SOCIAL SCIENCE RESEARCH COUNCIL | WORKING PAPERS

NUMBERS UNDER FIRE: THE CHALLENGES OF GATHERING QUANTITATIVE DATA IN HIGHLY VIOLENT SETTINGS

JAVIER OSORIO

DRUGS, SECURITY AND DEMOCRACY PROGRAM
DSD WORKING PAPERS ON RESEARCH SECURITY: NO. 6

ABOUT THE PROGRAM

The Drugs, Security and Democracy (DSD) Program strives to create a stronger, more systematized knowledge base on drugs, security, and democracy in Latin America and the Caribbean; to build capacity—both institutional and individual—by supporting relevant research; and to encourage policy-relevant, evidence-based research that could lead to the development of alternatives to present-day drug policies. Support is provided for research across a variety of disciplines—anthropology, criminology, economics, history, international relations, journalism, legal studies, political science, public health, public policy, sociology, and other related fields—to create a network of scholars interested in developing alternative approaches to drug policy.

ABOUT THE SERIES

Over the last generation, activists, journalists, and researchers working in Latin America have increasingly faced the challenge of operating in areas affected by chronic police and non-state violence. Further, rising crime rates are leading a growing number of scholars to conduct research on high-risk topics, which involves gathering data on communities that experience conflict, writing and publishing on these difficult and sensitive issues, and developing and implementing programs to deal with the needs of communities affected by violence as well as the wider conflicts in which those communities are embedded. Despite these trends, the literature on safe practices for those working in high-risk environments remains thin. The DSD Working Papers on Research Security series seeks to address this deficit by examining a range of research security concerns, providing a framework to help those working in the region consider how they can enhance their own safety as well as the safety of their associates and research participants.

The DSD Program is funded by the Open Society Foundations. The program is a partnership between OSF, the SSRC, Universidad de los Andes in Bogotá, Colombia, and Centro de Investigación y Docencia Económicas in Mexico.

NUMBERS UNDER FIRE: THE CHALLENGES OF GATHERING QUANTITATIVE DATA IN HIGHLY VIOLENT SETTINGS

JAVIER OSORIO
JOHN JAY COLLEGE OF CRIMINAL JUSTICE, CITY UNIVERSITY OF NEW YORK

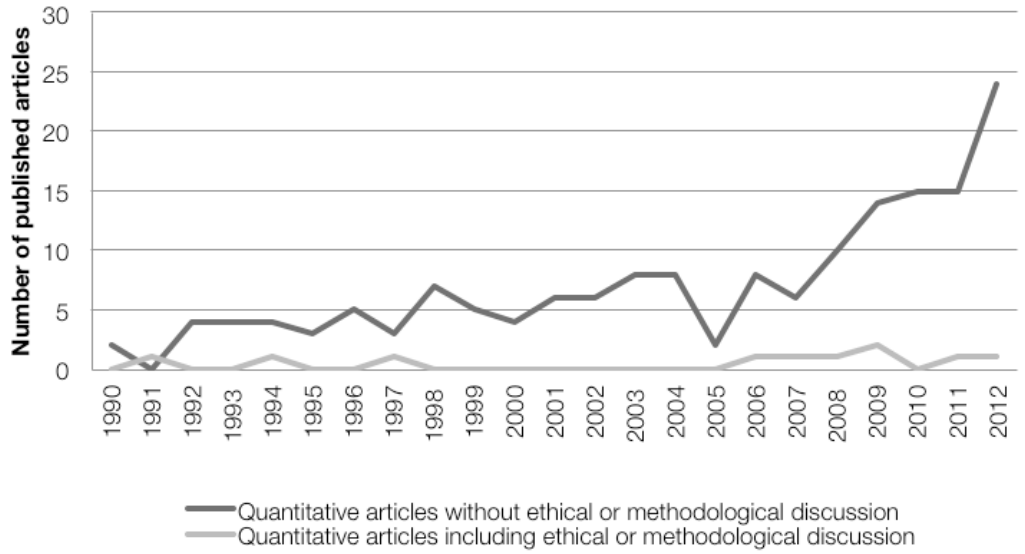
SEPTEMBER 2014

Areas torn by violence are characterized by precarious security, volatile conditions, the presence of armed actors, generalized fear among residents, traumatized combatants and civilians affected by human tragedies and material destruction, pervasive confusion about previous events, and general uncertainty about the future.[1] These factors exacerbate the ethical and methodological challenges of conducting quantitative research in violent settings. The risk of physical harm and emotional distress, both to human subjects participating in the research and to the principal investigator and research team, raises key ethical concerns, while violence increases methodological problems of bias, measure validity, and causal inference. Researchers undertaking quantitative projects in violent places can benefit greatly by taking these ethical and methodological challenges into account when designing and implementing their projects.

Researchers have devoted tremendous effort to understanding the causes and consequences of political violence from a quantitative perspective. Unfortunately, the surge of studies in this area has not included much consideration of the diverse ethical and methodological challenges associated with conducting research in violent settings. Figure 1 shows the numbers of articles published in seven political science journals between 1990 and 2012

that employed quantitative methods to analyze intrastate political violence. Of the 173 articles, only 10 included explicit discussions of these challenges. With a few insightful exceptions,[2] the quantitative literature on conflict has not engaged in ethical or methodological debates similar to those in the qualitative literature on political violence.[3] This dearth of discussion is particularly surprising, given that quantitative and qualitative research share similar challenges of validity, reliability, and inference.[4] Neglecting this concern can undermine our understanding of violence and compromise the physical and emotional integrity of both those who conduct and those who participate in such research.

FIGURE 1. Number of Quantitative Articles on Intrastate Political Violence in Selected Political Science Journals (1990–2012)



Note: The survey considered seven academic journals: *American Political Science Review*, *American Journal of Political Science*, *Journal of Conflict Resolution*, *Journal of Peace Research*, *International Organization*, *International Security*, and *Security Studies*. The articles selected used quantitative methods to analyze topics such as civil war onset and dynamics, terrorism, political violence, insurgencies, social movements, and human rights violations. Topics related to international warfare, such as interstate wars, nuclear weapons, diplomacy, or international alliances, were excluded.

The first part of this discussion proposes three general principles for conducting quantitative research in violent settings. The second discusses the need to obtain the necessary access for undertaking research activities. The third examines a set of ethical and methodological challenges associated with conducting quantitative research with human subjects, and the fourth

addresses research not involving human subjects. The fifth part outlines the basic requirements of a data management plan to help researchers in the design and implementation of quantitative projects in violent settings, and the last presents the conclusions.

KEY PRINCIPLES

Adherence to three main principles can help researchers[5] design and implement quantitative research in violent areas. The first refers to the ethical imperative "do no harm," the second recommends recognizing methodological limitations before, during, and after the research is conducted, and the third emphasizes the need for proper preparation to identify and deal with both ethical and methodological challenges.

Principle 1: Do No Harm

As in qualitative studies, the ethical imperative "do no harm" is of paramount importance in any quantitative research project, but it is especially so in those involving the participation of human subjects in areas subject to violence. Research participants may suffer physical attacks, damage to their property, coercion, sanctions, stigma, or ostracism for the simple fact of having participated in the research, and the researchers themselves may be exposed to dangerous situations during the conduct of research activities.

