Contents

1 Introduction

About Bebbo

Parents everywhere are in need of information on various aspects of child development from reliable and validated sources as well as guidance on how to support the health and development of their children. However, services providing this sort of information and support are often non-existent or inaccessible for a lot of parents in many places. Often, service providers, even when accessible, might lack necessary knowledge and skills to respond to the questions and concerns parents might have.

Mobile apps are one of the most convenient and easy ways to access information about child development and parenting. However, parenting apps are mainly in English and provide a limited thematic content without a possibility for parents to familiarize with, track, and support all aspects of their child's health and development. In addition, these apps are, naturally, not adapted to contexts of individual countries. Many apps are not free of charge, which presents a significant barrier, particularly for the most vulnerable families. At the same time, the majority of the existing apps operate only in online mode requiring good internet connectivity that is lacking in remote and rural areas.

To support parents to receive timely and quality guidance even when direct contact with service providers is not possible and overcome barriers in access to localized digital solutions with verified content, UNICEF Europe and Central Asia Regional Office (ECARO) developed a mobile parenting app, Bebbo. The mobile application also supports the most vulnerable parents/caregivers with lower education level, in terms of the navigation modalities, off-line operability and selection of the core content. The two main objectives of Bebbo, in line with the UNICEF ECARO Early Childhood Development Theory of Change, are: (1) Improving availability of information for parents on child development, and (2) Supporting parents for responsive caregiving and early intervention. Accordingly, Bebbo app provides users information and interactive tools to help nurture and aid their child's health and development. The launch of Bebbo in 11 countries in the ECA region is a direct response to the identified objective to engage parents and caregivers in nurturing care, positive parenting, stimulating, and learning.

What question does this evaluation answer?

The design of the study is set up to answer the following question in the positive: is asking parents to use Bebbo an effective policy to improve the parenting knowledge, attitudes, and practices of the general population in Bulgaria and Serbia? Note that the study cannot fully answer the question in the negative, it cannot prove that this intervention is ineffective, it can only fail to measure its effectiveness.

In order to understand this study, it's important to break down the question we are answering and understand how the design of the study limits or defines each component:

Asking parents to use Bebbo Study participants were randomly divided into two treatment groups. Each group was surveyed at baseline and then "treated" through an invitation to engage with parenting material. One group was asked to visit a website (this is the "treatment as usual", TAU, or the "control" group). The other group was asked to download Bebbo (the "treated" group). By comparing Bebbo to the existing TAU, we are asking the question: "does this new treatment offer something above and beyond the already existing treatments which parents might presumably already be asked to do?"

Why ask? This study follows a randomized encouragement design. This design is used whenever the following two conditions occur:

1. One is interested in the impact of a treatment on a population where individuals can choose whether or not to take the treatment (the "compliers")

2. The compilers and non-compliers might have different reactions to the treatment.

This design is not useful if the reaction to the treatment is independent of willingness to comply. This design is also not useful if one wants to know the theoretical impact that the treatment could have on those who choose to not take the treatment.

Effective An effective policy, in this study design, is one that can be measured to: Improve the way in which a treated individual responds to a set of survey questions, the second time they are asked, compared to the first time they were asked, and then compared to the treatment as usual.

Does so when the treatment is applied to the individual after being asked the survey questions the first time and 4-6 weeks before answering the survey questions the second time.

General Population The general population in these countries had significant pre-exposure to Bebbo before the beginning of the study. This population is different from a population that has no exposure to Bebbo. One would expect that asking people to download and use Bebbo would have a bigger impact on a population that has no exposure to Bebbo.

Additionally, no care was given to single out any particular subset of the population that might benefit the most (or the least) from Bebbo, nor those who would be most likely to use Bebbo.

What do we find?

We do not find evidence that asking this population to use Bebbo has any impact beyond that of asking them to visit a parenting website.

Given the design of the study, we can say that the following facts may have contributed to the lack of evidence of impact and are interesting facts in their own right, when considering the potential impact of this intervention:

1. The population was already very "good" in regards to the outcomes of interest. We measured the improvement of parents over time, under treatment, but many could not improve from their baseline scores, which were perfect.

