Citizens against corruption

Pre-analysis plan

Anne Hilger, World Bank

Quentin Stoeffler, Mohammed VI Polytechnic University

October 2021

1. Introduction

From 2020 to 2021, we are conducting research to measure the impact of a smartphone App which provides information on administrative processes in Burkina Faso with the aim to empower citizen and address bribery. The research was conducted in partnership with the social enterprise ONE in Burkina Faso, and with funding from the World Bank as part of an anti-corruption and citizen engagement initiative (3LC)[1]. The research received ethics approval from the Burkinabe *Comité d'éthique Institutionnel pour la recherche en sciences de la santé* (IRB number 48-2020/CEIRES on November 5, 2020) and was registered on AEA RCT registry (ID AEARCTR-0006543, November 2020). This document presents in detail the research design and the planned analysis.[2]

2. Intervention

Burkina Faso ranks 86 out of 179 countries on the 2020 Corruption Index produced by Transparency International. Citizens perceive the problem but also maintain hope to improve it. In a survey conducted in 2016, corruption has decreased in the country for 51.9% of the respondents.[3] However, one in three respondents said that they personally experienced corruption or witnessed acts of corruption. In the ranking of services according to their degree of corruption perceived by the respondents, the municipal police, public procurement, and customs occupy respectively the first three places. A vast majority of respondents (81%) believe in the possibility to reduce corruption in their country. For them, the Government must primarily impose exemplary sanctions and thereby dissuade perpetrators. These observations inspired the design of a mobile phone application to mobilize civil society and give citizens the tools to promote transparency in common areas of public life, thereby combating corruption.

The intervention studied consists in providing beneficiary with a free smartphone App designed to help them with administrative tasks in Burkina Faso. The App functions as a 'pocket lawyer' and provides information on several administrative processes, such as obtaining a national identity card

---

[2] This pre-analysis plan was finalized after collecting baseline data on study participants, and before collecting endline information.
[3] The data come from an annual survey about corruption perceptions from the leading national anti-corruption NGO REN-LAC.

or a driving license. The information provided includes the location of the relevant administration where a given administrative process needs to be conducted, the required documents or paperwork, the duration required for the administrative process, and the monetary cost of the process. As such, the App is also designed as an instrument to address day-to-day corruption, such as bribery, which is widespread in Burkina Faso.

Indeed, the App may address bribery in several manners: (i) because beneficiaries know the exact financial cost of obtaining a document, they may avoid *faux frais* (illegitimate fee requests); (ii) by having exact information on the process, they may not require bribes to facilitate the administrative task; (iii) for the same reason, they may not have to rely on costly intermediaries – through which brides are often paid; (iv) empowered by their knowledge, they may be able or willing to resist bribery requests better.

In addition, the App is thought to directly address financial and non-financial costs– primarily saving time and stress to beneficiaries– due to transportation, document provision, waiting time, etc. In particular, beneficiaries may be able to be more successful in achieving their administrative requests, to obtain their documents faster, to visit administrations fewer times, and to travel shorter distances overall.

Our research will test the hypothesis that the App reduces people's bribery payments and improves their knowledge of and ability to navigate administrative processes.

The App is provided by the Burkinabe start-up ONE, which developed and distributed the App, with funding from the World Bank. As part of a pilot, the App was provided to 465 participants in 2019.[4] Feedback was collected by ONE and through a qualitative study led by the research team in the fall of 2020, before the quantitative study started. The App was updated based on this feedback. It is now available in French and local languages, and has a voice-responsive version (which does not require literacy). It is only available for Android smartphones at the moment. ONE is providing technical support for installing the App, as well as a social marketing campaign to stimulate its usage by participants. In addition, phone credit was provided to participants with the aim of covering costs related to the download (500 CFA or 1 USD approximatively). The App was developed as part of a broader anti-corruption initiative (3LC) based on citizen engagement. ONE is considering its scale-up by providing the App for free on the Google Play Store upon completion of the study.[5]

3. Research design

3.1 Participant recruitment

Participants were recruited by ONE from several sources. The main source is a Facebook forum which had been animated by ONE on the topic of corruption in Burkina Faso. Facebook group

---

[4] These pilot participants were excluded from the quantitative study. They were not provided the App again and not part of the treatment or control groups.
[5] Access was restricted until now to avoid any contamination risk, among other reasons.