Informed consent and confidentiality are two basic measures usually employed to reduce the risk of physical and emotional harm to both subjects and members of the research team. Ideally, researchers would ask participants to sign an informed consent form prior to conducting the study. As mentioned by Elisabeth Wood, however, in some circumstances the existence of a signed consent form could actually endanger the subject.[6] Violent settings present a higher risk of documents, data, or any other research materials containing identifiable or sensitive information being seized or lost.[7] In such situations, it may be better to leave no trace of subjects' participation and simply request their unrecorded oral consent.

Principle 2: Acknowledge and Address Methodological Limitations

The usual methodological concerns of bias and validity characteristic of quantitative studies are more acute when data are gathered in violent settings. Obtaining access to systematic, untruncated, unbiased, and reliable data in such areas can be particularly challenging. Researchers who anticipate the challenges have a key opportunity to advance creative and effective strategies for measurement or identification that may contribute to their fields beyond what they uncover about the substantive objects of inquiry. If remedying or minimizing these concerns is not possible, researchers must still acknowledge them in their publications. Doing so does not undermine the quality of research; rather, it promotes academic standards of professionalism.

Principle 3: Remember "the Six P's"

Jerry McDermott, a noted war and crime journalist, often cites an adaptation of a famous British Army adage and refers to it as "the six *P*'s": "proper prior preparation prevents poor performance."[8] Researchers conducting quantitative work in violent areas could benefit from bearing this adage in mind throughout the design, implementation, and post-completion stages of their research. The volatile conditions of places affected by violence call for extreme care and precaution to ensure the safety, security, and well-being of all participants, whether subjects or researchers. Receiving appropriate training and securing the necessary organizational resources will help researchers meet the ethical and security challenges that may arise in the course of their projects.

ACCESS TO VIOLENT RESEARCH SITES

Quantitative projects often require the deployment of large research teams whose presence is not likely to pass unnoticed in areas torn by violence. Not having proper permission to enter the community may endanger researchers and compromise the security of research subjects. A key step in obtaining access to an area before going in is to identify which actors hold de facto control over the study site. Requesting permission could, in principle, be easier in an area controlled by a specific actor than in contested territories. In areas containing multiple armed actors, none of whom has full control over the territory, no organization can guarantee the necessary

security conditions for the researcher or human subjects. Even if the researcher manages to receive permission from one group to enter the area, there is no guarantee the rival group will respect the agreement.

A survey conducted in Afghanistan by Jason Lyall, Graeme Blair, and Kosuke Imai illustrates the challenges of gaining access to violent settings.[9] Lyall and his collaborators requested permission from both government authorities and the Taliban to enter the communities where their research was focused. In conservative districts, enumerators (also known as interviewers) were vetted by the elders based on their adherence to Islam. In some cases, having their permission did not protect the research team from threats made by younger community members.[10] In Central America, a common strategy for requesting access to gang-controlled neighborhoods is to indicate the researcher is interested in learning about the problems that most deeply affect the members of the community.[11] Sometimes enumerators have to pay a "war tax" to criminal gangs to allow them to conduct a survey in their territories.[12]

Even having permission to conduct research in dangerous areas may not be enough for coping with their unpredictability. In Mexico, a group of enumerators was kidnapped by criminals in the midst of a wave of drug-related violence affecting the country.[13] Heightened violence in certain areas may deter research teams from conducting quantitative studies and cause them to focus on qualitative research instead, as happened with the Afghan Research and Evaluation Unit (AREU).[14]

Sometimes the main risk is not to enumerators but to respondents, with strong social pressures affecting human interactions in violent settings and highlighting the need to address them to get safe access to research areas. This was illustrated by the research of Carlos Vilalta and Gustavo Fondevila.[15] For a series of surveys conducted with prison inmates in Mexico, these authors took special care to request permission from gang leaders inside the prison for inmates to participate in the study.[16]

When working in violent places, finding a person or local organization that can facilitate access to the research site may be crucial. When engaging with such local interlocutors, researchers must keep in mind the importance of reciprocity during and beyond the implementation of the project. As mentioned by Christopher Blattman, research conducted in violent settings often comes across as extractive and self-serving.[17] It makes matters

worse when, as usually happens, individuals and organizations who devote considerable time, knowledge, effort, and resources to helping researchers never hear back from them.

Researchers have many ways to reciprocate for contributions to their research. One basic way is by disseminating results in publications accessible to the population who participated in the research. This often requires translating publications into the native language and communicating ideas in a manner useful to the target audience. Researchers must, however, comply with the "do no harm" imperative and avoid endangering participants when disseminating results. Another form of repayment is to help empower individuals and organizations to develop their own professional and institutional capabilities. Researchers can offer professional training to individuals or help partner organizations improve the work they are already doing. This type of reciprocity is highly valuable, as communities affected by violence are usually located in areas where development opportunities are limited. Investing in human development through professional training could have significant and long-lasting benefits for individuals and their communities.

Access and vulnerability differ in important ways between researchers who are residents of the city or country of research and those who do not live there. Researchers residing in areas torn by violence are more likely to have effective networks that allow them access to difficult-to-reach areas. Their constant and long-term exposure to a violent environment may increase their vulnerability and that of those around them, however. Residents in violent areas conducting research in their own communities, cities, or countries should carefully assess the security conditions for themselves and their wider social and professional networks.

Nonresident researchers face different tradeoffs. As outsiders, they might need to invest greater effort into gaining access to violent areas. Depending on the nature of the project, however, they may be less exposed to aggression because they can conduct the research in a short period and leave after completing fieldwork. Nevertheless, their ethical responsibilities do not expire when they leave; they extend to the research team and the subjects who remain in the area after the project's completion. Nonresident researchers must assess the security conditions of local collaborators and participants, provide the necessary resources to minimize risks, and provide assistance when needed.

Finally, researchers must have an exit strategy that considers the security of the research team and those close to them, as well as protection of the data. Nonresident researchers may have more options and resources for planning their exits than resident researchers, whose options may be narrower and their difficulty in withdrawing themselves to safety greater, but residents are likely to have more robust support networks in place locally. Of course, all of this highlights the overriding ethical dilemma discussed by Sascha Helbardt and his collaborators—that in violent settings, most research subjects have no exit options.[18]

QUANTITATIVE RESEARCH INVOLVING HUMAN SUBJECTS

The most common type of quantitative research in the social sciences involving human subjects is the public opinion survey. In recent years, researchers also have focused increasingly on randomized experiments. Violence and insecurity increase concerns of bias and increase variance in both surveys and experimental studies. The following recommendations are primarily focused on face-to-face public opinion surveys and survey experiments, yet the analysis may be illustrative for other survey modes and experimental techniques as well.

Sampling and Randomization

The characteristics of the sampling frame and sampling selection procedures are key elements defining the quality of a survey and the prospects of expanding the conclusions of the study to the broader population of interest. Randomized treatment assignment for experimental designs is also essential for deriving causal inferences from a study. Violence may affect the availability and quality of sampling frames, as well as the implementation of sampling procedures and the reliability of random assignment to treatment conditions.

Sampling frame. Researchers rarely can reach every member of a population of interest. Instead, they select a sample representative of the overall population from which to derive inferences and then generalize the argument to the entire population. The accuracy of the sampling frame and the representativeness of the sample are key elements for supporting this generalization.

