Using quantitative data to study feminicide: challenges and opportunities

Maria Gargiulo*

30 April 2021

Summary

The growing availability of quantitative data on feminicide[1] presents new opportunities to study fundamental questions that have previously gone unanswered due to a lack of data. It also opens doors to expand the scale of inquiry from micro-level case studies towards macro-level analyses that fit individual cases into broader patterns of violence.

This report is organized into two main sections. The first outlines three difficulties practitioners will likely encounter when working with feminicide data: (1) despite the growing availability of data, feminicide is still fundamentally a missing data problem; (2) feminicide is difficult to measure due to lack of contextual information and shared definitions among the groups collecting the data; and (3) data on feminicide will likely be subject to selection biases causing data to be missing non-randomly, which can distort analyses over time, space, or other attributes. To address these challenges, we make two technical suggestions: (1) groups should collect data that is as disaggregated as possible; and (2) we should use statistical methods such as Multiple Systems Estimation to model what is missing from the data before examining patterns over time, space, or other attributes.

*Statistician, Human Rights Data Analysis Group, mariag@hrdag.org

[1] Throughout this report we choose to use the term "feminicide" proposed by Lagarde y De Los Ríos (2010), rather than the term "femicide" introduced by Radford and Russell (1992). We choose this terminology to situate these killings as part of the broader set of violations of the human rights of girls and women, including those that result from impunity. Following Lagarde y De Los Ríos (2010), when we use the term "feminicide", we understand the state to have a role in enabling this lethal violence.

Challenges of quantifying feminicide

Feminicide as missing data

Following D'Ignazio and Klein (2020), a useful way to understand feminicide is through the framing of "missing data." Much like many feminicide victims' bodies are missing, data about the circumstances of their deaths are missing too. Even in places where feminicide is explicitly criminalized, such as many countries in Latin America and the Caribbean (LAC), governments have largely failed to document and publish feminicide in official statistics. Existing administrative sources are often incomplete, updated infrequently, and lack the necessary contextual information to determine whether a particular homicide should be classified as a feminicide. Furthermore, reporting feminicide is stigmatized and dangerous in many contexts, and cultures of impunity often prevent proper investigations when feminicides are reported.

To fill the void in official statistics, civil society groups have begun assembling their own datasets of feminicide victims, often using media reports or crowdsourcing techniques.[2] These new datasets have surpassed the utility of official government data in many contexts. For example, in Mexico, "the most accurate records of feminicide are still kept by individuals, researchers, and journalists, rather than by the police or a state or federal institution" (Driver 2015, 15).

The historic lack of data means that many fundamental questions about feminicide have gone unanswered: How many feminicides have occurred? Where and when have they happened? Who were the perpetrators? While these questions speak to complex problems, quantification is at their core. Quantification alone cannot end feminicide; it cannot fully address past harms or single-handedly create more just societies. However, quantifying the magnitude of the problem and understanding patterns of violence are essential to the development of effective public policy to prevent future feminicides.

In particular, quantification provides at least three distinct use cases with applications to policy. First, having robust statistics about the magnitude and scope of feminicides makes the problem more difficult to ignore at the local, national, and international level. Second, quantification allows us to identify populations that are at higher risk for feminicide perpetration or victimization. This is an important consideration for the optimization of resource and service allocation and particularly important for situations where resources are heavily constrained. Finally, quantifying the problem allows for the evaluation of policies or interventions aimed at reducing violence. How can we know a policy is effective at reducing feminicides if we do not have an understanding of the size of the problem before and after the policy was instituted? As new data sources become available, we find ourselves closer to answers to these questions.

Feminicide and measurement error

While the quantification of feminicide may be a budding area of inquiry, the quantification of homicide is not. Many different methods have been used to quantify homicide in a variety of contexts, including excess death calculations, retrospective mortality studies, and multiple systems estimation. These same methods may also be useful for studying feminicide, but applications to feminicide introduce new measurement obstacles that are rarer in general homicide data. These issues stem from the difficulty in determining whether a particular killing should be considered

[2] See for example, María Salguero's map of feminicides in Mexico and Helena Suarez Val's project Feminicidio Uruguay.

a feminicide, especially in the absence of contextual information about the circumstances of the killing, knowledge of the perpetrators' motives, or a unified global taxonomy of feminicide.

This difficulty in classification has parallels to other subtypes of homicide, such as deaths of civilians in armed conflict. How do we know if someone was a civilian? Is it that they were wearing plain clothes at the time of their death? That they were killed in a certain location or with a certain weapon? That they do not appear on a list of known combatants? Different institutions might use different criteria to determine who they believe was a civilian, resulting in different conclusions of the burdens of the conflict on the civilian population—a task made more difficult when contextual information is missing from the available data. We can understand this variability in classification as a form of measurement error.