2. Participants improved from the first questionnaire to the second questionnaire, regardless of treatment arm and regardless of compliance. This seems to imply that the very act of asking the questions improves the way that parents answer them.

3. Very few people complied with treatment and used Bebbo. Of those who were asked to use the app, 28% used the app, 12% used the app more than one day, and only 3% used the app more than three days. With a small percentage using the app intensively, the impact on the entire population will be too small to measure with the size of this study. Significant pre-exposure could have led to the low initial compliance (the 28

2 Study Design

Experiment Design

This study follows a prepost design (Clifford, Sheagley, and Piston 2021) in which we measure the outcomes of interest before treatment (in a baseline survey) and after treatment (in an endline survey). We add an additional survey after the endline, referred to as a follow up, to look for longer-term impacts and test the impact of continued app usage.

Study participants are randomized, from the beginning, to one of two conditions:

1. **Treatment.** Participants in the treatment condition were told that there was one more step to qualify for the study and were then asked to download the app Bebbo and use it regularly, being encouraged that doing so will help them with their parenting.

2. **Control.** Participants in the control condition were told that there was one more step to qualify for the study and were then asked to visit a parenting website and use it regularly, being encouraged that doing so will help them with their parenting.

App usage in the treatment group was tracked, however website visits in the control group were not tracked.
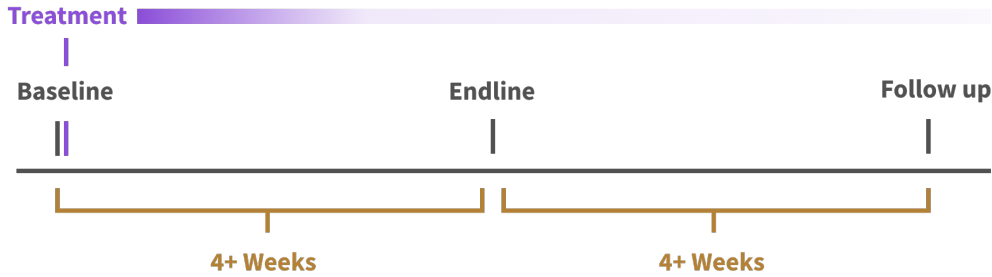

Figure 1: Study Design

Recruitment

Participants were recruited to the study with social media ads on the Meta platform (Facebook and Instagram) using the Virtual Lab platform to create and run the recruitment ads. In exchange for participating in the study, they were told they could receive gift cards worth up to 12 USD (in their local currency). See figure 2 for examples of the ad material used for recruiting.

Recruitment and survey administration was performed on a rolling basis between March and October, 2023. Each individual participant was treated at the end of the baseline survey and sent the endline survey 4 weeks after completing the baseline survey.

The survey was administered via a chatbot in Facebook Messenger, using the Virtual Lab platform. Respondents who clicked on the advertisements were directed to a Messenger chat with the Virtual Lab Facebook page, which did not contain any content or information related to this study. Consent was provided via chat, as well as all answers to the survey questions and the treatment condition. Gift cards were also provided via chat, using the Tremendous gift card platform to provide Visa international prepaid cards.

[TODO: add recruitment stats]

3 Descriptives

Respondent Characteristics by Country

Table 1 provides the baseline characteristics of the respondent population, separated by country.

Generally speaking, most respondents were themselves parents (not grandparents or other caregives), women, under 35 years of age, and spoke the dominant language of the country at home. A little over half had children 0-2, compared to 2-6 years of age.