A sampling frame is a complete list of all the members of the population of study, such as census data or voter registration records. In violence-torn communities, census or voter registration data may be inaccurate or simply not exist. Violence affects the size of the population and can substantially change its religious, ethnic, gender, or age composition, quickly rendering such information sources outdated.[19] Thus, finding valid, complete, and up-to-date sampling frames for quantitative projects in violent areas may be difficult. For example, census data in Guatemala and Honduras do not reflect the current characteristics of the population, since they were gathered in 2000 and 2001, respectively.[20]

When the population is poorly characterized by existing census data or other official list-based frames, researchers may, depending on the nature of the project, consider other types of lists, such as telephone directories, electric bill records, or registration records of local organizations, government offices, or international aid agencies. For example, the sampling frame for a study on ex-combatants and noncombatants in Sierra Leone came from a list of ex-combatants registered at the National Commission on Demobilization, Disarmament, and Reintegration (NCDDR).[21] The World Bank implemented a similar strategy to assess the reintegration needs of ex-combatants of the Aceh insurgency based on the records of a pardon and amnesty program in Indonesia.[22] To study forced displacement in Colombia, Klaus Deininger, Ana Maria Ibanez, and Pablo Querubin used lists of displaced people who requested assistance from the Catholic Church.[23] Researchers may even consider using non-list sampling frames, such as high-resolution satellite images or geographical area maps.[24]

Although creative, these sampling frames can raise concerns of bias, since they may systematically exclude some sectors of the population of interest.[25] In such a case, an explicit discussion of the limitations of the sampling frame in the published results of the study or in online appendices is highly valuable.[26]

Sample selection. Based on the sampling frame, researchers select a subgroup of members of the population using either probabilistic or nonprobabilistic sampling procedures. In probability sampling, all members of the sampling frame have some known probability of being included in the sample, and their selection depends on a random procedure. The numerous probability sampling techniques available include simple or stratified random sampling; ratio, regression, and difference estimation sampling;

systematic sampling; probability proportional to size sampling; and multistage sampling.[27] As indicated by Gary Henry, random does not mean arbitrary or haphazard.[28] Rather, randomized selection should guarantee the selection of each unit of analysis is not correlated with the selection of any other unit.

In contrast, nonprobability sampling involves selecting members on the basis of availability or some systematic criterion. The most common nonprobability sampling procedures include convenience sampling; most similar / most dissimilar sampling; typical or critical case sampling; snowball sampling; and quota sampling.[29] In nonprobability sampling, an unknown segment of the population is being excluded from the sample; consequently, the conclusions cannot be generalized to the entire population.

Although ideal, probability sampling is sometimes not feasible in contexts of violence. Precarious security conditions may prevent researchers from getting access to certain areas or interviewing some individuals who were selected for the sample. Researchers should consider the possible effects of violence on the sample selection procedure and design effective strategies and instructions for replacing selected Primary Sampling Units (PSU)[30] or individuals who cannot be reached. For instance, Macartan Humphreys and Jeremy Weinstein selected ex-combatants and noncombatants in Sierra Leone using several levels of randomization.[31] First, based on the NCDDR sampling frame, the researchers selected sixty-two clusters of subjects throughout the country, which fell into forty-five chiefdoms. Then enumerators worked with a variety of government authorities, nongovernmental organizations (NGOs), and other local sources to develop comprehensive lists of ex-combatants. After identifying two or three times the targeted number of potential respondents, the enumerators randomly selected ex-combatants, using a variety of methods. Finally, the enumerators selected noncombatants using random walks around the center of each PSU.

No matter how carefully made, decisions about the sampling procedure can have consequences for the results of the research. For example, the Human Security Report Project criticized the sampling technique used by Les Roberts in his mortality study of the Democratic Republic of the Congo's civil war, which led to unjustifiably high death estimates.[32]

Randomized treatment assignment. In general terms, experimental research relies on the random assignment of subjects to different treatment

conditions, in which each subject has an equal probability of receiving a particular treatment. Treatments refer to interventions or manipulations that may affect the outcome of interest.[33] Depending on the type of study, an experiment may also include a control group that receives no treatment.

Randomized treatment allows researchers to overcome the problem of causal inference—that is, the challenge of identifying a causal connection between changes in the effect variable as a response to a change in the cause variable. According to Rebecca Morton and Kenneth Williams, random assignment must comply with the following characteristics: subjects are recruited and assigned to all the different treatments simultaneously; assignment to a treatment is independent of random assignments of other treatments and the outcome variable; all subjects comply with the treatment—that is, no one withdraws from the experiment, and all subjects follow the instructions; and we can observe all the choices of the subjects.[34] In some cases, Gary King and his collaborators recommend assigning treatment conditions according to blocked randomization, which is based on some known characteristics of the population.[35] Although this approach would lead to more efficient estimates, the dearth of reliable data may make it difficult to implement in a conflict setting.

Although researchers pay careful attention to the random assignment process, the actual delivery of randomized treatment conditions to designated subjects may be problematic. Richard Berk, Gordon Smyth, and Lawrence Sherman, for example, show how pitfalls in the randomization treatment assignment of the Minneapolis Spouse Abuse Experiment generated misleading conclusions.[36] The process may be particularly challenging in violent settings. Based on his experience conducting surveys and field experiments with ex-combatants in Burundi, Côte d'Ivoire, and Liberia, Eric Mvukiyehe explains that the validity of randomized treatments can be compromised and/or attenuated by problems of compliance (for example, subjects not taking the treatment); attrition (subjects dropping out of the study); spillovers (non-selected subjects getting indirect treatment from treated individuals); and power (too few observations). To cope with these possibilities, he recommends researchers closely coordinate with the implementing partner in the field to guarantee compliance with research protocols. He also recommends limiting knowledge of the experiment among the staff to mitigate treatment distortions or spillovers.[37]

Systematic Nonresponse and Measurement Error

Quantitative research requiring the participation of human subjects usually involves some type of face-to-face interaction between the researcher and research participants. These interactions are likely to be affected by the environment in which the study takes place, especially if it is characterized by violence. Fear is a powerful feeling that can systematically affect the responses or measures gathered from research participants. Other sources of bias in violent settings include shame, stigma, and peer pressure. As indicated by Herbert Weisberg, when respondents are asked sensitive questions under these conditions, they are likely to distort their responses.[38] Researchers must assess the ways in which a violent setting could induce systematic measurement error at the unit and item levels and try to minimize their influence.

Unit-level nonresponse. Nonresponse error at the unit level—that is, the respondent level—occurs when individuals selected as part of the sample are not interviewed. Survey reports usually include the response rate of the study, which is the proportion of selected respondents who did participate. If a survey reports a low response rate, the information gathered might not be representative of the entire population because there may be factors systematically inhibiting or preventing individuals from participating. Weisberg describes three types of unit nonresponse: incapacity, noncontact, and noncooperation.[39]

Incapacity refers to cases in which physical or mental issues prevent the designated subject from participating in the survey. In areas not affected by conflict, incapacity is usually not a methodological problem, as it is considered a rare and random phenomenon. In contrast, in violent settings, unit nonresponse due to incapacity may be a source of bias. For example, in research studying the participation of child soldiers in civil war, some of the designated individuals may suffer from severe psychological trauma and emotional distress that could prevent their participation in the study.[40] Of course, this kind of systematic unit nonresponse could affect both combatant and noncombatant victims of violence.