One seemingly common form of measurement error that arises with feminicide occurs when social consensus about what constitutes feminicide differs from how the crime of feminicide is outlined in penal code. For example, in Nicaragua, a female homicide can only be considered a feminicide in the context of an intimate relationship (El Presidente de la República de Nicaragua 2014b, 2014a). This legal definition is much more stringent than broad social consensus in LAC. In this case, datasets documenting feminicides produced by government entities would likely exclude many cases of violence that would otherwise be included in datasets constructed by civil society groups.

This same issue of definitional differences appears again when comparing penal code across the countries in LAC where feminicide is criminalized. Each law uses different language to delineate what constitutes a feminicide ("Femicide and International Women's Rights" 2017). Some of these laws are compatible with each other in the sense that they apply to similar types of feminicide, but none are identical. These different definitions then result in different measurements. Then, when we use these measurements to compare feminicide prevalence across the region, we compare these completely different quantities as if they are the same—like comparing an apple to an orange—and draw possibly erroneous conclusions about where violence is better or worse. We might imagine that a country that defines feminicide very narrowly in the penal code could appear to have much a much lower feminicide prevalence than a country that defines feminicide more broadly, even if the underlying rates at which women and girls are killed because of their gender are the same.

Selection bias in feminicide data

Selection bias arises in the data collection process when data is missing non-randomly, resulting in a sample that is not statistically representative of a particular population of interest (e.g., feminicide victims in Mexico in 2018). Framing feminicide as a missing data problem also requires acknowledging the potential sampling biases in the data that is collected. Whose deaths are recorded in our databases, and perhaps more importantly, whose are not? If selection bias is large, conclusions drawn from available data may not describe the true reality of the problem. Civil society data collection on feminicide is an essential step towards understanding the problem in the context of incomplete or unreliable official government data. However, these data are often only snapshots—they tell true, but partial narratives about the violence.

The problem of selection bias is not unique to feminicide data. Studies of large-scale violence seldom have access to complete datasets documenting what happened. In the context of homicide, would-be witnesses might face stigmatization or retribution for reporting, the violence may leave no surviving witnesses, or a body might never be identified. Data collection might be limited to urban areas that are more easily accessible, administrative capacities may limit the quantity of data that can be collected, or data might be outright manipulated to be made consistent with a

particular political objective. News media reports, which form the source materials for many of the lists of feminicide victims kept by journalists, activists, and civil society groups, are not immune to selection bias. News reports may also suffer from urban biases, may systematically exclude certain populations (e.g., indigenous women or sex workers), and may choose to present only certain types of feminicides due to the agendas of publishers.[3]

Official data collected by governments often tells a different story about feminicide than data collected by activists. For example, in 2018, María Salguero's map documents 2,251 cases of feminicide across Mexico; official government data records 914.[4] Some differences are likely explainable through differences in measurement—María Salguero uses a broad social conceptualization of feminicide, whereas many state governments in Mexico only count a case as feminicide in official statistics if a case is opened with the prosecutor's office—still others will arise from the various forms of selection bias that the different sources are subject to. Each dataset likely documents some unique cases that the other does not, but the overlap cannot currently be known due to the lack of available microdata on feminicide victims from the government. Even so, these different datasets tell different stories about the magnitude and location of feminicide across the country. Choosing to believe one source over another has implications for policy-making, historical memory, and justice.

Rather than considering a single, incomplete and non-statistically representative source of feminicide data as the "truth", robust quantitative analyses should include information from multiple sources and then estimate the number of deaths that were not recorded in the available data to mitigate the impacts of selection bias in the source data. Understanding patterns of missingness in the data about feminicides is essential to answering the questions about the violence posed by civil society groups, policy makers, and international organizations. Failing to account for missing data in our analyses results in wrong answers to those questions.

Opportunities

Collecting disaggregated data

To respond to the difficulty in measuring feminicide, we need to adjust our data collection practices. We should aim to document all homicides with as much detail and disaggregation as possible so that feminicides can be identified in the data. Using this approach, cases that share attributes (e.g., signs of sexual violence, perpetrator was a co-worker) can be aggregated for study. It is not possible to investigate specific types of violence when they are grouped together under broad categories like "female homicide" without other contextual markers. Rather than pre-specifying a particular definition of feminicide, the disaggregation approach would allow researchers to identify cases consistent with their working definitions of feminicide. Additionally, the collection of disaggregated data would facilitate more rigorous comparisons of prevalence without the need for full agreement on a specific definition of what should be considered a feminicide across different data collectors or national contexts.