participants, as well as other Facebook users could register for the study via WhatsApp or SMS by providing their name and phone number. ONE also recruited participants among past participants of anti-corruption training that they had conducted among university students in the past, and through a new campaign in universities. Interested students could register in the same manner or by writing their contact information on a paper form. All participants were informed that they would enter a study and be allocated either to the treatment group (receiving the App now) or to the control group (receiving the App upon conclusion of the study in several months). Participants recruited through these four channels constitute an initial pool of 3,099 applicants to be interviewed during the baseline survey. Then, individuals who accepted to answer the baseline survey would be randomly allocated to the treatment or the control group.

3.2 Experimental design

Among the pool of applicants, the research team randomly selected half of the applicants for receiving the App, while the other half will remain in the control group. Randomization was conducted on the baseline survey data, by the PI, using a computer in the office. Randomization was conducted after collection of baseline survey data, so that treatment status remained unknown to all participants and to the intervention and research teams. Randomization was conducted at the individual level from the 2,476 individuals who answered the baseline survey, resulting in 1,238 individuals allocated to the treatment group, and 1,238 to the control group. There are no clusters.

3.3 Outcomes of interest & hypotheses

We will test outcomes related to three primary hypotheses: (i) the intervention facilitates administrative processes for the participant; (ii) the App reduces the amount of bribery paid by the participant; (iii) the App helps other individuals supported by the participant in their administrative tasks. To test these hypotheses, we measure the effect of the intervention on various outcomes which we decompose into five categories of primary and secondary outcomes. For each category of outcome, we measure the effect on several outcome variables. We will adjust p-values using multiple hypotheses testing corrections within each category of outcome (see section 5.2). Variable construction is described with additional details in section 4.3.

Category A: primary outcomes related to administrative processes
- administrative tasks succeeded (#)
- time to complete administrative tasks (days)
- number of visits to administration per task (#)
- perceived ease during the administrative process (see below)

Category B: primary outcomes related to bribery
- total amount paid to the administration (log, CFA)[6]

---

[6] Null values will be converted to 1 for log calculations. Note: CFA values are large (1 USD = 500 CFA approximately).

- bribery provided for administrative tasks (0/1)
- total bribery amount for administrative tasks (log, CFA)

Category C: primary outcomes related to supporting others in administrative processes
- support of other individuals for administrative tasks (# of tasks)
- administrative tasks succeeded by other individuals supported (#)
- bribery provided by individuals supported for administrative tasks (0/1)
- total bribery amount paid by individuals supported for administrative tasks (log, CFA)

Category D: secondary outcomes related to the administrative process
- total "other" costs for administrative tasks (transportation, etc.)
- distance travelled per task (km)
- support from peer for tasks (0/1)
- time for administrative process for individuals helped with administrative tasks (days)
- feels in control with respect to administrative tasks (index)

Category E: secondary outcomes related to bribery
- use of intermediary (0/1)
- initiation of bribery demands (0/1)
- reason for asking bribes (see below)

Category F: combined outcomes for oneself and network[7]
- administrative tasks succeeded by oneself or by individuals supported (#)
- time to complete administrative tasks for oneself or for individuals supported (days)
- bribery provided by oneself or by individuals supported for administrative tasks (0/1)
- total bribery amount paid by oneself and by individuals supported (log, CFA)

4. Data

4.1 Power calculations

We used data from the pilot test of the App to calculate power for this individually randomized trial. This data was collected using a different questionnaire and approach. However, it includes information on the time spent to obtain a birth certificate (related to our category of outcomes A), and on the amount of brides paid (related to our category of outcomes B). We used these variables for our power calculation in order to determine our sample size. These calculations were conducted using the *sampsi* command in Stata, for a power of 0.8, with one round of baseline data, a relatively low correlation between baseline and follow-up variables (0.3), and using an Ancova methodology

---

[7] The category allows us to increase the number of observations in terms of administrative tasks performed. Qualitative research suggested that the App was used in the wider social network. Because the roster for individuals supported in the social network differ from the roster for oneself, we focus on a smaller number of aggregated outcomes heres.