Noncontact refers to situations in which locating the designated respondents is not possible. This is often the case under normal circumstances when an individual is not at home at the time the interviewer reaches the selected household. In conflict settings, the impossibility of contacting subjects may

not be random. For example, a significant proportion of individuals selected for the sample may have fled because of violence, which could induce bias in the study.

Finally, noncooperation refers to instances in which designated subjects refuse to participate in the research. Precarious security conditions may instill fear in the population that prevents individuals from participating. For example, in areas heavily affected by criminal violence in Mexico, people often refuse to open the door when enumerators knock.[41] The very presence of research teams may also be sufficiently unusual and worrisome in a violent setting as to inhibit cooperation.[42] Individuals asked to participate in the study may also refuse because they are concerned about the confidentiality of their responses and the possibility of retribution from violent actors.

Receiving appropriate permission to conduct research may help increase the response rate in a violent setting, as individuals asked to participate would know it is acceptable to engage in this activity. Researchers may consider being escorted by the person or a representative of the person granting access to the community. This strategy could, however, backfire by generating perceptions of surveillance that induce noncooperation or biased responses.[43] For example, in a study on drug violence in Mexico,[44] enumerators reported being monitored by members of criminal organizations—a situation that might have affected the responses of research subjects.

In another approach to improving cooperation from subjects, researchers might consider training their enumerators, as skillful interviewers are able to raise respondents' confidence in participating.[45] Ana Arjona and Stathis Kalyvas, for instance, recruited and trained their own team of enumerators to conduct a survey of demobilized insurgents and paramilitary groups in Colombia.[46] Timothy Longman followed a similar strategy when assembling his research team to conduct a survey on postgenocide reconstruction in Rwanda.[47]

Some researchers use monetary[48] or nonmonetary[49] incentives to increase response rates. This may be a methodologically questionable approach to use in violent areas, because the harsh economic conditions people tend to experience there may prompt them to participate in the study to receive the incentive but not provide thoughtful responses. Incentives can also represent a logistical challenge for research teams carrying gifts or

money in insecure areas, and material incentives might endanger research subjects by constituting evidence of their having participated in the study.

Item nonresponse and measurement error. A frequent type of measurement error is caused by the respondent at the item (or survey question) level. This problem can manifest as item nonresponse or as error in the response process. Item nonresponse can occur when an individual agrees to participate in the study but offers "don't know" as an answer to a specific question. "Don't knows" do not induce bias if the answer reflects the respondent's genuine uncertainty. They can, however, generate biased estimates if they are systematically influenced by the environment where the study is being conducted.[50] For example, participants living in violent settings may systematically offer "don't knows" because they are afraid to express their views.

Measurement error in the response process occurs when respondents deliberately distort their answers. In violent settings, survey responses may be particularly affected by fear of disclosure and social desirability bias—that is, respondents might misrepresent their preferences because they are afraid to provide sincere answers or because they feel pressured to respond in ways that are socially acceptable. The difference between these two types of bias may depend on the intensity of the threat. Fear of disclosure may be associated with the perception of severe consequences, such as physical punishment, while social desirability bias may be related to the possibility of less severe sanctions, such as stigma, shame, or dishonor. In any case, these types of systematic measurement errors generate distorted responses and prevent researchers from identifying the true distribution of attitudes.

List and endorsement experiments have recently gained popularity as effective strategies to overcome problems of fear and social desirability bias. Instead of directly asking respondents to reveal their preferences about sensitive topics, these techniques help the researcher assess the respondent's true attitudes unobtrusively. List experiments work by asking respondents to indicate the number of elements from a list of nonsensitive items (control group) or from a list containing a sensitive item and the same nonsensitive elements (treatment group). List experiments have been used to analyze support for combatants,[51] perceptions of civilian safety,[52] presence of criminal organizations,[53] and support for torture.[54] Endorsement experiments work by asking respondents their opinions about a policy position endorsed by a sensitive actor (treatment group) or about the same policy without

mentioning such an actor (control group). Endorsement experiments have been used to assess support for insurgents[55] and for political violence.[56]

QUANTITATIVE RESEARCH NOT INVOLVING HUMAN SUBJECTS

Instead of gathering information directly from human subjects, researchers conducting some studies collect it from government documents, historical records, newspapers reports, or web-based social networks. Based on a set of predetermined coding rules, they systematize information into numerical categories for quantitative analysis. These kinds of projects are referred to here as nonparticipatory research. Despite not involving direct interaction between researchers and human subjects, nonparticipatory quantitative studies conducted in violent settings are not exempt from ethical and methodological challenges.

Modular Research Strategy

Collecting and coding data from sources other than human subjects usually requires large teams of researchers and prolonged stays in the field. In such situations, researchers may benefit from implementing a modular research strategy (MRS), which divides the investigation into two separate but related stages: information gathering and data systematization. Each stage can then be subdivided into a sequence of modules. In volatile conditions where researchers are uncertain about their ability to stay in the field until they finish their work, modular research designs allow them to have at least some discrete projects completed in the event of an emergency that forces them to leave the area.

An MRS also allows researchers to anticipate logistical and security challenges, especially for projects that may require organizing, coordinating, and managing large research teams. The need for multitasking and divided attention increases the risk of losing sight of important security measures, which could endanger the team or jeopardize the project. For this reason, an MRS may be especially helpful to reduce the risks of conducting research in violent settings. The remainder of this section addresses specific challenges to information gathering and data systematization using this approach.

Information-Gathering Stage

Researchers in violence-torn areas should try to anticipate the challenges to collecting research materials in an efficient and effective way while taking into account the risks that might arise. They also need to consider whether any sensitive information is contained in the materials they collect, as well as assess the different sources of bias in information sources and, when possible, try to minimize methodological concerns.

Information characteristics. When designing their data collection strategy, researchers should consider the type and format of information to be collected, as well as the logistical challenges of transporting, storing, and managing large volumes of documents. The use of technology such as portable scanners and high-resolution cameras reduces the burden but may also attract undesirable attention in violent areas. Scholars should emphasize researcher discretion, increase security measures, and request proper access to the area of study. Security protocols should also include appropriate measures and training to avoid health risks that can arise when precarious conditions at archival facilities lead to indoor fungal infestation and moldy research materials.[57]

Finally, the field researchers for a project should seek to ascertain what issues they may confront while gathering data before they enter the field. Combatants may have deliberately destroyed records of atrocities or entire archival facilities, and an exploratory fieldwork trip may be needed to assess the availability and quality of research materials. In cases where destruction of records has occurred, researchers must assess and acknowledge in their results the methodological problems of bias or truncation caused by incomplete information sources.

Sensitive information in research materials. Researchers conducting nonparticipatory quantitative studies in violent areas may encounter sensitive data whose use is governed by human subjects protocols. Scholars should assess the probability of damage to individuals mentioned in archival materials and how severe it might be and consider the appropriateness and importance of gathering personal identifiers that may increase the individuals' vulnerability. For example, reports and databases on sexual violence must preserve the identities of victims in complete confidentiality, because their dissemination may have deleterious consequences for them.[58] In line with the mandate to protect the identity of both victims and perpetrators of

violence and crime, the US Bureau of Justice Statistics implements different procedures, such as name-stripping, coding of data, or similar techniques.[59] Researchers might consider similar procedures to protect the confidentiality of their own data.

