH in which the mother prefers 80. in 3 months.

Figure 2: How observed baseline characteristics relate with the main outcome.



I.3 Intention to action gap

Another parameter of interest is the average intention to action gap $i.e.$ the expected individual difference $\mathbb{E}[Y_i - I_{0i}]$, and differences between social groups. to be clear, this variable equals 0 when people are consistent in their reported intention and application behaviours, equals 1 when the person applied but did'nt intend to (*Switchers*) and -1 when they intended to but did not apply (*Quitters*).

$$\begin{aligned}\mathbb{E}[Y_i - I_{0i}] =& Pr(Y_i > I_{0i}) \underbrace{\mathbb{E}[Y_i - I_{0i} > 0]}_{=1} + Pr(Y_i = I_{0i}) \underbrace{\mathbb{E}[Y_i - I_{0i} = 0]}_{=0} + Pr(Y_i < I_{0i}) \underbrace{\mathbb{E}[Y_i - I_{0i} < 0]}_{=-1} \\ =& \underbrace{Pr(Y_i > I_{0i})}_{switchers} - \underbrace{Pr(Y_i < I_{0i})}_{quitters}\end{aligned}$$

Proof. Starting with $\mathbb{E}[Y_i - I_{0i}]$, we can decompose this expectation by considering all possible cases for the difference $Y_i - I_{0i}$ and using the law of iterated expectation.

Since Y_i and I_{0i} are binary (0 or 1):

- When $Y_i > I_{0i}$, we must have $Y_i = 1$ and $I_{0i} = 0$, so $Y_i - I_{0i} = 1$
- When $Y_i = I_{0i}$, we have $Y_i - I_{0i} = 0$
- When $Y_i < I_{0i}$, we must have $Y_i = 0$ and $I_{0i} = 1$, so $Y_i - I_{0i} = -1$

Therefore:

- $\mathbb{E}[Y_i - I_{0i} | Y_i > I_{0i}] = 1$
- $\mathbb{E}[Y_i - I_{0i} | Y_i = I_{0i}] = 0$
- $\mathbb{E}[Y_i - I_{0i} | Y_i < I_{0i}] = -1$

Using the law of iterated expectation:

$$\begin{aligned}\mathbb{E}[Y_i - I_{0i}] = &\Pr(Y_i > I_{0i}) \cdot \mathbb{E}[Y_i - I_{0i} | Y_i > I_{0i}] + \Pr(Y_i = I_{0i}) \cdot \mathbb{E}[Y_i - I_{0i} | Y_i = I_{0i}] \\ &+ \Pr(Y_i < I0i) \cdot \mathbb{E}[Y_i - I_{0i} | Y_i < I_{0i}]\end{aligned}$$

Continuing with the derivation:

$$\mathbb{E}[Y_i - I_{0i}] = \Pr(Y_i > I_{0i}) \cdot 1 + \Pr(Y_i = I_{0i}) \cdot 0 + \Pr(Y_i < I_{0i}) \cdot (-1)$$

