

Pre-analysis plan.

Insecurity in Mali, experiences and opinions: A list experiment.

Olivia Bertelli^{*†} Thomas Calvo[†] Emmanuelle Lavallée[†] Marion Mercier[†]
Sandrine Mesplé-Somps[†]

December 3, 2021

Abstract

This project aims at measuring prevalence rates of experiences and opinions related to conflict and insecurity across Mali. Together with our local partner, the GREAT Mali,¹ we conduct a List Experiment over the phone with a sample of 1400 individuals, men and women in equal proportions. Half of the sample is randomly assigned to a treatment group and the other half to a control group. To the former, we administer a questionnaire containing 5 lists, each including a sensitive item about a violence-related experience or opinion. To the latter, we administer the same lists but without the sensitive items and we ask, instead, the sensitive questions in a traditional direct form. The result of the List Experiment will provide us with an accurate measure of the rate of exposure to conflict-related violence. The comparison between the prevalence rates measured with the List Experiment (LE) and with the Direct Questioning (DQ) will provide the size of the reporting bias present in our sample. We will then explore heterogeneities in violence-reporting across individual characteristics.

^{*}Corresponding author. Email: olivia.bertelli@dauphine.psl.eu

[†]DIAL-LEDa, UMR CNRS 8007 & IRD 260, Université Paris-Dauphine, PSL Research University.

¹ <http://www.greatmali.net/>

Introduction

This plan outlines the analysis of the measurement of violence-related experiences and opinions we intend to implement on a Malian selected sample. Section 1 describes the setting of the survey. Section 2 reviews the data sources. Section 3 presents the sample selection, replacement protocol and randomization process. Section 4 enumerates the hypotheses to be tested and primary outcomes, and discusses the causal chains involved. Section 5 outlines the specifications to be used in analyzing the data, and Section 6 provides details on variables construction.

1 The setting

This research project studies the prevalence of sensitive violence-related experiences and opinions among the Malian population using a List Experiment method. In a list experiment, respondents are given a list of statements and are asked to declare how many of these statements they experienced (or agree with), without revealing which particular statements they experienced (or agree with). Respondents are randomly allocated to two groups. Control-group respondents are asked about a list of J so-called “baseline” items, while treatment-group respondents are asked about a list of $J + 1$ items, the additional item being the sensitive item that the researcher is interested in. Given the random allocation of respondents into the two groups, the prevalence rates of the baseline items should be equal across groups. The difference between the average number of items declared in the treatment group and in the control group thus provides an estimate of the prevalence of the sensitive item in the population.

Our survey contained five list experiments in link with violence-related experiences and opinions, and five symmetric direct Yes or No questions. The lists and questions are reported in Table 1. For the first four lists, respondents were asked to declare the number of items that they agreed with. For the fifth one, respondents were asked to declare how many of the listed actors they trusted. In each list, the sensitive item, which was dropped from the control group’s list, appears in italics in Table 1. Direct questions were asked only to the respondents of the control group.

2 Data sources

The data used in this work calls on three different data sources: a harmonized national household survey, the *harmonized household survey on living standards* (hereafter, EHCVM), and two different event databases, the Armed Conflict and Location and Event Data (ACLED) Project and the Georeferenced Event Dataset (GED) from the Uppsala Conflict Data Program. We used the former to draw and call a sample of Malian households and will use the latter to measure households’ exposure to violence at local level. The ACLED and GED differ by their inclusion criteria of violent events. We will present these data sources in detail in the following sub-sections.

Table 1: Lists and questions about violence-related experiences and opinions.

List experiments (LEs)	Direct questioning (DQ)
<p>1. - Over the last 12 months, I was asked a bribe. - Over the last 12 months, the police was generally efficient in my locality. - Over the last 12 months, the number of road accidents increased in my locality. - <i>Over the last 12 months, me or a member of my household was physically assaulted outside home.</i></p>	<p>Over the last 12 months, were you or a member of your household physically assaulted outside home?</p>
<p>2. - Lately, local markets have been well supplied. - There are often arguments between the inhabitants of my neighborhood. - I feel safe at home. - <i>Me or a member of my household owns a firearm.</i></p>	<p>Do you or a member of your household own a firearm?</p>
<p>3. - Roads are well maintained in my area of residence. - Generally speaking, I trust strangers. - Lately, tensions have calmed down in my area of residence. - <i>Me or a member of my household is ready to take up arms to defend our community.</i></p>	<p>Are you or a member of my household ready to take up arms to defend your community?</p>
<p>4. - The electrical network is of poor quality. - The political class properly takes care of Malians' problems. - Lately, the public health system has improved. - <i>Today's Mali must be ruled by the army.</i></p>	<p>Do you think that today's Mali must be ruled by the army?</p>
<p>5. - The police. - The public health system. - The Malian political class. - <i>The foreign armed forces in Mali.</i></p>	<p>Do you trust the foreign armed forces in Mali?</p>