Gathering information about perpetrators of crimes carries particular risks for researchers. According to Jonathan Goodhand, violent actors will often not view researchers as neutral.[60] Presumed perpetrators may interpret the systematic gathering and analysis of data as a threat, and researchers should design appropriate strategies to minimize this risk. A simple way to reduce the danger is to substitute numeric codes for individual identifiers in survey questionnaires, records, data or any other research materials that may contain sensitive data. Researchers may also aggregate the data into larger units of analysis to introduce a level of generality that will help conceal subject identity. For example, instead of reporting violence perpetrated by specific individuals, researchers may report it at the organizational level; or instead of reporting daily events at the county level, they may report events monthly at the state level. Unfortunately, aggregating information at higher units of analysis may decrease the efficiency of statistical inferences and induce bias.[61] Researchers have to assess carefully the tradeoff between less precise aggregated estimates and the potential security threats associated with finer-grained data.

As indicated by Andrew March, an expert on terrorist violence, and Rich Nielsen, a specialist on the quantitative textual analysis of jihadist documents, researchers collecting and coding data may also need to take into account concerns associated with governments fighting terrorist organizations. Individuals identified by state authorities as engaged in reading, storing, sharing, or translating radical speech may suffer legal consequences of being considered potential or material supporters of terrorism.[62] Alternatively, a government may subpoena their data and seek to force them to identify research subjects. In some jurisdictions, a researcher in possession of evidence related to the commission of a crime who does not turn the information over to the government may be considered an accessory.

Bias from information sources. Nonparticipatory research projects usually rely on information extracted from a wide variety of sources. While newspapers are the predominant source for studying violence, researchers also use government documents, historical records, reports issued by NGOs, or web-based social networks. As noted by Christian Davenport and Patrick

Ball, different sources may cover the same events from varying perspectives, with important consequences for the inferences drawn from the information they provide.[63] A central concern is the problem of coverage bias. The footprints of newspapers, NGOs, and even government agencies differ across territory. Information sources are more likely to report events occurring within their areas of coverage, which are usually urban; this leads to what Kalyvas calls an urban bias in conflict research.[64] Similarly, the political orientation of an information source imprints a specific tone on the events reported. Description bias occurs when sources emphasize or downplay specific aspects of an event. For example, conservative and progressive newspapers are likely to portray events in different ways.

Researchers must carefully assess the profiles of specific sources of information.[65] A careful assessment of news sources is particularly important for conducting quantitative research in conflict areas, because the violence itself can exacerbate their ideological positioning or even suppress their freedom of expression. The best strategy for minimizing the limitations and biases of individual streams of information is to build data sets using multiple information sources. For example, the author's dissertation relied on 105 different information sources—including federal and local government agencies, as well as national and local newspapers—to gather news reports on drug-related violence in Mexico.[66]

Data Systematization Stage

After completing the information gathering, researchers usually focus on organizing and systematizing the data they have collected. Most nonparticipant quantitative research relies on teams of research assistants to code valuable information manually into predetermined categories that can be used for statistical analysis. The coding approach used in the study may require researchers to make decisions about whether to carry out the systematization stage in the field or code the information outside the violent setting.

Manual coding is the most common method for building databases in the social sciences. In conflict research, the Uppsala Conflict Data Program (UCDP) is probably the most commonly used data source on global armed conflicts. With more than twenty years of experience in systematizing events of political violence, UCDP has set the standard for designing and implementing large coding projects that rely primarily on human coders.[67]

As indicated by Frank Baumgartner, Bryan Jones, and Michael MacLeod, these types of projects may constitute a substantial challenge, as they require several activities, such as developing clear coding rules, training and managing coders, ensuring that coders are motivated to do the work properly, evaluating the quality and accuracy of sources, validating the data, updating the codebook, assessing coder reliability, and holding periodic meetings, among other tasks.[68] The substantial time, labor, and financial investment needed to build large databases using human coders make it difficult to achieve the research goals, and even harder to update or expand extant coding projects.

As a cheaper and more efficient alternative to human coding, researchers can use computer-based protocols to build databases. The seminal development of the Kansas Event Data System (KEDS) opened a research agenda using automated coding for building databases on international conflict and cooperation.[69] Later, KEDS evolved into the Textual Analysis by Augmented Replacement Instructions (TABARI) system, a more robust software package largely used for coding conflict data.[70] These projects served as building blocks for the Global Database of Events, Language, and Tone (GDELT), a coding project of unprecedented size documenting about a quarter billion events from all countries from 1979 to the present.[71] More recently, the author and Alejandro Reyes developed Eventus-ID for coding violent events from text written in Spanish.[72]

Although machine coding has some advantages over manual methods, computerized annotation is not a silver bullet. Natural language is highly complex, and computerized methods of textual analysis cannot fully capture the abstractions represented through it. This implies a tradeoff between manual and computerized coding protocols. Manual coding may be more precise in processing complex information, but it requires a substantial investment of resources. In contrast, computerized coding may be useful for less sophisticated coding tasks and can be executed in a fraction of the time and at low cost. Researchers should consider the advantages and limitations of each method.

DATA MANAGEMENT PLAN

A data management plan (DMP) is a proposal specifying the characteristics of the data that will be created in the research project and how they will be gathered, managed, stored, shared, and preserved. The creation of DMPs is becoming a standard quality practice of quantitative research in the social sciences. Since quantitative research usually requires substantial financial resources, DMPs are also becoming a key requirement in the application guidelines of funding agencies, such as the National Science Foundation (NSF).[73] Since its foundation in 1962, the Inter-university Consortium for Political and Social Research (ICPSR) has contributed to the development of standards for DMPs as well as data infrastructure and research networks on the topic.[74] Researchers might consider following those guidelines when designing their own DMPs.

According to Sarah Jones, DMPs should address several security, logistical, and methodological concerns.[75] Having clear procedures to protect sensitive information is vital for implementing the "do no harm" imperative so urgent in violent settings. Protecting information requires knowing what kinds of data are being created, who has access to them, and how information is managed, stored, preserved, shared, and disseminated. These precautions also benefit researchers, as they can help minimize the risk to those in charge of managing and securing sensitive data. DMPs further help researchers develop a strategy for organizing and managing data efficiently by providing standardized nomenclature and file versions for using, updating, and analyzing them. These regular procedures ensure the continuity of the project if researchers and assistants circulate in and out of it. DMPs also facilitate data sharing and research collaboration, and they help ensure transparency and scientific integrity when external researchers attempt to replicate or validate the results.

Several protocols and tools are available to help researchers develop DMPs. Besides the recommendations provided by ICPSR [76] and the Dataverse Network,[77] researchers may benefit from the insights of MIT Libraries[78] and the resources of the University of California Curation Center of the California Digital Library.[79] According to these sources, a basic DMP must include the following information: (1) types of data being produced; (2) data and metadata standards; (3) policies for accessing the data during the project; (4) procedures for backing up, storing, and preserving the data; and (5) policies for sharing and reusing the data.

Types of Data

The DMP must describe the kinds of data to be collected or produced in the project and give a description of their content. If data are going to be created by the researcher, the process for generating and capturing them needs to be described. If the data already exist, their source, format, and characteristics must be indicated. The DMP should contain naming conventions for files and folders and a protocol for controlling file versions. If applicable, the DMP should contain a codebook to guide the collection, systematization, and interpretation of data.