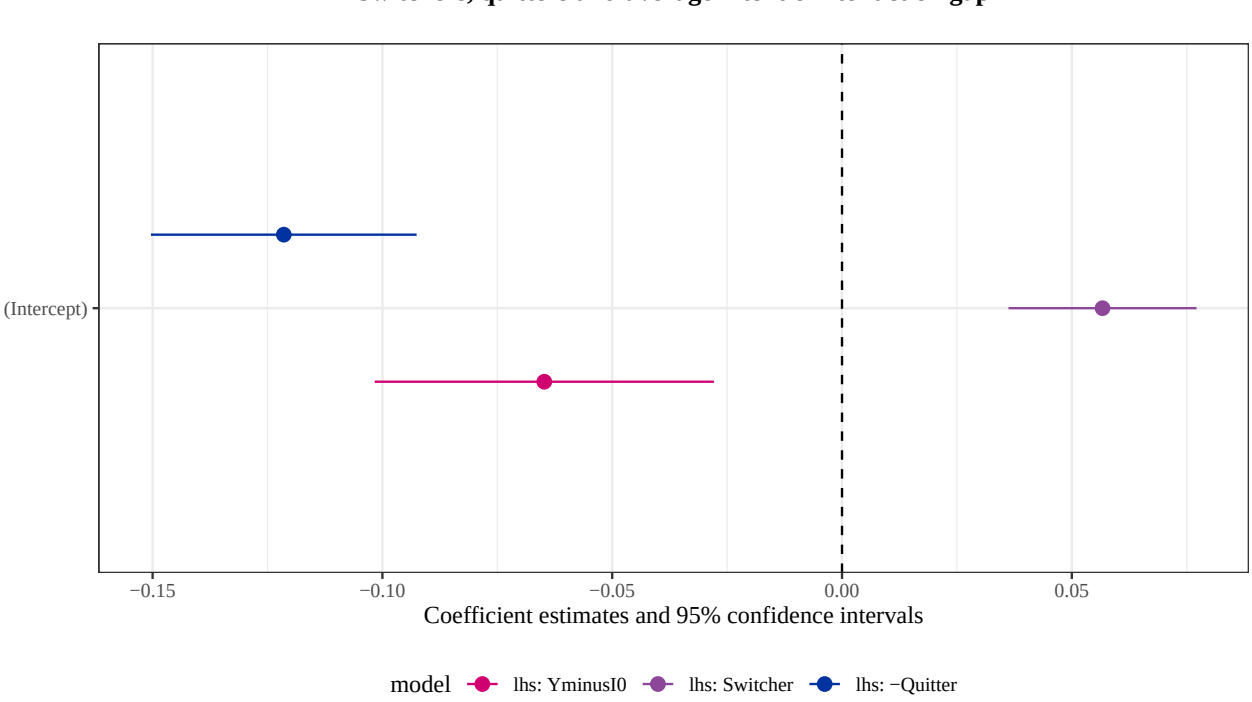
The term $\Pr(Y_i = I_{0i}) \cdot 0$ drops out, and we're left with:

$$\mathbb{E}[Y_i - I_{0i}] = \Pr(Y_i > I_{0i}) - \Pr(Y_i < I_{0i})$$

QED

3 below estimate these three quantities in the control group:

Figure 3: Intention to action gap: switchers and quitters



The intention-to-action gap is -6.5% ; it sums the 5.7% of *switchers* (who did not intend but applied) and the -12.1% of *quitters*.

We can go further into the decomposition using the same expansion. The probability of being a switcher can be divided into the probability of for high and low SES weighted by their relative share in the sample:

For each SES group, we can write:

$$\begin{aligned}\Pr(Y_i > I_{0i} | SES_i = High) = &\Pr(Y_i = 1, I_{0i} = 0 | SES_i = High) \\ \Pr(Y_i > I_{0i} | SES_i = Low) = &\Pr(Y_i = 1, I_{0i} = 0 | SES_i = Low)\end{aligned}$$

These conditional probabilities can be further decomposed using Bayes' rule:

$$\begin{aligned}\Pr(Y_i = 1, I0i = 0 | SES_i = High) &= \Pr(Y_i = 1 | I_{0i} = 0, SES_i = High) \times \Pr(I_{0i} = 0 | SES_i = High) \\ \Pr(Y_i = 1, I0i = 0 | SES_i = Low) &= \Pr(Y_i = 1 | I_{0i} = 0, SES_i = Low) \times \Pr(I_{0i} = 0 | SES_i = Low)\end{aligned}$$

Putting it all together, we get:

$$\begin{aligned}\Pr(Y_i > I0) = &[\Pr(Y_i = 1|I_{0i} = 0, SEi = High) \times \Pr(I_{0i} = 0|SES_i = High) \times Pr(SES_i = High)] + \\ &[\Pr(Y_i = 1|I_{0i} = 0, SES_i = Low) \times \Pr(I_{0i} = 0|SES_i = Low) \times Pr(SES_i = Low)]\end{aligned}$$