2.1 The *harmonized household survey on living standards* (EHCVM)

In 2018, all eight countries of the West African Economic and Monetary Union (WAEMU) launched a harmonized survey on households' living standards with the support of the World Bank. In Mali, the National Statistics Office (INSTAT) surveyed a representative sample of households which were drawn from a traditional two-way stratified random sampling design. First, enumeration areas are randomly drawn with selection proportional to regions' population size, taking into account urban/rural distribution based on the latest available national census and the Continuous Modular household survey (EMOP). Households are then drawn from the sample of enumeration areas, and constitute a representative sample of Mali's households at national, regional and urban/rural levels. This sample comprises 6,602 households interviewed during one of the two waves which took place, respectively, between October and December 2018 and between April and June 2019.

Among other things, when it exists, contact information of surveyed households in the form of a mobile phone number is available for the household head and another household member, in general the head's spouse. It corresponds to a sub-sample of 5,835 households. This information allows us to contact a random selection of the EHCVM sample for the purpose of our experiment.

2.2 Armed Conflict and Location and Event Data (ACLED)

ACLED makes a real-time census of all political violence and protests reported with their precise location (Raleigh et al., 2010). It defines simply political violence as "the use of force by a group with a political purpose or motivation." It compiles data from newspaper articles, press agencies dispatch to non-governmental and international organisations report. One credible source of information is enough to include an event in the database.

The ACLED program covers more than 100 countries. It has devoted a particular interest in African countries for which the available data time window is the longest, starting in 1997. It registers violent political events as well as related non-violent events including troops movement, or headquarters establishment, and riots and protests. They are included in the database if they match with one of the six types of events that ACLED considers, namely: battles involving two armed groups, remote violence which are one-sided, violence against unarmed civilians, strategic development that usually take place in a context of a war, protests/demonstration, and riots.

The unit of observation is the event. If an event lasts more than a day, it is registered as many times as the number of days it lasted. ACLED identifies the precise date of each event and locates each event through its GPS coordinates. Among the available information for each event, ACLED provides the type of event, itself disaggregated into sub-event types for a precise understanding of the reported violence. All the involved groups (perpetrators/dyad, their allies or targets) are also detailed for each event. ACLED also provides an estimated number of victims without specifying the side they were on. The source used to track the event is also available as well as a brief description of the event. The latter information allow the data user to look for more precise records of every event registered in the database.

In Mali, there are 3,959 violent events identified until July 2021, 97% of which are registered after the outburst of violence in 2012. Violence increased again significantly after the signature of the Ouagadougou peace agreements in 2015. This ‘second-wave’ of violence since 2016 takes a very different shape spatially but also in its nature. While violence located first in the North of the country (Tombouctou, Gao and Kidal regions), it is now concentrated in the centre, particularly in the region of Mopti. However, the whole country has now been exposed to violent events. The number of violence perpetrated against civilians exceeds the number of events characterised as battles since 2018.

2.3 Georeferenced Event Dataset (GED)

The georeferenced Event Dataset (GED) is part of the University of Uppsala Conflict Data Program (UCDP). It is similar to ACLED data to the extent that it records all violent events around the world since 1989. More precisely, it compiles and locates all occurrences of organised violence (Sundberg and Melander, 2013). The unit of observation remains the event but it differs from ACLED by its inclusion criteria. Indeed, GED includes only lethal violence, meaning that only events with at least one estimated fatality are reported. Furthermore, the use of arms (manufactured or not) by an organised actor (formal or not) are also conditions to the inclusion. The database is not updated on real-time and is available until December 31, 2020.

Available variables enable to characterise any event with precision within an existing conflict or not, the type of the corresponding conflict (whether it is state-based or not or one-sided). It also provides the involved groups for each event. Various information on the sources of the data is also available. At least two reliable sources need to relate the same event to be included in the database. It also differentiates, when it is possible, the number of fatalities of each side.