Respondents in Bulgaria were more likely to have a university education (42%) compared to those in Serbia (29%).
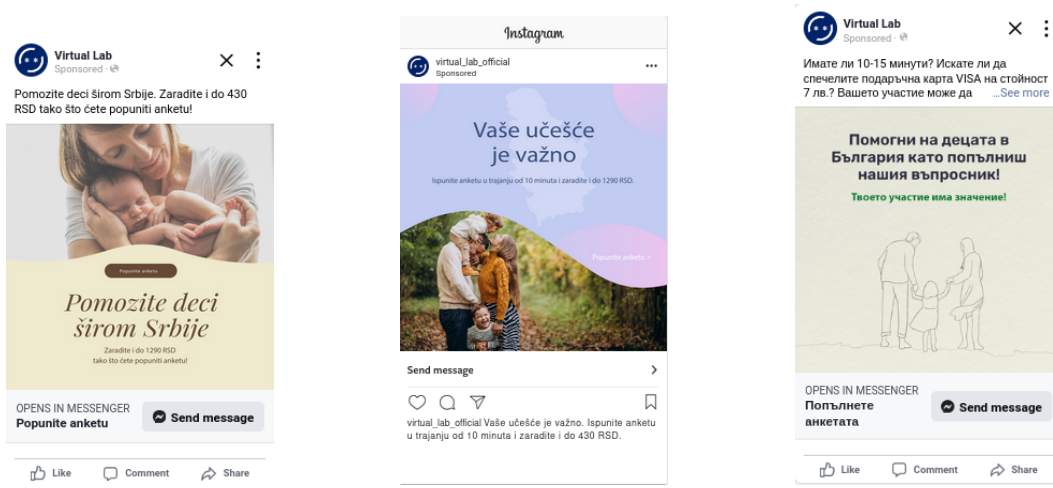
Figure 2: Recruitment Ads

Table 1: Baseline Respondent Characteristics

Variable	Value	Bulgaria	Bulgaria %	Serbia	Serbia %
woman	1	1418	0.83	2102	0.80
university	1	725	0.42	748	0.29
dominant_language	1	1571	0.92	2488	0.95
is_parent	1	1485	0.87	2374	0.91
age_flag	2-6	935	0.55	1561	0.60
children_count	4+	70	0.04	279	0.11
parent_age_flag	Over 35	365	0.21	550	0.21
urban	1	1059	0.62	941	0.36

Construct Variables

The outcomes of interest consist of eight constructs divided into three domains: knowledge and awareness, confidence and attitudes, and practices. The mapping between the constructs, domains, and questions that make up the constructs are laid out in table 2.

The constructs "health knowledge", "caregiver wellbeing", and "was breastfed" are made up of only one question. The construct "practices 24" consists of a count of the number of activities, within the previous 24 hours, that the respondent has done. The rest of the constructs are created by averaging of a set of likert variables.

Descriptive statistics regarding the baseline responses for the outcomes are shown in table 3. Note that many of the constructs have quite high means and medians [TODO: add percent right-truncated to the table].

Table 2: Construct Variable Mapping

Domain	construct_variable	question
Knowledge and awareness	health_knw	I know which vaccine {{field:child_name}} needs to take next.
Knowledge and awareness	dev_knw_recog	I would be able to recognize if {{field:child_name}} lags behind in social-emotional development (expressing and recognizing feelings and emotions, engaging in interactions, etc.).
Knowledge and awareness	dev_knw_recog	I would be able to recognize if {{field:child_name}} lags behind in cognitive development (mental development, intellectual development).
Knowledge and awareness	dev_knw_recog	I would be able to recognize if {{field:child_name}} lags behind in physical development.
Knowledge and awareness	dev_knw_recog	I would be able to recognize if {{field:child_name}} lags behind in language development.
Confidence and attitudes	confidence	How confident do you feel in your ability to deal with {{field:child_name}}'s emotions?
Confidence and attitudes	confidence	How confident do you feel in your ability to respond properly when {{field:child_name}} misbehaves?
Confidence and attitudes	attitude	Do you agree that in order to bring up, raise, or educate a child properly, the child needs to be physically punished?
Confidence and attitudes	caregiver_well_being	How often can you handle stressful parenting situations successfully?
Practices	was_breastfed	Has {{field:child_name}} been breastfed in the last 24 hours?
Practices	practices_24	In the past 24 hours, did you read books or look at picture books with {{field:child_name}}?
Practices	practices_24	In the past 24 hours, did you tell stories with {{field:child_name}}?
Practices	practices_24	In the past 24 hours, did you sing songs (including lullabies) to or with {{field:child_name}}?
Practices	practices_24	In the past 24 hours, did you take {{field:child_name}} outside the home?
Practices	practices_24	In the past 24 hours, did you play with {{field:child_name}}?
Practices	practices_24	In the past 24 hours, did you name, count or draw things with or for {{field:child_name}}?
Practices	practices_agree	When {{field:child_name}} and I play together, we laugh a lot.
Practices	practices_agree	I joke around with {{field:child_name}}.
Practices	practices_agree	I often smile when I'm around {{field:child_name}}.
Practices	practices_agree	{{field:child_name}} and I play together on the floor.
Practices	practices_hostility	I snap at {{field:child_name}} when he/she gets on my nerves.
Practices	practices_hostility	When {{field:child_name}} upsets me, I lose my patience and punish him/her more severely than I really mean to.
Practices	practices_hostility	When {{field:child_name}} does something wrong, I sometimes threaten him/her.
Practices	practices_hostility	I sometimes make fun of {{field:child_name}}.