This decomposition allows us to quantify:

The contribution of the High SES group to the overall switching probability:

$$\Pr(Y_i > I_{0i}, SES_i = High) = \Pr(Y_i = 1|I_{0i} = 0, SES_i = High) \times \Pr(I_{0i} = 0|SES_i = High) \times \Pr(SES_i = High) \tag{5}$$

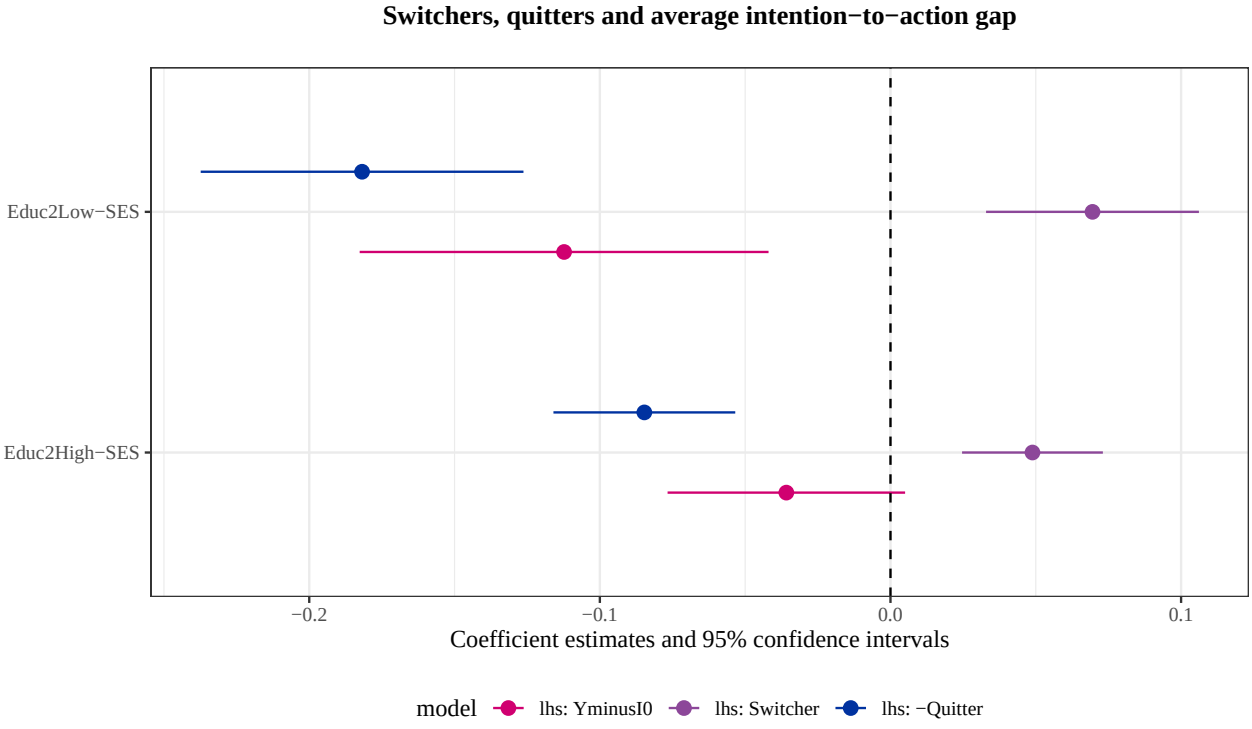
The contribution of the Low SES group to the overall switching probability:

$$Pr(Y > I0, SES = Low) = Pr(Y = 1|I0 = 0, SES = Low) \times Pr(I0 = 0|SES = Low) \times Pr(SES = Low) \tag{6}$$

This decomposition gives hiw much of the overall switching behaviour is driven by each SES group, accounting for both their population share and their propensity to switch. We can do the same derivation for quitters.