Efforts to prevent feminicide will be most effective if they can be tailored to the specific harms experienced by victims and their communities. For example, interventions in response to sex-selective foeticide will likely differ from those responding to targeted killings of women due to

[3] See Table 1 in Dawkins (2020) for examples of media biases in newswire data on conflict-related casualties in South Sudan. Although a different context, many of the same biases may be applicable to news reporting on feminicides.

[4] 2018 data from María Salguero's map is available at https://feminicidiosmx.crowdmap.com/. Mexican Government data is available at: https://drive.google.com/file/d/1iD7VsognaA9Lt40gzgz977GkHGbGenmL/view (p. 1).

organized crime because the underlying drivers of violence differ. Disaggregated data has the potential to tangibly impact the development of these policies and interventions by describing the characteristics specific to different forms of violence.

The Latin American Open Data Initiative's (ILDA by its Spanish acronym) guide on systematizing feminicide identification and registration provides an example of collecting disaggregated feminicide data in the Argentine context (Fumega 2019). Among other information, their protocol suggests standardized variables for recording information about the victim, the accused, and the event leading to the death. The adoption and use of protocols like ILDA's will be critical for the generation of better official data on feminicide, as well as the harmonization with data from other sources, like lists of victims maintained by civil society groups.

The collection of disaggregated data also offers an opportunity to use sensitivity analysis to examine whether legal conceptualizations of feminicide are sufficiently covering the violence experienced by the population. If our conclusions about the impacts of feminicide change drastically based on the specification of what is considered a feminicide, this might be cause to re-evaluate whether current policy, justice, and accountability efforts are adequate to address past harms and prevent future violence. Additionally, sensitivity analyses might help to identify specific types of violence that may be overlooked by current laws and policies.

Estimating undocumented feminicide victims

Multiple systems estimation (MSE, also known as capture-recapture) is one statistical method that could be applied to study feminicide data. In the context of lethal violence, MSE models use multiple incomplete lists of victims—like the ones being assembled by civil society groups—in order to estimate the likely number of deaths that were not observed by any of the sources. This method has previously been applied to study patterns of lethal violence in other contexts (e.g., Ball et al. (2001), Ball, Rodríguez, and Rozo (2018), Ball et al. (2019), Hoover Green and Ball (2019)), but has not been used to study feminicide largely due to the difficulty in obtaining data. As more and higher quality data on feminicide becomes available, MSE has the potential to clarify crucial questions of fact necessary to understand the impacts of feminicide.

Chao (2001) offers a technical introduction to MSE, while Bird and King (2018) presents an overview of how MSE has previously been used to study human populations and inform public policy. The intuition behind the method is the following:[5] imagine two dark rooms. We would like to know their sizes, but we cannot see inside the rooms. The only tool we have to investigate the size of the rooms are rubber balls. These rubber balls have a particular property that they do not make any sounds when they hit the walls, floor, or ceiling of the room, but when two or more balls collide, they make a clicking noise. When we throw the rubber balls into the first room, we hear many clicks. We then repeat this action in the second room, using the same number of balls and the same amount of force as we did the first time. This time we hear fewer clicks and they are less frequent. In this case, our intuition would be that the second room is larger than the first. The balls spread out more and collided less frequency because the room had more space.

Translating this analogy back to the context of feminicide, the size of the dark room is the size of the population of feminicide victims that we want to estimate and we are "throwing" the different data sources at the population of victims. When two or more sources document the same victim, it is as if they "collided" making a clicking noise. We then use these documentation pattern, counting

[5] HRDAG uses this analogy in many reports that use MSE.

the number of datasets every victim appears in, to estimate the total number of victims, including those that were never documented in our sources.

Using MSE methods, we can draw more accurate conclusions about the number of feminicides that have occurred—both those that have been documented and those that have not. Additionally, using these estimates allows for statistically valid comparisons over time and space, allowing us to identify areas where more or less violence has occurred and to establish baselines for evaluating the effectiveness of interventions aimed at violence reduction. Additionally, if demographic information about victims or perpetrators, or contextual details surrounding the death are available, we can also use estimates to explore hypotheses about patterns of victimization or perpetration.

Though many challenges exist, it is encouraging to see growing attention among academic, government, and civil society organizations on this important problem. We look forward to working alongside practitioners and scholars to use quantitative data to advance our understanding of feminicide and inform policy discussions about violence prevention.

