

Selection bias and the statistical patterns of mortality in conflict¹

Megan Price* and Patrick Ball

Human Rights Data Analysis Group, San Francisco, CA, USA

Abstract. This paper explores how information is generated about killings in conflict, and how the process of generation shapes the statistical patterns in the observed data. The difference between the observed patterns and the true patterns is called bias, two examples of which will be examined. First, we compare multiple individual sources reporting identifiable killings in Syria, highlighting variations in the likely probabilities of reporting for events of different sizes. Second, we conduct a similar analysis examining the number of sources reporting events of varying sizes in the Iraq Body Count public dataset. In both cases we explore how depending on the observed data without accounting for bias caused by missing data could mislead policy. The paper closes with recommendations about the use of data and analysis in the development of policy.

Keywords: Bias, casualty counting, human rights, missing data

1. Introduction

Emerging technology has greatly increased the amount and availability of data in a wide variety of fields. In particular, the notion of “big data” has gained popularity in a number of business and industry applications, enabling companies to track products, measure marketing results, and in some cases, successfully predict customer behavior [1]. These successes have, understandably, led to increased excitement about the potential to apply these and other quantitative methods in an increasing number of disciplines.

Although we share this excitement about the potential power of data analysis, our experience over the past 25 years analyzing data about conflict-related violence motivates us to proceed with caution. The data available to human rights researchers is fundamentally different from the data available to business and indus-

try. The difference is whether the data are complete. In most business processes, an organization has access to all the data: every item sold in the past 12 months, every customer who clicked through their website, etc. In the exceptional cases where complete data are unavailable, often industry analysts are able to generate a representative sample of the data of interest.

In our analyses of mass violence in Guatemala, Kosovo, Perú, Colombia, and Timor-Leste, we have repeatedly found that the amount of missing data can be surprisingly large: in the chaos and fear that surrounds conflict, violence often goes unreported and consequently victims remain hidden from view. As statisticians supporting human rights analysts, we have found that understanding patterns of missing data about violence is crucial to answering the questions that are brought to us by human rights groups, international tribunals, and truth and reconciliation commissions. Failing to account for missing data means getting the answers to those questions wrong.

In human rights, and more specifically in studies of conflict violence, we rarely have access to complete data. What we have instead are snapshots of violence: a few videos of public killings posted to YouTube, a particular set of events retrospectively recorded by a truth commission, stories covered in the local or inter-

¹Parts of this article have been excerpted with permission from SAIS, where they were originally published as part of a different article, see M. Price and P. Ball, Big Data, Selection Bias, and the Statistical Patterns of Mortality in Conflict, SAIS Review of International Affairs, Volume 34, Issue 1, Winter-Spring 2014, pp. 9–20.

*Corresponding author: Megan Price, 109 Bartlett St Suite 204, San Francisco, CA 94110, USA. E-mail: meganp@hrdag.org.

national press, protesters' SMS messages aggregated onto a map, or victims' testimonies recorded by non-governmental human rights organizations (NGOs) are typical sources. Statistically speaking, these snapshots are convenience samples, and they cover an unknown proportion of the total number of cases of violence. It is mathematically difficult, often impossible, to know how much is undocumented, and consequently missing from the sample.

Worse, in most contexts, the visible cases differ in fundamental ways from the unseen cases, and the pattern of seen vs unseen is often related to substantive questions of interest. The challenge is that researchers, and advocates, naturally want to address questions that require either the total number or a representative subset of cases of violence: How many people have been killed? What proportion were from a vulnerable population? Were more victims killed last week or this week? Which perpetrator(s) are committing the majority of the violence? Answering these types of questions, and basing policy decisions, on naïve analyses of such snapshots can prove to be dramatically misleading. This is not a criticism of the data – in most cases data were not collected with the intention of addressing these kinds of questions, and collecting complete or representative data under conflict conditions is generally impossible. These concerns should not deter researchers from asking questions of data, rather it should caution them against basing conclusions on naïve analyses. We hope this discussion will generate ideas about appropriate statistical models to address data limitations.