(see section 5.1). Using the time spent to obtain a birth certification, we obtain a sample size of 2,858 for a 0.1 s.d. drop in the time spent. For observing a 0.1 s.d. change in the likelihood to pay brides, we need a sample size of 2,836, and for a drop in the average bride amount paid by 0.1 s.d., we need a sample size of 2,790. We aimed at collecting baseline information on the 3,099 participants who registered for a possible participation to the experiment. Note that the sample size required to obtain a power of 0.8 is much smaller if we aim at detecting a 0.15 s.d. change, even without any correlation between baseline and follow-up covariates (e.g. only 1,392 observations to detect a change of 0.15 s.d. in bribery amount).

4.2 Data collection

Baseline data was collected in November 2020. 3,099 participants who registered for a possible participation in the experiment were contacted by telephone by a local survey firm (see above). Of those, 2,476 individuals agreed to complete the baseline survey. The survey took about 20 minutes and was conducted by enumerators who had previously completed a 3-day training workshop. Enumerators had been advised to call each potential participant at least three times at different hours of the day to maximize response rates. Informed consent was obtained from all survey participants.

The baseline survey contained the following sections: socio-demographic background of the participant; connectivity, e.g. frequency of using the internet and network coverage at the participant's home; perception of the public administration and corruption in Burkina Faso and a roster of administrative tasks for the respondent or household members. Another administrative roster was then conducted for any other individuals that the respondent might have assisted (e.g. friends or neighbors).

For the roster of administrative tasks, participants were asked which administrative task they had undertaken in the last 12 months. Enumerators were advised not to offer any examples, but to choose the administrative tasks mentioned by the participant from a comprehensive list of the most common day to day administrative tasks. Everyday administrative tasks include items such as obtaining a birth certificate, enrolling a child in elementary school, or obtaining a national identity card. For each administrative task mentioned by the participant, a range of follow-up questions then ensued. These included questions on the duration of the process; whether or not the administrative process had been successfully completed; how often and how far the individual had to travel to complete the process; how much the individual had to pay to complete the process and whether the individual encountered any bribery demands or offers (monetary, gifts, and so on). The roster was then repeated for any administrative tasks that the individual might have assisted someone outside of their own household with. The roster thus provides a comprehensive overview of everyday administrative tasks that Burkinabé face on a regular basis and which have oftentimes been subject to demands for bribes (*pot de vin*) by low-ranking public officials.

After the baseline survey data collection was completed, individuals who had completed the survey were randomized into a treatment and a control group. Those in the treatment group received an invitation to download the App, as well as a FCFA 500 phone credit to cover the cost. Individuals

could then interact with the App for 12 months. An endline survey will be conducted in November 2021.

4.3 Data processing [desc var gen here?]

We provide more details on how we will generate variables here, including how we aggregate values at the individual level for variables collected through the roster of administrative task (see above).

Category A: primary outcomes related to administrative processes

- administrative tasks succeeded (#): count of administrative tasks succeeded by the individual from the admin roster.

- time to complete administrative tasks (days): average number of days to complete tasks started (aggregated from admin roster questionnaire).

- number of visits to administration per task (#): average number of visits for completing a task started (aggregate from admin roster questionnaire).

- perceived ease the administrative process: average subjective easiness of the administrative task, aggregated from the admin roster (question 4.1p from the baseline questionnaire).

Category B: primary outcomes related to bribery

- total amount paid to the administration (log, CFA)[8]: total aggregated from the administrative task roster; this amount includes all the fees actually paid (whether legitimate or not).

- bribery paid for administrative tasks (0/1): indicator variable equal to 1 when the individual declared that she paid any sort of bride (from the admin roster and from the direct question; 4.3b in questionnaire).

- total bribery amount for administrative tasks (log, CFA): total aggregated from the administrative task roster and from the direct question (4.3d in questionnaire). For brides in nature, individuals are asked to estimate the monetary equivalent in CFA.

Category C: primary outcomes related to supporting others in administrative processes

- support of other individuals for administrative tasks (# tasks): total of tasks recorded in the roster of administrative tasks in which the individual assisted others.

---

[8] Null values will be converted to 1 for log calculations. Note: CFA values are large (1 USD = 500 CFA approximately).

- administrative tasks succeeded by other individuals supported (#): count of administrative tasks succeeded by others from the roster of tasks supported by the individual.