3 Sample, replacement protocol and randomization

3.1 Sample selection

We selected our sample from a database provided by the Institute of Statistics of Mali (INSTAT) containing 5,835 households and 6,415 individuals with a telephone number. This sample is representative of households in Mali owning a mobile phone number at the national and regional levels.²

We selected a random sample of 2000 individuals among those reporting their telephone number. The selected sample has the same proportional size in each region as the main initial sample and contains men and women in equal proportion.

² 92% of households own a mobile phone number in Mali in 2018. 90% of Malian adult population owned a mobile phone in 2020.

3.2 Replacement protocol

All the remaining individuals constituted a replacement list to peak from in case the primary one contained erroneous or non-existent telephone numbers or the person did not consent to answer the survey.

We designed a replacement protocol in case the person answering the phone is not the expected one. The enumerators are instructed to conduct the interview with anyone answering the phone, provided that the person is aged above eighteen years old. The interview does not take place if the person over the phone is under eighteen years old.

The replacement protocol works as follows. Once the person picks up the phone:

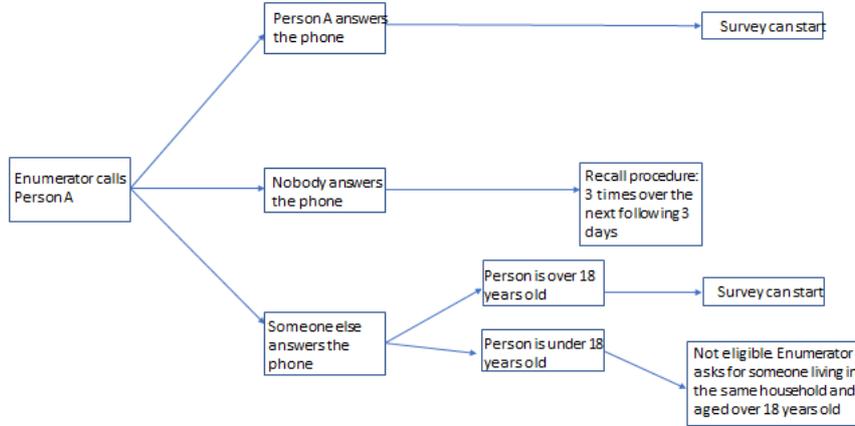
1. He/She is the person we aimed to interview → the interview begins.
2. He/She is not the person we aimed to interview. Then:
 - (a) He/She is at least 18 years old → the interview begins.
 - (b) He/She is under 18 years old → the enumerator asks to speak to someone else in the household aged at least 18 years old:
 - i. An eligible individual is available → the interview begins.
 - ii. An eligible individual is willing to answer the questionnaire but not immediately available → An appointment is set in the next days.
 - iii. There is no eligible person → the enumerator contacts the first person on the replacement list living in the same region and assigned to the same group.

The replacement protocol takes place also if:

1. Nobody picks up the telephone → the enumerator calls back in the morning, afternoon and early evening of the three following days. If no one ever picks up the telephone, then the individual is dropped from the sample and the first person on the replacement list living in the same region and assigned to the same group is contacted.
2. The contacted person does not give consent to the interview → the person is replaced with the first person on the replacement list living in the same region and assigned to the same group is contacted.

This replacement protocol possibly leads to a sample of respondents slightly different from the initially selected sample. Respondents might, in fact, be living in a different household and a different location. A set of the respondents' socio-demographic characteristics is collected by the enumerator to update the data. Note that data from the 2018 initial sample will not be available in the case of respondents being different from the ones initially selected. In this case we cannot ascertain whether the surveyed person belongs to the same household as the one initially selected.

Figure 1: Schema of replacement protocol



3.3 Randomization

Assignment to the Treatment group and to the Control group was randomly determined by the research team using the Stata software. Among the 1,000 men initially selected, 500 were assigned to the Treatment group and 500 to the Control group. The same was done separately for women. Each randomization was stratified on regions.

Individuals in the Treatment group were administrated a questionnaire with the List Experiment containing the sensitive items. Individuals in the Control group were assigned a questionnaire containing the List Experiment without the sensitive items. They were, in addition, administrated a more traditional Direct Questioning section containing the sensitive items formulated as direct questions.