2. The problem of bias

Nearly all samples are partial, and in samples not collected randomly, the patterns of omission may have structure that influences the patterns observed in the data. We have seen countless examples in our own work: violence in urban areas may be more visible to conventional media sources; violence committed by the state may be more likely than insurgent violence to be reported in truth commission testimonies; violent events that kill children may be more likely to be remembered and recorded relative to violence against adults [2].

These are all examples of “selection bias,” the phenomenon through which some events, due to the characteristics of the event, are more likely to be “selected” for the non-random sample than other events. This is a

problem in general, but the implications become particularly costly when the characteristics that make an event more or less likely to be selected are correlated with (or perhaps are exactly the same as) the characteristics under study. For example, in our work for the truth commission in Perú in 2003, we found that killings attributed to the government had a much higher probability of documentation than killings attributed to the Shining Path insurgency, yet questions of accountability hinged precisely on determining which group perpetrated the majority of the violence [3]. A naïve analysis of the observed data, without accounting for selection bias, would have incorrectly held the state responsible for a larger proportion of the violence.

There are many ways that selection bias can affect human rights data collection [4]. In this article, we focus on a particular kind of selection bias called event size bias. This kind of bias addresses the situation in which events that involve only one victim are less likely to be documented than events that involve larger groups of victims. For example, a market bombing may involve the deaths of many people. The very public nature of the attack means that the event is likely to attract extensive media attention, and consequently, it will be covered by many media organizations. By contrast, the assassination of a single person, at night, by perpetrators who hide the victim's body may be unreported by anyone. The victim's family may be too afraid to report the event, and the body may not be discovered until much later, if at all. Event size bias is the variation in the probability that a given event is reported related to the size of the event: big events are likely to be known, small events are less likely to be known. These differences in the likelihood of observing information about an event can skew the available data and result in misleading interpretations about patterns of violence [5].

We use the word “bias” in the statistical sense, meaning a statistical difference between what is observed and what is “truth” or reality. “Bias” in this sense is not used to connote judgment. Rather the point is to focus attention on empirical, calculable differences between what is observed and what actually happened.

3. Case studies

We present here two examples from relatively well-documented conflicts. We present these examples specifically because some analysts have argued that information about conflict-related killings in Syria and

Iraq is complete, or at least sufficient for detailed statistical analysis. Given the data that are available about conflict violence, it might appear reasonable to assume that the data available in Syria or Iraq are telling a relatively complete and therefore unbiased picture of the violence that occurred (or is occurring) in those countries. However, we find that when examined closely, the multiple data sources that cover these conflicts tell conflicting narratives and are not appropriate for quantitative analyses aimed at answering questions about patterns of violence.

To be clear, this is not a criticism of any of these sources. In our experience over the past two decades working in more than thirty countries, we have found that different sources tend to have access to different aspects of each conflict, and consequently, each source offers unique and valuable insights. However, these different aspects, resulting in potential biases, must be accounted for if quantitative analyses are to build defensible answers about patterns of violence.

3.1. Syria

Many civilian groups are currently carrying out efforts to document and identify victims of violence in the midst of the ongoing conflict in Syria. In early 2012, the United Nations Office for the High Commissioner for Human Rights (UN-OHCHR) commissioned HRDAG to examine datasets from several of these groups. In three reports, Price et al. [6] provide in-depth descriptions of these sources, which are in essence lists of people who have been killed.

In this section, we focus our attention on four sources which cover the entire length of the ongoing conflict and which have continued to provide us with updated records of victims of killing. These sources are:

- Syrian Center for Statistics and Research [7] (CSR-SY)
- Syrian Network for Human Rights [8] (SNHR)
- Syria Shuhada Website [9] (SS)
- Violations Documentation Centre [10] (VDC).