- bribery paid by individuals supported for administrative tasks (0/1): indicator variable equal to 1 when the individual declared that the person assisted paid any sort of bride (from the admin roster of tasks supported by the individual).

- total bribery amount paid by individuals supported for administrative tasks (log, CFA): total aggregated from the admin roster of tasks supported by the individual).

Category D: secondary outcomes related to the administrative process

- total "other" costs for administrative tasks (transportation, etc.): total aggregated from the admin tasks roster

- distance travelled per task (km): total aggregated from the admin tasks roster; outliers will be dealt with by censoring at the 99th percentile.

- support from peer for tasks (0/1): an indicator variable equal to 1 when the individual obtained support from peers for an administrative task (from admin tasks roster).

- time for administrative process for individuals helped with administrative tasks (days): average time to complete the administrative tasks for the supported individual (from the roster of admin tasks for supported others).

- feels in control with respect to administrative tasks (index): an index related to stress and control with respect to administrative tasks computed from 4 variables asked in the endline survey.

Category E: secondary outcomes related to bribery

- use of intermediary (0/1): an indicator equal to 1 when the individual had helped from an "intermediary" in any of the administrative tasks initiated (from admin tasks roster).

- initiation of bribery demands (0/1): an indicator equal to 1 when the individual initiated any bribery request herself/himself.

- reason for asking bribes (see below): an indicator equal to 1 when the reason for asking a bride is "to speed up the process" (from admin tasks roster).

Category F: combined outcomes for oneself and network

- administrative tasks succeeded by oneself or by individuals supported (#): count of administrative tasks succeeded by the individual or the network from the two admin rosters.

- time to complete administrative tasks for oneself or for individuals supported (days): average number of days to complete tasks started (aggregated from admin roster and the network roster).

- bribery provided by oneself or by individuals supported for administrative tasks (0/1): an indicator equal to 1 when the individual or the person supported by the individual initiated any bribery request herself/himself.

- total bribery amount paid by oneself and by individuals supported (log, CFA): total aggregated from the administrative task rosters (for oneself and for the network) and from the direct question.

5. Econometric approach

5.1 Econometric specifications

Our main conclusions will be drawn from intention to treat (ITT) estimations, since these specifications are less susceptible to econometric biases. However, to gain precision and estimate parameters of interest from an academic or policy perspective, we will also try to estimate treatment for App users only, and treatment for individuals which actually engaged with the administration (aimed at conducting administrative tasks). We describe these three approaches here.

A. ITT (policy treatment, whole population)

For each outcome, we will test the effect of being in the treatment group (i.e. being offered the possibility to download the App) on the outcomes of interest. Because not all individuals downloaded, opened or used the App, this gives us "Intention to treat" (ITT) estimates. These estimates are conservative compared to the effect of the App on individuals who actually used it. However, it provides an estimation of the policy consisting in offering a free App to individuals who expressed an initial interest in the product. For the ITT specification, we estimate the following cross-sectional model at endline (t_1):

$$y_{i,t1} = \beta_0 + \beta_1 T_i + \beta_2 y_{i,t0} + \varepsilon_i \tag{1}$$

where i is the individual and T_i indicates the treatment status of the individual, and $y_{i,t0}$ is the value of the outcome of interest at baseline. This widely used ANCOVA specification is preferred since we expect a relatively small autocorrelation for most of our outcomes of interest. Thus, the power is expected to be greater compared to a difference-in-difference specification. For this specification, outcomes are all aggregated at the individual level (see section 4).

In addition to this base specification, we will perform robustness analyses adding control variables. Although balance should be obtained from randomization, control variables may address potential remaining imbalance between treatment and control groups and increase precision.

$$y_{i,t1} = \beta_0 + \beta_1 T_i + \beta_2 y_{i,t0} + \boldsymbol{\beta_k X_{k,i}} + \varepsilon_i \tag{2}$$

where $\mathbf{X_{k,i}}$ indicates k control variables at the individual level. $\mathbf{X_k}$ includes indicators for the recruitment channel of participants (see section 3.1) and individual demographic characteristics such as age, language, and network quality.