All individuals in the replacement list were also randomly assigned to either groups, so that replacements were done within the same group. The replacement list was also stratified on gender and regions, as for the primary list.

The test of randomization balance will be conducted using the (few) characteristics we collected about the interviewees: age, gender, finished primary school, marital status, ethnic group, migration status and region of residence. In addition, few more characteristics at the household level measured in 2018 can be recovered from the initial database: household size, urban/rural household and household’s poverty status.

The same set of observable characteristics will be used to test whether non-response rate (in our case limited to interviewees not completing the questionnaire) is correlated with certain characteristics and, in particular, with treatment assignment.

4 Hypotheses, primary outcomes and causal chain

The hypotheses to be tested are the following:

1. Survey methods affect respondents' reporting of their violence-related experiences and opinions.
2. Individual characteristics influence the reporting differences of violence-related experiences and opinions across survey methods.
3. The prevalence of violence-related experiences and opinions varies across population sub-groups.

Our key outcomes of interest are the prevalence of five violence-related experiences and opinions. They will be measured in two ways, depending on the survey technique used. For the LE part of the survey, the prevalence rate of each sensitive item will be retrieved based on the comparison between the counts declared by respondents in the treatment and control groups. For the DQ part of the survey, they will correspond to the average answers to the five Yes/No questions.

The causal chain involved in hypotheses 1 and 2 lies upon the fact that DQ, by asking respondents to fully disclose personal information, generates declarative biases which are mitigated in the case of LE (hyp. 1), and which can vary across sub-groups of the population (hyp. 2). DQ techniques are known to suffer from measurement error due, among others, to social desirability bias and to the risk of limited confidentiality (and/or lack of trust in the surveying organization). In particular, [Tourangeau and Yan \(2007\)](#) find that misreporting in DQ is mostly due to respondents trying to avoid embarrassing themselves (social desirability bias), and to avoid repercussions from third parties, which can happen in case of lack of confidentiality. LE techniques, which do not require respondents to reveal any specific personal information, are deemed to avoid these two sources of declarative bias.³ We argue that violence-related experiences and opinions in Mali are sensitive, and thus likely to be affected by those biases and misreported by (some categories of) respondents when asked directly, while more accurately measured through LE. In [Table 2](#), we provide examples of how social desirability bias and the risk of limited confidentiality can translate into misreporting in DQ, and through which channels, in the case of violence-related behaviours and opinions.

These declarative biases are likely to be shaped by individual characteristics. In particular, we expect the following regarding the characteristics that are available in our data:

- ▷ On social desirability bias:

The region of residence, as well as gender, education, age, ethnic group and migration status are expected to influence the social norm to which individuals aim at complying; the degree to which they are attached to this norm; and the feeling of shame that they associate with certain experiences. Hence the social desirability bias arising in case of DQ is expected to vary across regions, gender, age, ethnic groups, migration status, and depending on whether the individual completed primary school or not.

³ Other sources of declarative bias, linked for instance to the misunderstanding of the questions or to the lack of memory, can exist in both types of survey methods.

Table 2: From the sources of declarative bias to expected misreporting.

Source of bias	Channel	Expected misreporting in DQ
Social desirability bias	Willingness to comply with the social norm	Under-reporting of opinions and experiences that deviate from the norm Over-reporting of socially desirable opinions and experiences
	Shame	Under-reporting of victimization
Risk of limited confidentiality	Willingness to keep information private	Under-reporting of personal experiences
	Fear of future violence and/or retaliation from third parties	Under-reporting of victimization that could signal weakness Over-reporting of experiences or opinions that could signal strength Under-reporting of politically sensitive opinions
	Hope to attract supportive public policies	Over-reporting of victimization

▷ On the risk of limited confidentiality:

Regions are characterized by various past and current levels of violence, and by various public policies. The repercussions that respondents expect in case their answers are disclosed to third parties, both in terms of possible future violence and retaliation, and in terms of supportive local public policies, should thus vary across regions.

Local violence should also affect respondents' expected repercussions in case their answers are disclosed to third parties, in particular in terms of future victimization and possible retaliation.

Respondents' individual characteristics (in particular their level of education, gender, marital status, ethnic group, migration status and age) should affect their capacity to internalize the possible consequences if their answers are disclosed to third parties and/or their willingness to keep their personal information private.

As our randomization procedure was stratified on region and gender, we will be able to measure the heterogeneity in the propensity to misreport our five violence-related experiences and opinions across these two characteristics. We will also document heterogeneity across education level, age, ethnic group, and local violence if the balance tests performed ex-post authorize it.