For brevity, each list will be referred to by its acronym. Three of these sources (CSR-SY, SNHR, and VDC) rely primarily on trusted networks within Syria to provide and verify information about victims. The fourth source (SS) aggregates information from a variety of publicly available sources, including traditional and social media (see Price et al. for further details about these, and other, sources).

One way to investigate potential bias is to compare documentation patterns across sources. Figure 1 shows the number of victims documented by each of the four sources over time within the Syrian governorate of Tartus. The large peak visible in all four lines in May 2013 corresponds to an alleged massacre in Baniyas [11]. It appears that all four sources documented some portion of this event. Many victims were recorded in the alleged massacre, this event was very well-reported, and all four sources reflect this event in their lists. However, three out of the four sources document very little violence occurring before or after May 2013 in Tartus. The fourth source, VDC, shows the peak of violence in May as the culmination of a year of nearly consistent month-to-month increases in the number of reported killings.

As reported by the Oxford Research Group [12] and confirmed in personal communications with these documentation groups, it is difficult to access information in Tartus because this region is considered primarily loyal to the regime, and these documentation groups, despite their positioning as independent human rights groups, are frequently considered to side with the opposition. Consequently, the information available about this region varies dramatically, as indicated in Fig. 1. VDC clearly has more access than the other groups, but it is impossible to conclude solely from the observed data what portion of the violence in Tartus is covered by VDC.

Another way to investigate these documentation patterns is to consider the number of sources that document each victim. We use record linkage methods to identify multiple records that refer to the same victim, so we know that each source documents some victims included by one or more of the other sources in addition to some victims that are only documented by a single source. In other words, all of the sources are contributing substantial numbers of unique records of victims undocumented by the other sources [13].

This information is presented in Fig. 2, which shows the number of sources that document each victim (in contrast to the number of victims documented by each source, as shown in Fig. 1). Figure 2 shows the number of victims documented by all four sources (the darkest lower portion of each bar), three sources (the next lightest shade), two sources, and just one source (the lightest grey at the top of each bar). The information in Fig. 2 indicates that the different sources are not only documenting different *numbers* of victims, but in different months, they are documenting different mixes of unique and duplicated victims.