Table 3: Outcome Construct Descriptives Pooled Baseline

Subdomain	name	mean	median	min	max	sd	prop_na
Attitude	attitude	3.13	3.0	1	4	0.84	0.00
Child development knowledge	dev_knw_recog	0.86	1.0	0	1	0.28	0.00
Confidence in parenting	confidence	3.34	3.5	1	4	0.65	0.00
ECD engagement	practices_24	4.92	5.0	0	6	1.23	0.00
Health knowledge	health_knw	0.72	1.0	0	1	0.45	0.58
Nutrition	was_breastfed	0.37	0.0	0	1	0.48	0.58
Responsive parenting Hostility	practices_hostility	3.04	3.0	1	4	0.69	0.00
Responsive parenting PA	practices_agree	3.20	3.5	1	4	0.75	0.00

Baseline Balance

To test for balance between our randomly assigned treatment and control groups, we run an omnibus test, following Hansen and Bowers (2008), to observe standardized differences at baseline and the associated omnibus p-value. Results are found in table 4.

The most notable variable is the practices_24 and was_breastfed, which differ a bit between the groups although they are still under 0.1 standard deviations in their diffrence. The omnibus test, with a p-value above 0.13, implies that the two groups are not different in a statistically significant manner.

Table 4: Baseline Balance Pooled

	control_mean	treatment_mean	standardized_diff	z_score
health_knw	0.73	0.71	-0.06	-1.20
dev_knw_recog	0.86	0.85	-0.04	-1.29
confidence	3.33	3.34	0.01	0.37
attitude	3.14	3.11	-0.04	-1.23
was_breastfed	0.35	0.39	0.08	1.79
practices_24	4.97	4.87	-0.07	-2.46
practices_agree	3.20	3.19	-0.01	-0.39
practices_hostility	3.03	3.04	0.004	0.12
(health_knw)	0.42	0.43	0.02	0.67
(was_breastfed)	0.42	0.43	0.02	0.61

Overall P-Value: 0.134132865981628

Pre-Exposure to Bebbo

To check for either pre-exposure or contamination, we ask control group users, at the end of the final follow up survey, if they have ever heard of Bebbo or used Bebbo.

55% of respondents said that they had heard about the app Bebbo and 23% said that they had downloaded and used the app Bebbo. It's worth noting that there might be some social desireability bias or acquiesence bias in these responses and we do not have a good way to detect that in this instance. However, despite those potential biases, this is strong suggestive evidence that there was pre-exposure to the treatment in our sample.

Pre-Post Descriptives

One of the dangers of a prepost design is that you are priming your respondents with the first survey and that priming may impact how they answer the questions in the post-treatment survey(s) (Stantcheva 2023).

Given this particular study design, where our control is a "treatment as usual" (TAU) that involved sharing a website and we do not have data regarding the takeup, or usage, of the website, it is difficult to isolate a priming effect.