We end-up here with a decomposition of the intention-to-action gap as a weighted average of the switching and quitting share in high and low SES group with weights proportional to the share of household who did not intend to use daycare in each group times the share of each group in the sample.

Figure 4: Intention to action gap: switchers and quitters



These results show that for High SES, the intention-to-action gap (I.A.G) is small and not significant with that sample size. It is significant for Low-SES group and we later test if the two are different. The I.A.G of High-SES group averages 5% of switchers and the -8%n of quitters.

The I.A.G of Low-SES group averages 7% of switchers and the -18%n of quitters.

Note that there is a direct link between the intention gap and application gap by social group defined earlier:

Figure 5: Intention to action gap: switchers and quitters by SES

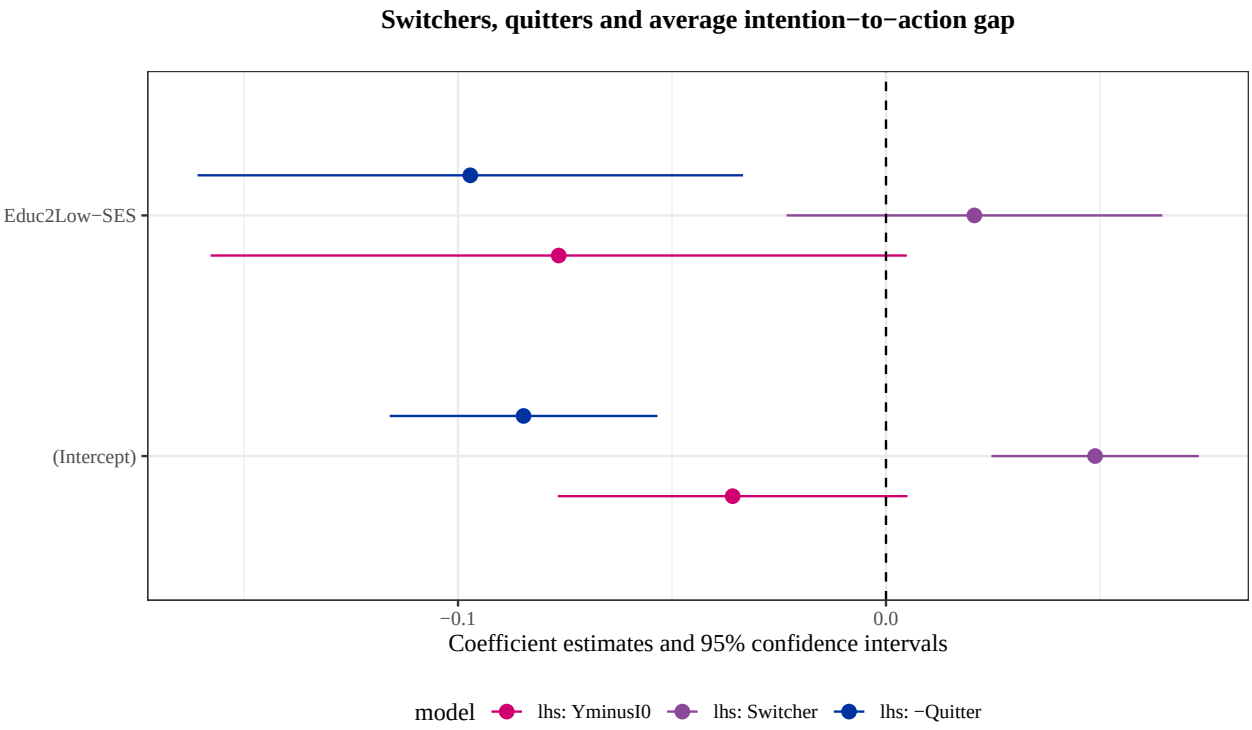


Table 1: Intention to action gap by SES: tests

	lhs: YminusI0	lhs: Switcher	lhs: -Quitter
(Intercept)	−0.036*	0.049***	−0.085***
	(0.021)	(0.012)	(0.016)
Educ2Low-SES	−0.076*	0.021	−0.097***
	(0.041)	(0.022)	(0.032)
Num.Obs.	494	494	494
R2	0.008	0.002	0.021
R2 Adj.	0.006	0.000	0.019
Std.Errors	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust

* p <0.1, ** p <0.05, *** p <0.01

$$\begin{aligned}\Delta_{x,x'}(\tilde{Y}_i) - \Delta_{x,x'}(I_0) = &\quad \mathbb{E}\Big[\tilde{Y}_i|\mathbf{X}=\mathbf{x}\Big] - \mathbb{E}\Big[\tilde{Y}_i|\mathbf{X}=\mathbf{x}'\Big] - (\mathbb{E}\Big[I_{0i}|\mathbf{X}=\mathbf{x}\Big] - \mathbb{E}\Big[I_{0i}|\mathbf{X}=\mathbf{x}'\Big]) \\ = &\quad \mathbb{E}\Big[\tilde{Y}_i - I_{0i}|\mathbf{X}=\mathbf{x}\Big] - \mathbb{E}\Big[\tilde{Y}_i - I_{0i}|\mathbf{X}=\mathbf{x}'\Big]\end{aligned}$$

$$\Delta_{x,x'}(\tilde{Y}_i - I_{0i}) = \quad \mathbb{E}\Big[\tilde{Y}_i - I_{0i}|\mathbf{X}=\mathbf{x}\Big] - \mathbb{E}\Big[\tilde{Y}_i - I_{0i}|\mathbf{X}=\mathbf{x}'\Big] \tag{7}$$

Thus, these parameters can be estimated either by regressing the difference between application and intention on social group dummies or by:

1) stacking baseline (intention) and endline databases (application)
2) defining an outcome Y which equals I_0 in the baseline and $\tilde{Y}$ in the endline.
3) regressing Y on an endline dummy, an interaction with the group indicator, mother fixed effects and clustered SE

The coefficient of the interaction is an estimate of $\Delta_{x,x'}(\tilde{Y}_i - I_{0i})$ i.e. the intention-to-action gap between high and low SES.

Table **??** shows these estimates by high/low education. The average individual gap among high SES is -3.6pp, which average switchers from intention to no use and from no intention to use among high SES households. It decomposes the total intention to action gap. This gap is wider for Low SES Family ; the difference in intention to action gap between the two groups is driven by the second and third columns, both estimating the difference in difference in gaps, either by using the actual difference in variable as outcome, or stacking databases with interactions.

In Table **??**, we test the difference in the share of quitters and switchers by social groups. For switchers, there are no difference by SES status But for quitters, Low-SES status who are already less likely to intend to use childcare are twice more likely to quit !

Table 2: Intention to application gap by SES status

	Mean gap	Group average	OLS	DID
(Intercept)	−0.065***		−0.036*	
	(0.019)		(0.021)	
Educ2High-SES		−0.036*		
		(0.021)		
Educ2Low-SES		−0.112***	−0.076*	
		(0.036)	(0.041)	
Baseline				−0.036*
				(0.021)
Baseline × Educ2Low-SES				−0.076*
				(0.041)
Num.Obs.	494	494	494	1117
R2		0.006	0.008	0.781
R2 Adj.		0.004	0.006	0.503
Std.Errors	Heteroskedasticity-robust	Heteroskedasticity-robust	Heteroskedasticity-robust	by: ResponseId
FE: ResponseId				X