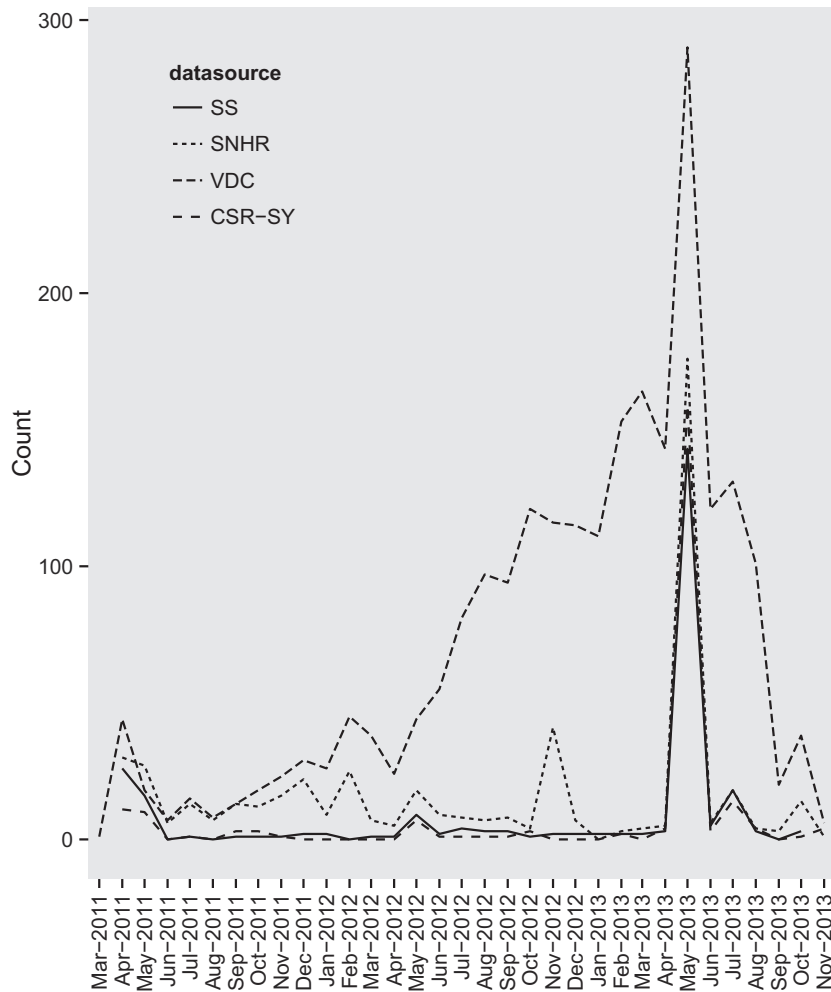


Fig. 1. Number of victims documented by four sources, over time, in Tartus.

This information could be the input to a statistical method called multiple systems estimation (MSE) (also often referred to as capture-recapture) which can be used to estimate the total number of victims, both those documented by one or more of these sources and those missing from current documentation efforts (i.e., undocumented victims) [14]. Presentation of MSE methods and results is beyond the scope of this paper, but the method motivates a consideration of what data may be missing from all of these sources.

The presence of event size bias is detectable in this particular example because all four of the sources obviously captured a similar event (or set of events) in May 2013. During the preceding months, one of those sources captured a very different subset of events. Were we to only have access to the three non-VDC sources, our conclusion about conflict violence in Tartus would incorrectly be that the alleged massacre in

May 2013 was an isolated event surrounded by relatively low levels of violence. Without a statistical model to estimate the unobserved killings, it is impossible to know which of these datasets tells the true story of the pattern of conflict killings in Tartus.

The conclusion from Figs 1 and 2 should not be that VDC is doing a “better” job of documenting victims in this region (however “better” might be measured). VDC is clearly capturing some events that are not captured by the other sources, but there is no way to tell how many events are not being captured by VDC. Furthermore, the other three sources are each capturing killings that VDC is not capturing.

To underline this crucial point: despite the availability of a large amount of data describing violence in Tartus, without using a model, there is no mathematically sound method to draw conclusions about the patterns of violence directly from the data (though it is possi-