The causal chain involved in assumption 3 is that socio-demographic characteristics – in particular, age, gender, education, marital status, migration status, region of residence, ethnic group, the area of residence (rural/urban), household size and poverty status – as well as the level of local violence determine the individual risk of having undergone violence-related experiences, and participate in building individual violence-related opinions.

5 Specifications

To test assumption 1, we will compare the prevalence rates computed from the LEs with those computed from DQ.

The prevalence rate of sensitive item k from the DQ, \hat{g}_k , will be obtained from a regression of the direct answers on a constant:

$$Z_{k,i} = g_k, \quad (1)$$

where $Z_{k,i}$ equals 1 if individual i answers ‘Yes’ to sensitive question k , and 0 otherwise.

The prevalence rate of sensitive item k from the LE, $\hat{\gamma}_k$, will be measured based on the comparison of the control and treatment groups. In particular, we will estimate the following equation:

$$Y_{k,i} = \alpha_k + \gamma_k T_i + \epsilon_{k,i}, \quad (2)$$

where $Y_{k,i}$ denotes the response given by individual i to the list of items containing k ($Y_{k,i}$ ranges from 0 to 3 for respondents of the control group and from 0 to 4 for respondents of the treatment group), T_i is a dummy variable equal to 1 if i is in the treatment group and to 0 if i is in the control group, and $\epsilon_{k,i}$ is the error term.

Testing assumption 1 will then consist in testing whether \hat{g}_k and $\hat{\gamma}_k$ are significantly different. For each of the five sensitive questions, the difference between the LE-based and DQ-based prevalence rates can be interpreted as the size of the declarative biases that LE allows to avoid, and gives an idea of the value-added of using LE rather than DQ in terms of estimates accuracy.⁴ Moreover, analyzing the differences in the gap between LE-based and DQ-based prevalence rates across the five sensitive items may help us to understand why people misreport some information, and shed light on the existing sources of declarative biases.

Equation 2 yields an accurate estimate of the prevalence of item k provided that the assignment to treatment is random. Balance tests of randomization will be performed on the set of observed individual characteristics (gender, age, whether the individual completed primary school, marital status, region of residence, internal migration status and ethnic group proxied by the spoken language; as well as household size, poverty status and rural or urban residence for individuals who were part of the initial 2018 sample⁵). Variables for which the randomization proves unbalanced will be added as control variables in Equation 2.

To test assumption 2, we will follow the same approach over specific sub-samples, so as to compare prevalence rates by sub-groups across survey methods. This will allow us to identify whether different categories of respondents differ in their level of reluctance to reveal their experiences or opinions. We plan to compute prevalence rates by region of residence and gender and, if the randomization balance tests allow, by education level (whether primary education

⁴ An important advantage of the LE is that it mitigates respondents’ potential discomfort in answering direct questions about traumatizing experiences. On the other hand, measurement errors due to respondents’ misunderstanding of the exercise is a potential drawback.

⁵ In case a telephone number did not allow the surveyor to reach the expected person, but yielded a contact with an eligible person, we decided to proceed to the interview. Some respondents are thus absent from the initial 2018 sample.

was completed), age, ethnic group (proxied by spoken language), and local violence (relying on the ACLED and/or GED data).

To test assumption 3, we will undertake a multivariate analysis *à la* Imai (2011) on the answers obtained to the five LEs. In particular, we will estimate the following equation:

$$Y_{k,i} = \alpha_k + \beta_k X_i' + \gamma_k T_i + \delta_k X_i' \times T_i + \epsilon_{k,i}, \quad (3)$$

which interacts the treatment status in the list experiment T_i with each individual characteristics contained in a vector X_i (namely gender, age, education, marital status, ethnic group, region of residence and internal migration status). We will also include in X_i an indicator of local-level violence computed using the ACLED or GED data. Analyzing $\hat{\delta}_k$ will allow us to enlighten individual characteristics that correlate with violence-related experiences or opinions. Given that there are five different item-count questions, five regressions will be performed.

For comparison purposes, five symmetric multivariate regressions, controlling for the same vector of individual characteristics X_i , will be ran on the responses to the DQ (which were asked to the respondents of the control group only). We will estimate the following equation:

$$Z_{k,i} = g_k + \mu_k X_i' + \epsilon_{k,i}, \quad (4)$$

where $Z_{k,i}$ denotes the answer given by individual i to sensitive question k . The estimated $\hat{\mu}_k$ will provide information on which individual characteristics correlate with violence-related experiences or opinions.