[TODO: Add an analysis of priming impact –¿ t-test for pre/post in the control group]

Plots of pre-post means are available in

Power Analysis
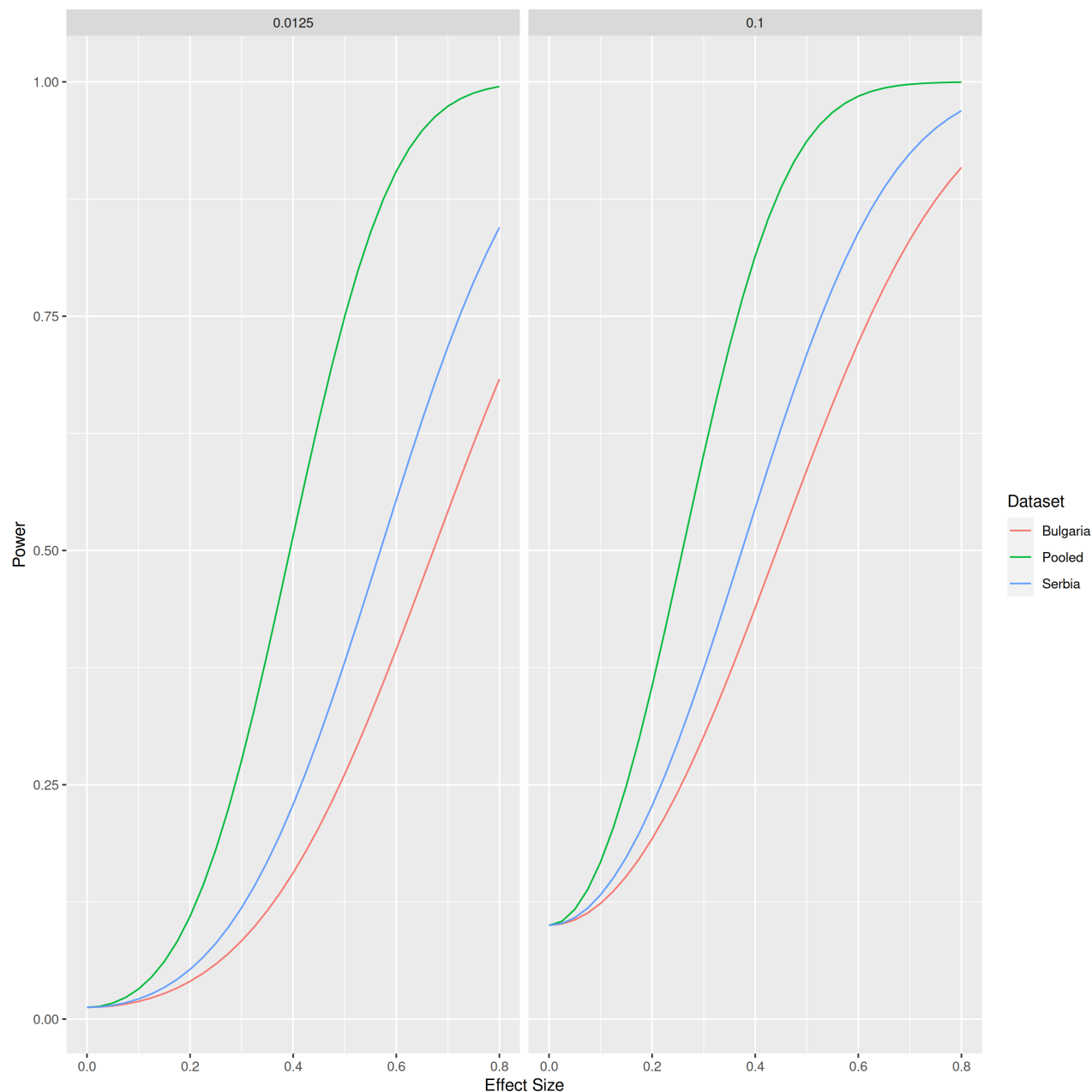
App Usage Over Time

[TODO: Add app usage over time plots and stats]x

Attrition

We define attrition as a respondent present in baseline but not present in endline. Attrition is indicated with a binary flag at the user level. For all respondents, we fit a logistic regression model to whether the respondent attrited using the respondents' baseline characteristics, demographic variables, and a flag for treatment (Table 5).