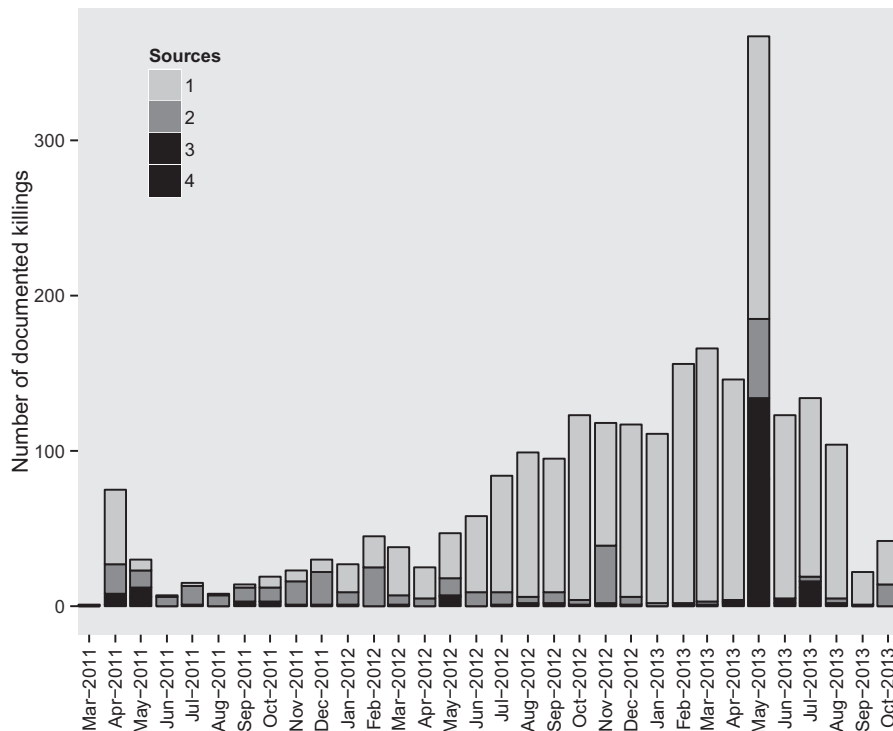


Fig. 2. Documented killings by month and number of sources per killing in Tartus.

ble to use the data and statistical models to estimate how many events are missing). The differences in the four sources available to us make it possible to detect the event size bias occurring in May 2013, but what other biases might also be present in this observed data and hidden from view? What new events might a fifth, sixth, or seventh source document? Are there enough undocumented events such that if they were included, our interpretation of the patterns would change? These are the crucial questions that must be examined when interpreting perceived patterns in observed data.

3.2. Iraq

We detect a subtler form of event size bias in data from the Iraq Body Count [15] (IBC), which indexes media and other sources that report on violent deaths in Iraq since the Allied invasion in March 2003. Our analysis is motivated by a recent study by Carpenter et al. [16], which found evidence of substantial event size bias. Their approach was to compare the US military's "significant acts" (SIGACTS) database to the IBC records. SIGACTS is based on daily "Significant Activity Reports" which include "...known attacks on Coalition forces, Iraqi Security Forces, the civilian population, and infrastructure. It does not include

criminal activity, nor does it include attacks initiated by Coalition or Iraqi Security Forces" [17].

Carpenter et al. report that their comparison of the two sources showed that "[e]vents that killed more people were far more likely to appear in both datasets, with 94.1% of events in which ≥ 20 people were killed being likely matches, as compared with 17.4% of singleton killings" [18]. We should note that IBC also conducted their own matching of events in their database with those in SIGACTS and found a substantially different match rate among smaller events. Specifically, IBC finds that 74% of the deaths in incidents with a single victim in SIGACTS's "civilian" category are found in other IBC records (while Carpenter et al., as reported above, find only 17% of single-victim events from SIGACTS in other IBC sources) [19].

Importantly, both studies find a much higher match rate for larger events than smaller events. IBC expected this to be true, prior to carrying out their own analysis: "We had good reasons to expect that there would be a strong relationship between sizes of events, measured by the number of deaths, and the rate at which these would match against the IBC data, with deaths in larger events matching more frequently than smaller ones" [20]. This implies that IBC, SIGACTS, or both

capture a higher fraction of large events than small events.

Motivated by this analysis, we considered other ways to examine IBC records for evidence of potential event size bias. Since IBC aggregates records from multiple sources, updated IBC data [21] already incorporates many records from SIGACTS. In contrast to the work of Carpenter et al. and IBC, both of whom treated IBC and SIGACTS as two separate data sources and conducted their own independent record linkage between the two sources, we examined only records in the IBC database, including those labeled as from SIGACTS.

It should be noted that we conducted this analysis on a subset of the data after filtering out very large events with more than 50 victims. We made this choice because, on inspection, many of the records with larger numbers of reported victims are data released by institutions (e.g., by morgues) or incidents aggregated over a period, rather than specific, individual events [22].

We began by identifying the top 100 data sources [23]; one or more of the top 100 sources cover 99.4% of the incidents in IBC. Given these sources, we counted the number of sources (up to 100) for each event. An earlier version of this analysis, published in *SAIS Review* [24], incorrectly assumed that all available sources for each event were included in the publicly available IBC data. The analyses presented here correctly include all available sources for a subset of data, as provided to us by IBC. See <https://hrdag.org/event-size-bias-iraq-body-count/> from November 2014 for a detailed discussion and correction of our previous assumption.