Researchers conducting quantitative projects in violent settings will benefit from carefully assessing the degree of sensitivity of each type of data being produced or gathered. The objective is to evaluate the nature, severity, and probability of risks associated with the sensitive information. A DMP for research pertaining to information from human subjects must specify how informed consent will be handled and thoroughly explain the measures that will be used to protect the privacy, confidentiality, and security of the human subjects. Finally, scholars may seek to have security protocols externally evaluated.

Data and Metadata Standards

Establishing an explicit structure for understanding the structure, meaning, semantics, and attributes of the data in a quantitative study ensures the correct use and interpretation of the data by any user. This structure is known as metadata and is particularly useful for sharing the data and collaborating with external researchers. The usual practice is to create a "README.txt" file indicating the file organization and structure of the data, explaining variables, and specifying system requirements.

Metadata specifications may be structured in many ways, but the End User Model developed by the Data Documentation Initiative (DDI) is widely used.[80] It allows researchers to identify and use relevant data at the following stages of the data "life cycle": (1) study concept, indicating key elements, definitions and concepts; (2) data collection, including questionnaires and coding instruments; (3) data processing, containing the data and specifying the content of the information; (4) data archiving, indicating procedures to guarantee the preservation of data and confidentiality; (5) data distribution, indicating the terms of use and citation; (6) data analysis, providing replication codes and

publications; and (7) data repurposing, indicating the procedures for post hoc harmonization and data transformation.

Policies for Accessing the Data

This section of the DMP describes who will have access to the information and how, at each stage of the data life cycle, specifying the level of access and data management privileges that will be accorded to each person. In violent settings, designing and implementing specific protocols for protecting, storing, and accessing sensitive data is of paramount importance. Controlling access will help reduce the risk of compromising the security of sensitive information. To do so, researchers may implement different strategies, such as using password protection, encrypted folders and files, and systems for secure data transfer and storage. Importantly, access and privilege protocols should be assessed and renewed on a regular basis.

Backing Up, Storing, and Preserving the Data

DMPs should specify the procedures and systems for managing, storing, and preserving the data, including a detailed description of the online and physical media, resources, and facilities involved. The DMP should also describe the procedures for archiving and preserving the data for the long term. ICPSR and Dataverse offer virtual infrastructure for the preservation and sharing of data. Depending on the nature of the project, the DMP should indicate whether the data gathering will take place as a one-step process (for instance, conducting a survey) or be repeated over regular intervals (such as conducting a panel survey with several waves), or if it will be maintained over time (daily media monitoring during a specific period). Research projects carrying out data collection in more than one stage should describe the process for archiving and updating the data for the various iterations.

Having an adequate system for information storage and preservation is especially important for projects conducted in violent settings, where the unpredictability of security conditions poses the risk of research data being lost, seized, deleted, or destroyed.

Policies for Sharing and Reusing the Data

The DMP should include a description of how and when the data will become available after the project is completed. It should specify the form in which the data will be accessible (for example, through data visualization and download from a website, by e-mail request, or through open access to an online repository, among other means). If applicable, the plan should specify how long the investigator will retain the right to use the data before they are released and the reasons for the embargo period (security, political, commercial, patent). Over time, the researcher should update this information.

Special attention should be paid to protecting sensitive information when releasing the data. Sometimes public access to this type of data may raise ethical and security concerns. The DMP should reflect means and procedures to guarantee anonymity and/or confidentiality; describe the restrictions for accessing, using, and disseminating the data; and explain how users will be informed about these restrictions.

CONCLUSION

This discussion contributes to the quantitative literature on violence by calling the attention of researchers to the diverse ethical and methodological challenges associated with conducting research in areas affected by violent conflict. Ethical concerns include the risk of physical and emotional harm to those participating in the research, whether as human subjects or as members of the research team. Methodological concerns are related to problems of bias, measurement error, and validity of inferences. These concerns apply both to research projects that include the participation of human subjects and those that rely primarily on documents rather than people as information sources.

Researchers should be aware of and address a variety of ethical conundrums and methodological tradeoffs throughout the design, implementation, and post-completion stages of quantitative research projects conducted in violent settings. An adequate understanding of the ethical challenges can reduce the risk of causing more damage to people who have already experienced considerable suffering, as well as risk to the research team. A clear awareness of the methodological challenges of conducting quantitative studies in conflict zones can improve our measurement tools and

our ability to derive valid causal inferences from accurate data. Ultimately, improvements in our understanding of violence can help prevent the brutality, death, and destruction inherent in violent conflict.

NOTES

1. This research was possible thanks to the support of the Social Science Research Council—Open Society Foundations. The author appreciates the insightful comments provided by Jason Lyall, Víctor Borge, Francisco Abundis, Carlos Vilalta, and the participants of the Drugs, Security and Democracy Security Protocol meeting at the Social Science Research Council, New York City, March 2013. The author thanks Fátima Ávila for her research assistance. The usual caveat applies. Contact: josorio@jjay.cuny.edu.

2. Sascha Helbardt, Dagmar Hellmann-Rajanayagam, and Rudiger Korff, "War's Dark Glamour: Ethics of Research in War and Conflict Zones," *Cambridge Review of International Affairs* 23, no. 2 (2010): 349–69; Roos Haer and Inna Becher, "A Methodological Note on Quantitative Field Research in Conflict Zones: Get Your Hands Dirty," *International Journal of Social Research Methodology* 15, no. 1 (2012): 1–13; Kristine Hoglund and Magnus Oberg, *Understanding Peace Research: Methods and Challenges* (New York: Routledge, 2011).

3. Chandra Lekha Sriram, John C. King, Julie A. Mertus, Olga Martin-Ortega, and Johanna Herman, eds., *Surviving Field Research: Working in Violent and Difficult Situations* (New York: Routledge, 2009); Susan Thomson, An Ansoms, and Jude Murison, eds., *Emotional and Ethical Challenges for Field Research in Africa: The Story behind the Findings* (New York: Palgrave MacMillan, 2013); Elisabeth Wood, "The Ethical Challenges of Field Research in Conflict Zones," *Qualitative Sociology* 29 no. 3 (2006): 373–86.

4. Gary King, Robert O. Keohane, and Sidney Verba, *Designing Social Inquiry* (Princeton, NJ: Princeton University Press, 1994); David Collier and Henry E. Brady, *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Maryland: Rowman & Littlefield Publishers, 2004).

5. In this discussion, the term "researcher" refers equally to scholars and practitioners.

6. Wood, "The Ethical Challenges of Field Research in Conflict Zones."

7. Jason Lyall, "Process Tracing, Causal Inference, and Civil War," in *Process Tracing: From Metaphor to Analytic Tool*, ed. Andrew Bennett and Jeffrey Checkel (Cambridge: Cambridge University Press, forthcoming).

8. Information provided by Jerry McDermott in a security training session for the Drugs, Security, and Democracy Fellowship Program, Bogota, Colombia, August 8, 2011. The original British Army adage has seven *P*'s. For more information, see http://en.wikipedia.org/wiki/7_Ps_(military_adage).

9. Jason Lyall, Graeme Blair, and Kosuke Imai, "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan," *American Political Science Review*, forthcoming.

10. Information provided by Jason Lyall in a personal interview, April 9, 2013.

11. Information provided by Víctor Borge, manager of a survey firm in Central America, in a personal interview, June 28, 2013.