B. Treatment: Admin services users (policy treatment on relevant population)

Only a fraction of treated individuals conducted an administrative task in the past twelve months. Mechanically, the App cannot have an effect on individuals who did not interact with the administration for most of the outcomes considered. For these reasons, we explore a second series of specification for individuals who did interact with the administration. In addition to the specifications at the individual level (A.) where the information from the administrative roster is aggregated, we estimate a specification at the administrative task level, where a unit of observation corresponds to a given administrative task (see section 4). This allows us to control for the type of administrative task performed, and potentially gain precision. We estimate the following specification:

$$y_{i,j,t1} = \beta_0 + \beta_1 T_i + \mathbf{\beta_k X_{k,i}} + \mathbf{\beta_m X_{m,i,j}} + \varepsilon_{i,j} \qquad (3)$$

where i is an individual and j is an administrative process, and $\mathbf{X_{k,i}}$ indicates k control variables at the individual level and $\mathbf{X_{m,i,j}}$ m control variables at the administrative task level. This specification does not include baseline values for the dependent variable ($y_{i,t0}$) because although we have a panel of individuals (at baseline and endline), there is no panel equivalent for administrative tasks (an administrative process is performed only once). $\mathbf{X_k}$ includes indicators for the recruitment channel of participants (see section 3.1) and individual demographic characteristics such as age, language and network quality. $\mathbf{X_{m,i,j}}$ includes indicators for the type of administrative task conducted (e.g. obtaining a driving license).

C. Treatment: App users (treatment on the treated, whole population)

Given that not all households in the treatment group will have installed the App, the point estimates obtained through equation (1) and (3) will be conservative. For this reason, we will also estimate the effect of using the App through an instrumental variable (IV) approach. For equations (1) and (3), we will instrument App usage with the treatment variable and a series of covariates $\mathbf{X_k}$. This will give us an estimate of the effect of using the App. Actual App usage will be obtained from administrative records from the implementing firm (ONE).[9] For these specifications, we will either treat App usage as a dummy variable (0/1, i.e. focusing on individuals who used the App at least once), or as a continuous variable (based on the number of times that the individual used the App). Instruments will include the random assignment to the treatment group, as well as other

---

[9] We will compare this administrative information with responses from survey respondent as well, since we will ask them about App usage in the endline survey.

determinants such as connectivity (e.g. question 2.6 in baseline questionnaire about network quality in the area).

5.2 Multiple hypotheses testing & Attrition

For each category of outcomes specified in 3.3, we will correct p-values for multiple hypotheses testing. Specifically, we plan on correcting for multiple hypotheses testing by following the recent recommendations from David McKenzie on the World Bank Development Impact blog[10] and using the rwolf2 Stata command. This produces Romano-Wolf stepdown adjusted p-values, controlling for the familywise error rate (FWER). According to McKenzie, "this currently seems the theoretically best option for FWER correction at the moment".

Regarding attrition, we plan on testing whether attrition is substantial and differential between the treatment and control group. If it is, we will employ the Kling and Liebman (2004) sensitivity bounds approach to correct for attrition. In addition, also in case of differential attrition between the treatment and control group, we will explore alternative approaches to address attrition such as the one described in Behaghel et al. (2012). This approach consists in exploiting the number of calls needed to reach sampled individuals during the follow-up survey, if individuals in the treatment group are easier to reach.

5.3 Heterogeneity

We are considering several heterogeneity analyses on various sub-groups for which impacts are likely to differ. To limit the number of outcomes considered, we will focus our interpretation of the heterogeneity results on the ITT specification (1) for the list of outcomes A and B. We will consider the following subgroup analyses:

- Urban vs rural

- People using social media

- People having good internet connection

- Women vs men

- Younger vs older participants

- Education

- Language skills

---

[10] https://blogs.worldbank.org/impactevaluations/updated-overview-multiple-hypothesis-testing-commands-stata, July 20, 2021.

6. Timeline of activities

Development and pre-tests

First version of App developed : October 2019

Pilot intake data collection: November 2019

Pilot follow-up survey: March 2020

Qualitative research conducted: August-September 2020

Second version of App developed taking into account lessons learned from pilot and qualitative work: October 2020

Impact evaluation:

Recruitment of potential participants: September-October 2020

Impact evaluation baseline survey: November 2020

Randomization and beginning of the intervention: December 2020

Impact evaluation endline survey: November 2021