Event size was defined as the mean (rounded to the nearest integer) of the reported maximum and minimum event size values. Then the data were divided into four categories: events with one victim, events with 2–5 victims, events with 6–14 victims, and events with 15+ victims. The analysis was performed on these groups.

Figure 2 summarizes our findings. The shading of each bar in Fig. 2 indicates the proportion of events of that size reported by one, two, 3–4, 5–6, 7–10, 11–14, or 15 or more sources. The events with one victim have a mean of 2.8 sources per record, whereas events with 6–14 victims have 7.6 sources per record, and events with fifteen or more victims have an average of 12.5 sources per record. That is, the largest events have on average about 170%–350% more sources than the smallest events. Figure 2 shows, for example, that more than a quarter of events with only one victim

have only one source. By contrast, nearly half of the events with fifteen or more victims have fifteen or more sources. Clearly, larger events get more sources. This reinforces the finding by Carpenter et al. that larger events are more likely to be captured by both IBC and SIGACTS. We have generalized this finding to the top 100 sources; larger events are more likely to be captured by multiple sources.

The number of sources covering an event is an indicator of how ‘interesting’ an event is to a community of documentation groups – in this case, media organizations. The pattern shown in Fig. 2 implies that media sources are more interested in larger events than smaller events. Greater interest in the larger events implies that larger events are less likely to be ignored by every source, i.e., to be unobserved relative to smaller events. Since a larger proportion of small events are covered by only a single source, it is likely that more small events are missed, and therefore excluded from IBC [25].

As noted by Carpenter et al., “[t]he possibility that large events, or certain kinds of events (e.g., car bombs) are overrepresented might allow attribution that one side in a conflict was more recklessly killing civilians, when in fact, that is just an artifact of the data collection process” [26]. Put another way, the correlation between event attributes and the likely reporting of those events can result in highly misleading interpretation of apparent patterns in the data. As a relatively neutral example, analysts might erroneously conclude that most victims in Iraq were killed in large events, whereas this may be an artifact of the data collection. A potentially more damaging, incorrect conclusion might be reached if large events are centered in certain geographic regions or attributed to certain perpetrators; in these cases, reading the raw data directly would mistake the event size bias for a true pattern, and thereby mislead the analyst. Inappropriate interpretations could result in incorrect decisions regarding security measures, intervention strategies, and ultimately, accountability.

4. Discussion

Event size bias is one of many kinds of selection and reporting bias that are common to human rights data collection. It is important to recall that we refer here to biases in the statistical sense: a measurable difference between the observed sample and the underlying population of interest. As such, the biases that worry