- The goodness of fit measure used is the percentage improvement in deviance over the null deviance (pseudo R^2). The pseudo R^2 for this model is 0.95.

-

We also fit a logistic regression model to whether the respondent attrited using the respondents' app usage (Table 6). This is fit only to the treated users who have app usage. The pseudo R^2 for this model is 0.007. A respondent's *usage_count* and *days_used* are weak signals of future attrition.

Table 5: user attrition (binary) predicted by baseline characteristics

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.47	2.12	-0.69	0.49
treated	129.00	8,782.77	0.01	0.99
dev_knw_recog	0.14	0.70	0.20	0.84
confidence	-0.41	0.33	-1.27	0.20
attitude	-0.23	0.22	-1.07	0.28
caregiver_well_being	0.02	0.40	0.06	0.95
practices_24	1.70	1.17	1.45	0.15
practices_agree	-0.24	0.33	-0.72	0.47
practices_hostility	0.35	0.29	1.20	0.23
parent_age	0.001	0.002	0.38	0.70
number_children	0	0	-0.01	0.99
parent_genderWoman	-0.04	0.53	-0.07	0.94
survey_duration	-0.26	0.10	-2.70	0.01
education	-0.79	0.51	-1.54	0.12
age_flag2-6	0.09	0.43	0.20	0.84
countryserbia	-1.26	0.66	-1.92	0.05

Table 6: user attrition (binary) predicted by app usage

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.70	0.16	16.85	0
home_opened	-0.43	0.39	-1.11	0.27
usage_count	-0.001	0.001	-1.77	0.08
days_used	0.18	0.09	1.94	0.05
downloaded	0.26	0.37	0.69	0.49

4 Results

We run the following regression model to measure the intent-to-treat effect (ITT) of assignment to the treatment arm on the outcome construct (c) of interest for individual i:

$$y_i = \alpha + \beta T_i + \gamma_1 X_i + \gamma_2 Z_i + \epsilon$$

Note that due to the relatively large number of separate outcomes (8), we adjust p-values of the treatment variable to control the false discovery rate (FDR), using Benjamini-Hochberg, reported as the "Adjusted Treatment p-value."

Table 7: Pooled: OLS - Endline - Knowledge and Awareness

	Dependent variable:	
	Vaccine Knowledge	Child Dev. Knowledge
	(1)	(2)
Treatment	0.03	−0.004
	(0.03)	(0.01)
Adjusted Treatment p-value	0.695	0.714
Observations	696	1,931
R^2	0.01	0.01

Note: *p<0.1; **p<0.05; ***p<0.01

Table 8: Pooled: OLS - Endline - Confidence and Attitudes

	Dependent variable:	
	Parenting Confidence	Attitude to Phys. Punishment
	(1)	(2)
Treatment	−0.02	0.08
	(0.03)	(0.04)
Adjusted Treatment p-value	0.695	0.133
Observations	1,905	1,892
R^2	0.003	0.01

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9: Pooled: OLS - Endline - Practices

	Dependent variable:			
	Breastfed	Activities Past 24h	Positive Practices	Hostile Practices
	(1)	(2)	(3)	(4)
Treatment	−0.02	0.12	0.02	0.02
	(0.03)	(0.05)	(0.03)	(0.03)
Adjusted Treatment p-value	0.695	0.133	0.695	0.695
Observations	657	1,832	1,832	1,827
R^2	0.02	0.01	0.01	0.01

Note: *p<0.1; **p<0.05; ***p<0.01

Table 10: Pooled: 2SLS - Endline - Knowledge and Awareness

	Dependent variable:	
	Vaccine Knowledge	Child Dev. Knowledge
	(1)	(2)
Used App	0.10	−0.02
	(0.11)	(0.04)
Adjusted Treatment p-value	0.695	0.714
Weak instruments p-value	1.17e-28	2.44e-77
Wu-Hausman p-value	0.849	0.355
Observations	696	1,931
R^2	0.01	0.01