* p <0.1, ** p <0.05, *** p <0.01

Table 3: Intention to application gap by SES status

	Group lhs: Switcher	Group lhs: -Quitter	Diff lhs: Switcher	Diff lhs: -Quitter
Educ2High-SES	0.049***	−0.085***		
	(0.012)	(0.016)		
Educ2Low-SES	0.070***	−0.182***	0.021	−0.097***
	(0.019)	(0.028)	(0.022)	(0.032)
(Intercept)			0.049***	−0.085***
			(0.012)	(0.016)
Num.Obs.	494	494	494	494
R2	0.000	0.019	0.002	0.021
R2 Adj.	−0.002	0.017	0.000	0.019

* p <0.1, ** p <0.05, *** p <0.01

II Treatment effects on quitter and switchers.

Now that we have seen that there are more quitters than switchers and that the gap in quitters is twice as large among low SES group, we look at the effect of information and administrative support on these probabilities.

II.1 Average estimations

Table 4: ITT on Switching and quitting status

	Quitter	Switchers
Information-only vs Control	-0.00 (0.02)	-0.01 (0.01)
	[-0.05, 0.04]	[-0.03, 0.01]
	adj.p.val. = 0.829	adj.p.val. = 0.553
Information + Support vs Control	-0.04** (0.02)	0.00 (0.01)
	[-0.08, 0.00]	[-0.03, 0.03]
	adj.p.val. = 0.060	adj.p.val. = 0.719
Information + support vs Information-only	-0.04* (0.02)	0.01 (0.01)
	[-0.08, 0.01]	[-0.02, 0.04]
	adj.p.val. = 0.094	adj.p.val. = 0.672
Mean control group	0.12 (0.02)	0.05 (0.01)
	[0.08, 0.16]	[0.02, 0.09]

Table 4: ITT on Switching and quitting status

	Quitter	Switchers
Num.Obs.	2906	2906
R2	0.144	0.354
R2 Adj.	0.062	0.292
Fixed effects	X	X
Chi 2	6.43	4.45
P-value	0.092	0.217

Sources: stacked database of pairwise comparisons.
*= p<.1, **= p<.05, ***= p<.01 based on point-wise p-value.
Standard errors are cluster-heteroskedasticity robust adjusted at the block x wave level.
Adjusted p-value and confidence intervals account for simultaneous inference using the Westfall method.
Each column estimates jointly the effects of the program using fully-saturated stacked regressions. Control means estimated separately by OLS.
Joint significance test of null effect using Chi-2 test and p-value are reported at the bottom of the table.

II.2 Analysis by SES Status:

Group		Quitters	Switchers
High-SES	Information-only vs Control	0.01 (0.02)	-0.01 (0.01)
		[-0.04, 0.05]	[-0.02, 0.01]
		adj.p.val. = 0.626	adj.p.val. = 0.422
	Information + Support vs Control	-0.02 (0.02)	-0.01 (0.01)
		[-0.07, 0.03]	[-0.03, 0.02]
		adj.p.val. = 0.528	adj.p.val. = 0.706
	Information + support vs Information-only	-0.03 (0.02)	0.00 (0.01)
		[-0.08, 0.02]	[-0.02, 0.03]
		adj.p.val. = 0.460	adj.p.val. = 0.779

Group		Quitters	Switchers
Low-SES	Information-only vs Control	-0.03 (0.04)	-0.01 (0.02)
		[-0.12, 0.06]	[-0.07, 0.04]
		adj.p.val. = 0.483	adj.p.val. = 0.728
	Information + Support vs Control	-0.07** (0.03)	0.02 (0.03)
		[-0.14, -0.00]	[-0.05, 0.09]
		adj.p.val. = 0.046	adj.p.val. = 0.728
	Information + support vs Information-only	-0.06 (0.04)	0.02 (0.03)
		[-0.14, 0.03]	[-0.04, 0.08]
		adj.p.val. = 0.205	adj.p.val. = 0.728
	Num.Obs.	2906	2906
	R2	0.145	0.355
	R2 Adj.	0.062	0.292
	Fixed effects	X	X