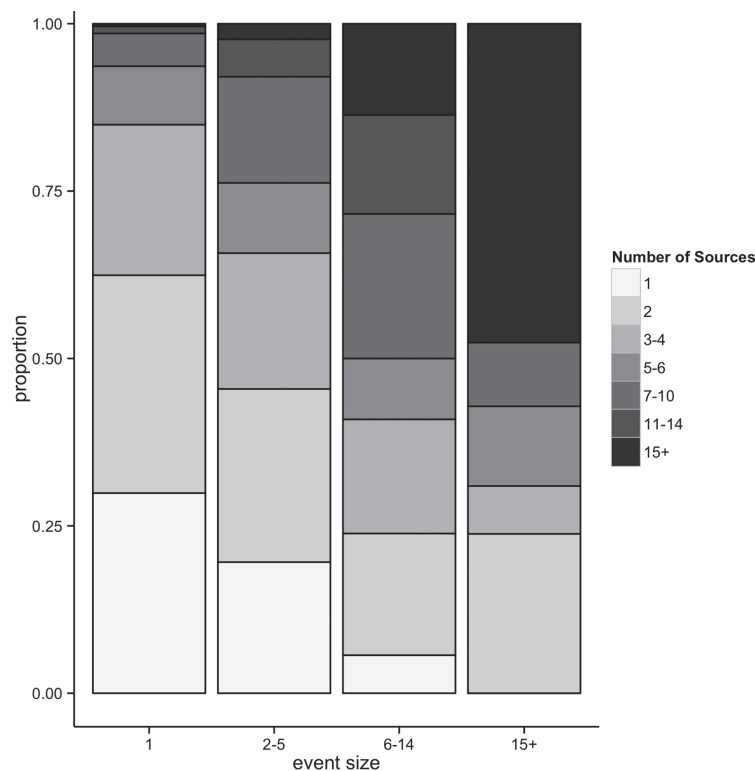


Fig. 3. Proportion of events covered by one to 15 or more sources.

us here affect statistics and quantitative analyses; we are not criticizing the content of the records which may provide valuable contextual details for qualitative analyses.

In the context of conflict violence, meaningful statistical analysis involves comparisons to answer questions such as: Did more violence occur this month or last month? Were there more victims of ethnicity A or B? Did the majority of the violence occur in the north or the south of the country? The concern about bias focuses on how the data collection process may more effectively document one month relative to another, creating the appearance of a difference between the months. Unfortunately, the apparent difference is the result of changes in the documentation process, not real changes in the patterns of violence.

To make sense of such comparisons, the observed data must in some way be adjusted to represent the true rates. There are a number of methods for making this adjustment if the observed data were collected at random. This is rarely true, and there are relatively few models that can correctly adjust data from convenience samples.

In order to compare nonrandom data across categories like months or regions, the analyst must assume

that the rate at which events from each category are observed is the same in each category (e.g., 60% of the true, total killings were collected in March, and 60% of the total killings were collected in April); this rate is called the coverage rate, and it is unknown unless somehow the true number of events were known or estimated. If the coverage rates for different categories differ, the observed data tell only the story of the documentation, they do not indicate an accurate pattern. For example, if victims of ethnicity A are killed in large-scale violent events with many witnesses, while victims of ethnicity B are killed in targeted, isolated violent events, we may receive more *reports* of victims of ethnicity A and erroneously conclude that the violence is targeted at ethnicity A. Until we adjust for the event size bias resulting in more reports of victims of ethnicity A, we cannot draw conclusions about the true relationship between the number of victims from ethnicity A versus B.

There are many other kinds of selection bias. As an example, when relying on media sources, journalists make decisions about what is considered newsworthy. Sometimes their decisions may create event size bias, as large events are frequently considered newsworthy. But the death of individual, prominent members

of a society are frequently also considered newsworthy. Conversely, media “fatigue” may result in underdocumentation later in a conflict, or when other newsworthy stories may limit the amount of time and space available to cover victims of a specific conflict [27]. Many other characteristics of both the documentation groups and the conflict can result in these kinds of biases: logistical or budgetary limitations, trust or affinity variations within the community, the security and stability of the situation on the ground, to name just a few [28]. As each of these factors changes, coverage rates are likely to change as well.

The fundamental reason why biases are so problematic for quantitative analyses is that bias often correlates with other dimensions that are interesting to analysts. As in the example of ethnicities A and B above, the event size bias is correlated with the kind of event. Failing to adjust for the reporting bias leads to the wrong conclusion. As another example, consider the Iraq case described above: if event size is correlated with the events’ perpetrators, then bias on event size means bias on perpetrator, and a naïve reading of the data could lead to security officials trying to solve the wrong security problems. Or in the Syria case: if decisions about resource allocation to Tartus were made in near-real time, they may have inaccurately concluded that violence documented in May 2013 represented an isolated event.

It is important to note that these challenges frequently lack a scientific solution [29]. We do not need to capture more data. Instead, what we need is to appropriately recognize and adjust for the biases present in the available data. Indeed, as indicated in the Iraq example, where multiple media sources appear to share similar biases, the addition of more data perpetuates and in some cases amplifies the event size bias.