Note: *p<0.1; **p<0.05; ***p<0.01

Table 11: Pooled: 2SLS - Endline - Confidence and Attitudes

	Dependent variable:	
	Parenting Confidence	Attitude to Phys. Punishment
	(1)	(2)
Used App	−0.06	0.29
	(0.10)	(0.14)
Adjusted Treatment p-value	0.695	0.139
Weak instruments p-value	7.65e-76	2.16e-76
Wu-Hausman p-value	0.353	0.0374
Observations	1,905	1,892
R^2	0.001	−0.01

Note: *p<0.1; **p<0.05; ***p<0.01

Table 12: Pooled: 2SLS - Endline - Practices

	Dependent variable:			
	Breastfed	Activities Past 24h	Positive Practices	Hostile Practices
	(1)	(2)	(3)	(4)
Used App	−0.07	0.41	0.07	0.06
	(0.09)	(0.20)	(0.12)	(0.11)
Adjusted Treatment p-value	0.695	0.139	0.695	0.695
Weak instruments p-value	3.8e-28	9.16e-75	6.29e-75	3.87e-75
Wu-Hausman p-value	0.805	0.027	0.671	0.306
Observations	657	1,832	1,832	1,827
R^2	0.03	−0.01	0.01	0.005

Note: *p<0.1; **p<0.05; ***p<0.01

Table 13: Pooled for Follow Up: OLS - Follow Up - Knowledge and Awareness

	Dependent variable:	
	Vaccine Knowledge	Child Dev. Knowledge
	(1)	(2)
Treatment	−0.003	−0.01
	(0.07)	(0.02)
Adjusted Treatment p-value	0.999	0.999
Observations	134	420
R^2	0.03	0.04
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 14: Pooled for Follow Up: OLS - Follow Up - Confidence and Attitudes

	Dependent variable:	
	Parenting Confidence	Attitude to Phys. Punishment
	(1)	(2)
Treatment	−0.05	0.0001
	(0.06)	(0.08)
Adjusted Treatment p-value	0.999	0.999
Observations	419	419
R^2	0.04	0.04
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 15: Pooled for Follow Up: OLS - Follow Up - Practices

	Dependent variable:			
	Breastfed	Activities Past 24h	Positive Practices	Hostile Practices
	(1)	(2)	(3)	(4)
Treatment	−0.11	0.07	0.004	−0.03
	(0.07)	(0.12)	(0.08)	(0.07)
Adjusted Treatment p-value	0.796	0.999	0.999	0.999
Observations	133	417	417	416
R^2	0.04	0.004	0.03	0.02
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 16: Pooled for Follow Up: 2SLS - Follow Up - Knowledge and Awareness

	Dependent variable:	
	Vaccine Knowledge	Child Dev. Knowledge
	(1)	(2)
Used App	−0.02	−0.04
	(0.35)	(0.08)
Adjusted Treatment p-value	0.999	0.999
Weak instruments p-value	6.1e-05	1.87e-17
Wu-Hausman p-value	0.443	0.522
Observations	134	420
R^2	0.03	0.03

Note: *p<0.1; **p<0.05; ***p<0.01

Table 17: Pooled for Follow Up: 2SLS - Follow Up - Confidence and Attitudes

	Dependent variable:	
	Parenting Confidence	Attitude to Phys. Punishment
	(1)	(2)
Used App	−0.17	0.0003
	(0.24)	(0.31)
Adjusted Treatment p-value	0.999	0.999
Weak instruments p-value	4.56e-17	4.56e-17
Wu-Hausman p-value	0.587	0.973
Observations	419	419
R^2	0.03	0.04

Note: *p<0.1; **p<0.05; ***p<0.01

Table 18: Pooled for Follow Up: 2SLS - Follow Up - Practices

	Dependent variable:			
	Breastfed	Activities Past 24h	Positive Practices	Hostile Practices
	(1)	(2)	(3)	(4)
Used App	−0.53	0.26	0.01	−0.13
	(0.33)	(0.47)	(0.30)	(0.26)
Adjusted Treatment p-value	0.861	0.999	0.999	0.999
Weak instruments p-value	5.05e-05	1.18e-16	1.18e-16	1e-16
Wu-Hausman p-value	0.308	0.804	0.709	0.977
Observations	133	417	417	416
R^2	−0.01	0.004	0.03	0.02

Note: *p<0.1; **p<0.05; ***p<0.01

5 User Characteristics Correlated with App Usage

In this analysis, we attempt to answer the questions -

1. Who are the respondents who used the app?

2. Do more knowledgeable parents use the app more?

We do so by regressing respondents' app usage activity against their characteristics.