12. Personal communication with Carlos Meléndez, December 10, 2013.

13. Information provided by Francisco Abundis, manager of a Mexican survey firm, in a personal interview, June 28, 2013.

14. Paul Fishstein and Andrew Wilder, "Establishing a Policy Research Organization in a Conflict Zone: The Case of the Afghanistan Research and Evaluation Unit," in *Research Methods in Conflict Settings: A View from Below*, ed. Dyan Mazurana, Karen Jacobsen, and Lacey Andrews Gale, 223–53 (Cambridge: Cambridge University Press, 2013).

15. Carlos Vilalta and Gustavo Fondevila, *Perfiles de la población penitenciaria I: Frecuencias y descriptivos*. [Profiles of the prison population I: Frequencies and descriptions] (Mexico: Centro de Investigación y Estudios en Seguridad—Secretaría de Seguridad Pública, 2012).

16. Information provided by Carlos Vilalta in a personal interview, April 25, 2013.

17. Christopher Blattman, "So You Want to Go to a (Post) War Zone?" International Development, Politics, Economics, and Policy Blog, March 10, 2008, http://chrisblattman.com/2008/03/10/so-you-want-to-go-to-a-post-war-zone.

18. Helbardt et al., "War's Dark Glamour."

19. Haer and Becher, "A Methodological Note."

20. Information provided by Víctor Borge in a personal interview, June 28, 2013.

21. Macartan Humphreys and Jeremy Weinstein, "What the Fighters Say: A Survey on Ex-Combatants in Sierra Leone, June–August 2003," Columbia University–Stanford University–The Post-Conflict Reintegration Initiative for Development and Empowerment, July, 2004, http://www.columbia.edu/~mh2245/Report1_BW.pdf; Macartan Humphreys and Jeremy Weinstein, "Demobilization and Reintegration," *Journal of Conflict Resolution* 51, no. 4 (2007): 531–67; Macartan Humphreys and Jeremy Weinstein, "Who Fights? The Determinants of Participation in Civil War," *American Journal of Political Science* 52, no. 2 (2008): 436–55.

22. World Bank, "GAM Reintegration Needs Assessment: Enhancing Peace through Community-level Development Programming," Annex A, "Research Methodology," World Bank, Aceh, 2006, http://www.columbia.edu/~mh2245/XCSURVEYS/ACEH_METHOD.pdf.

23. Klaus Deininger, Ana Maria Ibanez, and Pablo Querubin, "Towards Sustainable Return Policies for the Displaced Population: Why Are Some Displaced Households More Willing to Return than Others?," Households in Conflict Network, HiCN Working Paper 7, Brighton, 2004, http://www.hicn.org/wordpress/wp-content/uploads/2012/06/wp07.pdf.

24. Naresh Kumar, "Spatial Sampling Design for a Demographic and Health Survey," *Population Research and Policy Review* 26, no. 5–6 (2007): 581–99.

25. Tilman Bruck, Patricia Justino, Philip Verwimp, and Alexandra Avdeenko, "Identifying Conflict and Violence in Micro-level Surveys," Institute for the Study of Labor, Discussion Paper 5067, Bonn, 2010, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1648352.

26. Humphreys and Weinstein, "What the Fighters Say."

27. Richard L. Scheaffer, William III Mendenhall, Lyman R. Ott, and Kenneth G. Gerow, *Elementary Survey Sampling*, 7th ed. (Boston: Cengage, 2012).

28, Gary Henry, *Practical Sampling* (London: Sage, 1990).

29. Ibid.

30. James M. Lepkowski, "Primary Sampling Unit (PSU)," in *Encyclopedia of Survey Research Methods*, ed. Paul J. Lavrakas, 611-12. Thousand Oaks, CA: Sage Publications, 2008.

31. Humphreys and Weinstein, "What the Fighters Say"; Humphreys and Weinstein, "Who Fights?"

32. Human Security Report Project, "The Death Toll in the Democratic Republic of the Congo," in *Human Security Report 2009/2010: The Causes of Peace and the Shrinking Costs of War*, Human Security Report Project, Vancouver, 2010, chapter 7, 123–31, http://www.hsrgroup.org/human-security-reports/20092010/overview.aspx; Les Roberts, "Mortality in Eastern Democratic Republic of Congo: Results from Eleven Mortality Surveys," technical report, International Rescue Committee, New York, 2001, http://www.rescue.org/sites/default/files/resource-file/2001%20Mortality%20Survey.pdf.

33. The treatment is the principal variable manipulated by the researcher that is expected to have a causal impact on the outcome variable. For example, in medical research the treatment is the specific dose of a drug administered to patients that is expected to have an effect on their health.

34. Rebecca B. Morton and Kenneth C. Williams, *Experimental Political Science and the Study of Causality* (Cambridge: Cambridge University Press, 2010).

35. Gary King, Richard Nielsen, Carter Coberley, James E Pope, and AaronWells, "Avoiding Randomization Failure in Program Evaluation, with Application to the Medicare Health Support Program," *Population Health Management* 14, no. 1 (2011): S11–22, http://www.ncbi.nlm.nih.gov/pubmed/21323615.

36. Richard A. Berk, Gordon K. Smyth, and Lawrence W. Sherman, "When Random Assignment Fails: Some Lessons from the Minneapolis Spouse Abuse Experiment," *Journal of Quantitative Criminology* 4, no. 3 (1988): 209–23.

37. Eric N. Mvukiyehe, "From Design to Implementation: Forging Partnerships and Mitigating Threats to Validity in Field Experiments," 2011, http://www.columbia.edu/~enm2105/docs/liberia/From_design_to_implementation.pdf.

38. Herbert F. Weisberg, *The Total Survey Error Approach* (Chicago: University of Chicago Press, 2005).

39. Ibid.

40. Christopher Blattman, "From Violence to Voting: War and Political Participation in Uganda," *American Political Science Review* 103, no. 2 (2009): 231.

41. Alberto Díaz-Cayeros, Beatriz Magaloni, Alia Matanock, and Vidal Romero, "Living in Fear: Mapping the Social Embeddedness of Drug Gangs and Violence in Mexico," 2011, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1963836.

42. John Brehm, *The Phantom Respondents* (Ann Arbor, MI: University of Michigan Press, 1993).

43. Norman N. Bradburn and Seymour Sudman, *Improving Interview Method and Questionnaire Design* (San Francisco, CA: Jossey-Bass, 1979).

44. Diaz-Cayeros et al., "Living in Fear."

45. Robert M. Groves and Mick P. Couper, *Nonresponse in Household Interview Surveys* (New York: Wiley, 1998).

46. Ana M. Arjona and Stathis N. Kalyvas, "Preliminary Results of a Survey of Demobilized Combatants in Colombia," 2006, http://stathis.research.yale.edu/documents/Report5-06.pdf; Ana M. Arjona and Stathis N. Kalyvas, "A Micro-level Approach to Armed Organizations: Results of a Survey of Demobilized Guerrilla and Paramilitary Fighters in Colombia," 2008, http://www.anamarjona.net/docs/Arjona_Kalyvas_2008.pdf.

47. Timothy Longman, "Conducting Research in Conflict Zones: Lessons from the Great African Lakes Region, in *Research Methods in Conflict Settings: A View from Below*, ed. Dyan Mazurana, Karen Jacobsen, and Lacey Andrews Gale, 254–92 (Cambridge: Cambridge University Press, 2013).

48. Jeannine James and Richard Bolstein, "The Effect of Monetary Incentives and Follow-Up Mailings on the Response Rate and Response Quality in Mail Surveys," *Public Opinion Quarterly* 54 (1990): 346–61.