6 Variable construction

The variables that will be used either to check the balance between the treated and control groups or to define individual characteristics and sub-groups will be mainly dummy variables: gender, whether the individual completed primary school, marital status, region of residence, internal migration status, ethnic group proxied by the spoken language, rural/urban place of residence and poverty status. With regard to ethnicity, the ethnic groups representing 10% or fewer observations in our sample will be grouped in a single category called "other ethnic groups". Similarly, due to the potentially low number of respondents in the regions of Kidal, Gao, Tombouktou, Ménaka and Taoudénit, these regions will be possibly grouped into one region called "North". The continuous variable informing about individual's age or household size will be transformed into categorical variables with 4 or 5 modalities.

The prevalence of violence events will be computed using ACLED or GED data. Violence events will be defined as events covering all types of "battles", "remote violence" and "violence against civilians" detailed by the ACLED typology or as the number of deaths due to violence events. Different indicators of violence events will be computed, depending on period p when the

events occurred. We will test if the effect of violence events on the magnitude of response bias due to risk of limited confidentiality (hyp. 2) depends on long term cumulative measurement of violence events or on more recent violence events. Concerning the role of exposure to violence on violence-related experiences and opinions (hyp. 3), we plan to test different variables of violence events. For instance, we hypothesize that trust in the foreign armed forces is more likely to be correlated to cumulative violence events since 2012 (first year of the presence of international armed forces), whereas the willingness to take up arms to defend the community is more likely to be correlated to more recent violence events.

Variables about the prevalence of violence will be computed at the level of district (*cercle*) or commune or using buffers around commune or district of residence. We will investigate the extensive margin of violence with the use of dummy variables: living in a place with or without violence events that occurred in period p , or living in place where the prevalence of violence events in period p is lower or higher than the national or regional median level, for instance. In addition, we will also investigate the intensive margin effect of violence-event prevalence, taking into account the number of non-null violence-events.

No imputation for missing data will be performed. We will check whether non-responses are correlated with treatment status and with sensitive questions. For each of the five sensitive issues, the non-response rates will be analyzed firstly by checking if they are statistically different between LE treated and LE control group and between LE treated group and DQ module. Secondly, we will investigate if non-responses are linked to the type of questionnaire - LE with and without sensitive items on the one hand, and LE with sensitive item and direct question on the other hand, controlling for observable characteristics⁶. Consequently, ten regressions will be run.

To more thoroughly investigate non-response, we will resort to two different methods. We will first run regressions with inverse probability weights (IPW) in the spirit of Moffit et al. (1999) that allows to control for non-response bias linked to observable variables. The model used to predict the probability is a probit with a set of socio-demographic characteristics as independent variables: sex, age, level of education, marital status, migration status, region of residence, local violent event indicators and surveyor dummies.

A second, more robust look at non-response will be performed. Following Lee (2009), we will estimate sharp bounds on the treatment effect. The assumption is that the treatment selection is monotonous, i.e. that assignment only affects non-response in one direction. The method relies on a trimming procedure so that the share of observations with an observed outcome is equal for the treated and control groups. The lower and upper bounds of the treatment effect are computed and correspond to extreme assumptions about missing information that are consistent with the observed data. In practice, the treatment arm with less non-response is being trimmed from below or above (whichever fits the data), removing the highest (or lowest) values

⁶ Sex, age, level of education, marital status, migration status, region of residence, local violent event indicators and surveyor dummies.

and assuming complete bias.

References

- IMAI, K. 2011. Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association* 106:407–416.
- LEE, D. S. 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies* 76:1071–1102.
- MOFFIT, R., FITZGERALD, J., AND GOTTSCHALK, P. 1999. Sample attrition in panel data: The role of selection on observables. *Annales d'Economie et de Statistique* 55/56:129–152.
- RALEIGH, C., LINKE, A., HEGRE, H., AND KARLSEN, J. 2010. Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research* 47:651–660.
- SUNDBERG, R. AND MELANDER, E. 2013. Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research* 50:523–532.
- TOURANGEAU, R. AND YAN, T. 2007. Sensitive questions in surveys. *Psychological bulletin* 133:859.