Independent variables :

- **Baseline characteristics** - respondents' scores on the construct variables. We only use the construct variables that have sufficiently high internal consistency. We drop the constructs that are highly correlated with other constructs. We find *parent_knw* to be correlated with *caregiver_well_being*. We drop the construct with lower reliability *caregiver_well_being*.

- Demographics variables - parents' age flag (categorical), university flag (binary), gender (categorical), and number of children (numeric)

- Survey response variables - survey duration, start week, country flag

Dependent variables :

- Home opened - binary variable indicating whether the respondent had a home opened event logged

- Home opens - continuous count variable indicating the respondents' count of total home opens

Home Opens and User Characteristics

For respondents who downloaded the app, we regress their number of home opens against their baseline characteristics, demographic variables, and survey response variables. The R^2 for this model is 0.065. Note that the country flag is not significant which means that app usage does not differ significantly across the two countries after holding constant user characteristics (Table 19).

Table 19: home opens predicted by baseline characteristics

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	10.35	1.79	5.78	0
start_week	-0.15	0.03	-4.35	0
dev_knw_recog	0.36	0.59	0.60	0.55
confidence	-0.31	0.28	-1.08	0.28
attitude	-0.07	0.22	-0.30	0.76
caregiver_well_being	-0.90	0.34	-2.64	0.01
practices_24	-2.23	0.78	-2.85	0.005
practices_agree	0.31	0.26	1.19	0.23
practices_hostility	0.31	0.26	1.19	0.24
parent_age	-0.01	0.02	-0.82	0.41
number_children	0	0	0.36	0.72
parent_genderWoman	0.63	0.45	1.41	0.16
survey_duration	0.09	0.05	1.82	0.07
education	0.02	0.36	0.04	0.96
age_flag2-6	-1.22	0.34	-3.55	0
countryserbia	-0.09	0.58	-0.15	0.88

Home Opened and User Characteristics

For respondents who were treated, i.e., asked to download the Bebbo app, we fit a logistic regression model to whether the respondent had a home opened event logged using the respondents' baseline characteristics, demographic variables, and their survey response variables (Table **??**). The goodness of fit measure used is the percentage improvement in deviance over the null deviance (pseudo R^2). The pseudo R^2 for this model is 0.03.

Table 20: home opened (binary) predicted by baseline characteristics

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.11	0.52	2.15	0.03
start_week	-0.04	0.01	-3.87	0
dev_knw_recog	0.31	0.17	1.84	0.06
confidence	-0.15	0.08	-1.89	0.06
attitude	0.01	0.06	0.22	0.82
caregiver_well_being	-0.20	0.09	-2.12	0.03
practices_24	-0.62	0.22	-2.77	0.01
practices_agree	0.13	0.07	1.84	0.07
practices_hostility	-0.10	0.07	-1.44	0.15
parent_age	-0.02	0.01	-3.71	0
number_children	0	0	0.24	0.81
parent_genderWoman	0.31	0.13	2.47	0.01
survey_duration	0.06	0.01	4.59	0
education	0.11	0.10	1.12	0.26
age_flag2-6	0.11	0.10	1.15	0.25
countryserbia	0.20	0.16	1.29	0.20

References

[1] Scott Clifford, Geoffrey Sheagley, and Spencer Piston. "Increasing Precision without Altering Treatment Effects: Repeated Measures Designs in Survey Experiments". In: *American Political Science Review* 115.3 (2021), pp. 1048–1065. ISSN: 15375943. DOI: 10.1017/S0003055421000241.

[2] Stefanie Stantcheva. "How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible". In: *Annual Review of Economics* 15 (2023), pp. 205–234. ISSN: 19411391. DOI: 10.1146/annurev-economics-091622-010157.

A Additional Plots

Pre-Post Plots