49. Eleanor Singer and John Van Hoewyk, "Does the Payment of Incentives Create Expectation Effects?" *Public Opinion Quarterly* 62 (1998): 152–64.

50. Adam Berinsky, *Silent Voices: Public Opinion and Political Representation in America* (Princeton: Princeton University Press, 2004).

51. Graeme Blair, Kosuke Imai, and Jason Lyall, "Comparing and Combining List and Endorsement Experiments: Evidence from Afghanistan," *American Journal of Political Science* (2014). Article first published online. http://onlinelibrary.wiley.com/doi/10.1111/ajps.12086/abstract.

52. Dinuk S. Jayasuriya and John Gibson, "Elephants, Tigers and Safety in Post-conflict Sri Lanka," Australian National University, Crawford School of Public Policy, Development Policy Centre Discussion Paper No 27, 2013, http://devpolicy.anu.edu.au/pdf/papers/DP-27-Elephants-tigers-and-safety-in-post--conflict-Sri-Lanka.pdf.

53. Díaz-Cayeros et al., "Living in Fear."

54. Javier Osorio, "Support for Torture: Experimental Evidence from the Mexican War on Drugs," 2014. http://www.javierosorio.net/#!research/c240r.

55. Lyall, Blair, and Imai, "Explaining Support for Combatants"; W. Bullock, K. Imai, and J. N. Shapiro, "Statistical Analysis of Endorsement Experiments: Measuring Support for Militant Groups in Pakistan," *Political Analysis* 19, no. 4 (2011): 363–84.

56. C. C. Fair, N. Malhotra, and J. N. Shapiro, "Faith or Doctrine? Religion and Support for Political Violence in Pakistan," *Public Opinion Quarterly* 76, no. 4 (2012): 688–720.

57. Beth Lindblom Patkus, "Emergency Salvage of Moldy Books and Paper," Northeast Document Conservation Center, 2013, https://www.nedcc.org/free-resources/preservation-leaflets/3.-emergency-management/3.8-emergency-salvage-of-moldy-books-and-paper.

58. Matthias Olufemi Dada Ojo, "How Rape Should Be Reported by the Victims: Opinion Survey of the Female Students of Crawford University in Nigeria," *Open Journal of Social Science Research* 1, no. 6 (2013): 128–32.

59. Laurence A. Greenfield, "Data Quality Guidelines for the Bureau of Justice Statistics," US Department of Justice, 2002, http://www.bjs.gov/content/pub/pdf/dataquality.pdf.

60. Jonathan Goodhand, "Research in Conflict Zones: Ethics and Accountability," *Forced Migration Review* 8 (2000): 12–15.

61. S. M. Shellman, C. Hatfield, and M. J. Mills, "Disaggregating Actors in Intranational Conflict," *Journal of Peace Research* 47, no. 1 (2010): 83–90.

62. Andrew March, "A Dangerous Mind?" *New York Times*, April 22, 2012, http://www.nytimes.com/2012/04/22/opinion/sunday/a-dangerous-mind.html from Rich Nielson to the author, December 9, 2013.

63. Christian Davenport and Patrick Ball, "Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977–1995," *Journal of Conflict Resolution* 46, no. 3 (2002): 427–50; Christian Davenport, *Media Bias, Perspective, and State Repression: The Black Panther Party* (New York: Cambridge University Press, 2009).

64. Stathis N. Kalyvas, "Urban Bias in Research on Civil Wars," *Security Studies* 13, no. 3 (2004): 1–31.

65. Magnus Oberg and Margareta Sollenber offer advice on evaluating news resources as sources of conflict data in "Gathering Conflict Information Using News Resources," in *Understanding Peace Research: Methods and Challenges*, ed. Kristine Hoglund and Magnus Oberg, 47–73 (New York: Routledge, 2011).

66. Javier Osorio, "Hobbes on Drugs: Understanding Drug Violence in Mexico," 2013, http://media.wix.com/ugd/76b255_675f00d2a50ceebfd89750ae3b6969ae.pdf.

67. Ralph Sundberg and Lotta Harbom, "Systematic Data Collection: Experiences from the Uppsala Conflict Data Program," in *Understanding Peace Research: Methods and Challenges*, ed. Kristine Hoglund and Magnus Oberg, 91–113 (New York: Routledge, 2011).

68. Frank Baumgartner, Bryan Jones, and Michael MacLeod, "Lessons from the Trenches: Ensuring Quality, Reliability, and Usability in the Creation of a New Data Source," *Political Methodologist* 8, no. 2 (1998): 1–10.

69. Philip A. Schrodt, Shannon G. Davis, and Judith L. Weddle, "Political Science: KEDS—A Program for the Machine Coding of Event Data," *Social Science Computer Review* 12, no. 4 (1994): 561–87.

70. Philip A. Schrodt, "TABARI: Textual Analysis by Augmented Replacement Instructions," 2014, http://eventdata.parusanalytics.com/tabari.dir/TABARI.0.8.4b2.manual.pdf.

71. Kalev Leetaru and Philip A. Schrodt, "GDELT: Global Data on Events, Location and Tone, 1979–2012," 2013, http://data.gdeltproject.org/documentation/ISA.2013.GDELT.pdf.

72. Javier Osorio and Alejandro Reyes, "Eventus ID: Supervised Event Coding from Text Written in Spanish," 2014, http://www.javierosorio.net/#!software/cqbi.

73. National Science Foundation, "NSF Data Management Plan Requirements," 2013, http://www.nsf.gov/eng/general/dmp.jsp.

74. Inter-university Consortium for Political and Social Research, "Guidelines for Effective Data Management Plans," 2013, http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/index.html.

75. Sarah Jones, "How to Develop a Data Management and Sharing Plan," DCC How-to Guides. Edinburgh: Digital Curation Centre, 2011, http://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How%20to%20Develop.pdf.

76. Inter-university Consortium for Political and Social Research, "Guidelines."

77. The Dataverse Network Project, "Template for Data Management Plan," 2013, http://thedata.org/book/data-management-plan-template.

78. MIT Libraries, "Data Management Plans," 2013, https://libraries.mit.edu/guides/subjects/data-management/plans.html.

79. University of California Curation Center of the California Digital Library, "DMPTool," 2013, https://dmp.cdlib.org.

80. Karl Dinkelmann, Michelle Edwards, Jane Fry, Chuck Humphrey, Ron Nakao, and Wendy Thomas, "Best Practices," Data Documentation Initiative, 2009, http://dx.doi.org/10.3886/DDIBestPractices02.

ABOUT THE AUTHOR

Javier Osorio is an assistant professor in the Department of Political Science at John Jay College of Criminal Justice at the City University of New York. He obtained his PhD in political science from the University of Notre Dame. In general, his research interests are located at the intersection of political violence, quantitative research methods, and natural language processing. In particular, his academic agenda is focused on understanding the causes and consequences of large-scale criminal violence in Latin America with a special emphasis on Mexico. Osorio is the codeveloper of Eventus ID, a software application for automated coding of event data from text written in Spanish. To conduct his research, he has received support from grants and fellowships provided by the National Science Foundation, the Social Science Research Council–Open Society Foundations, the United States Institute of Peace, the Harry Frank Guggenheim Foundation, the University of Notre Dame, Yale University, and Cornell University.