Acknowledgements

We thank Dr. Megan Price for her review of this report and Dr. Elisabeth Wood for her comments on an earlier draft.

About HRDAG

The Human Rights Data Analysis Group is a non-profit, non-partisan organization[6] that applies scientific methods to the analysis of human rights violations around the world. This work began in 1991 when Patrick Ball began developing databases for human rights groups in El Salvador. HRDAG grew at the American Association for the Advancement of Science from 1994–2003, and at the Benetech Initiative from 2003– 2013. In February 2013, HRDAG became an independent organization based in San Francisco, California; contact details and more information is available on HRDAG's website (https://hrdag.org).

HRDAG is composed of applied and mathematical statisticians, computer scientists, demographers, and social scientists. HRDAG supports the protections established in the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and other international human rights treaties and instruments. HRDAG scientists provide unbiased, scientific results to human rights advocates to clarify human rights violence. The human rights movement is sometimes described as "speaking truth to power:" HRDAG believes that statistics about violence need to be as true as possible, with the best possible data and science.

The materials contained herein represent the opinions of the authors and editors and should not be construed to be the view of HRDAG, any of HRDAG's constituent projects, the HRDAG Board of Advisers, the donors to HRDAG or to this project.

References

Ball, Patrick, W Betts, F Scheuren, J Dudukovich, and J Asher. 2001. "Political Killings in Kosovo, March-June 1999." American Association for the Advancement of Science. https://hrdag.org/wp-

[6]Formally, HRDAG is a fiscally sponsored project of Community Partners, see https://communitypartners.org/.

content/uploads/2015/07/kosovo-politicalkillings-2000-cc.pdf.

Ball, Patrick, Sheila Coronel, Mariel Padilla, and David Mora. 2019. "Drug-Related Killings in the Philippines." Human Rights Data Analysis Group. https://hrdag.org/wp-content/uploads/2019/07/2019-HRDAG-killings-philippines.pdf.

Ball, Patrick, César Rodríguez, and Valentina Rozo. 2018. "Asesinatos de Líderes Sociales En Colombia En 2016–2017: Una Estimación Del Universo." Dejusticia. https://cdn.dejusticia.org/wp-content/uploads/2018/09/Asesinatos-de-l%C3%ADderes-sociales-en-Colombia-en-2016-2017-una-estimaci%C3%B3n-del-universo.pdf.

Bird, Sheila M., and Ruth King. 2018. "Multiple Systems Estimation (or Capture-Recapture Estimation) to Inform Public Policy." *Annual Review of Statistics and Its Application* 5 (1): 95–118. https://doi.org/10.1146/annurev-statistics-031017-100641.

Chao, Anne. 2001. "An Overview of Closed Capture-Recapture Models." *Journal of Agricultural, Biological, and Environmental Statistics* 6 (2): 158–75. https://doi.org/10.1198/108571101750524670.

Dawkins, Sophia. 2020. "The Problem of the Missing Dead." *Journal of Peace Research*. SAGE Publications Sage UK: London, England. https://doi.org/10.1177/0022343320962159.

D'Ignazio, Catherine, and Lauren F Klein. 2020. *Data Feminism*. MIT Press.

Driver, Alice. 2015. *More or Less Dead: Feminicide, Haunting, and the Ethics of Representation in Mexico*. University of Arizona Press.

El Presidente de la República de Nicaragua. 2014a. "Decreto No. 42-2014." http://legislacion.asamblea.gob.ni/Normaweb.nsf/%28%24All%29/3387522EDDAD4A2F06257D3A00768A5A.

———. 2014b. "Ley No. 779." http://extwprlegs1.fao.org/docs/pdf/nic138659.pdf.

"Femicide and International Women's Rights." 2017. *Global Americans*. https://theglobalamericans.org/reports/femicide-international-womens-rights/.

Fumega, Silvana. 2019. "Guía para protocolizar procesos de identificación de femicidios para su posterior registro." Zenodo. https://doi.org/10.5281/zenodo.1659629.

Hoover Green, Amelia, and Patrick Ball. 2019. "Civilian Killings and Disappearances During Civil War in El Salvador (1980–1992)." *Demographic Research* 41: 781–814. https://doi.org/10.4054/DemRes.2019.41.27.

Lagarde y De Los Ríos, Marcela. 2010. "Preface: Feminist Keys for Understanding Feminicide." In *Terrorizing Women*, xi–xxvi. Duke University Press.

Radford, Jill, and Diana EH Russell. 1992. *Femicide: The Politics of Woman Killing*. Twayne Pub.