Detection of and adjustment for bias requires statistical estimation – not just more data. A wide variety of statistical methods can be used to adjust for bias and estimate what is missing from observed data [30]. Each method has limitations and requires assumptions, which may or may not be reasonable. But formal statistical models provide a way to make those assumptions explicit, and in some cases, to test whether they are appropriate. Comparisons from raw data implicitly but necessarily assume that such snapshots are statistically representative. This assumption may sometimes be true, but only by coincidence.

5. Conclusions

Carpenter et al. warn that “press members and scientists alike should be cautious about assuming the completeness and representativeness of tallies for which no formal evaluation of sensitivity has been conducted. Citing partial tallies as if they were scientific samples confuses the public, and opens the press and scholars to being manipulated in the interests of warring parties.” In a back-of-the-envelope description elsewhere [31], we have shown that small variations in coverage rates can lead to an exactly wrong conclusion from raw data.

To re-emphasize: our findings here are not meant to criticize data producers. Iraq Body Count, the Syrian Center for Statistics and Research, the Syrian Network for Human Rights, the Syria Shuhada website, and the Violations Documentation Centre are collecting invaluable data, and they are doing so systematically, and with principled discipline. The data is extraordinarily valuable, and we urge these and other groups to continue to collate and share data as a fundamental record of the past. The data can also be used in qualitative research about specific cases, and in some circumstances, for use in statistical models that can adjust for biases.

Our goal in this paper is to warn against the naïve use of observed data to understand patterns of mass violence. It is tempting, particularly in politically and emotionally charged research such as studies of conflict-related violence, to search available data for answers. It is intuitive to create infographics, to draw maps, and to calculate statistics and draft graphs to look for patterns in the data. Unfortunately, all people – even statisticians – tend to draw conclusions even when we know that the data are inadequate to support comparisons. Weakly founded statistics tend to mislead the reader.

Statistics, graphs, and maps are seductive because they seem to promise a solid basis for conclusions. The current obsession with data, evidence-based policy, and similar ideas increase the pressure to use statistics, even as new doubts emerge about whether “big data” predictions about social conditions are accurate [32,33]. When the statistics have been calculated in a way that enables a mathematical foundation for statistical inference, statistics deliver on the promise of an objective measurement of a specific question. But analysis with inadequate data is very hard even for subject matter experts to interpret. In the worst case, it offers a falsely precise view, a view that may be completely wrong. In the best case, it invites speculation

about what's missing, what biases are uncontrolled, creating more questions than answers, and ultimately, a distraction. When policymakers turn to statistical analysis to address key questions, they must assure that the analysis gives the right answers.

6. About HRDAG

The Human Rights Data Analysis Group is a non-profit, non-partisan organization [34] 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 and Facebook page.

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

- [1] One extreme example includes Target successfully predicting a customer's pregnancy, as reported in the New York Times and Forbes. In particular, Target noticed that pregnant women buy specific kinds of products at regular points in their pregnancy, and the company used this information to build marketing campaigns.
- [2] C. Davenport and P. Ball, Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977–1996, *Journal of Conflict Resolution* **46**(3) (2002), 427–450.
- [3] P. Ball, J. Asher, D. Sulmont and D. Manrique-Vallier, How Many Peruvians Have Died? American Association for the Advancement of Science, 2003.
- [4] Another common kind of bias that affects human rights data is *reporting* bias. Whereas selection bias focuses on how the data collection process identifies events to sample, reporting bias describes how some points become hidden, while others become visible, as a result of the actions and decisions of the witnesses and interviewees. For an overview of the impact of selection bias on human rights data collection, see Jule Krüger, Patrick Ball, Megan Price, and Amelia Hoover Green. "It Doesn't Add Up: Methodological and Policy Implications of Conflicting Casualty Data." In *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict*, T.B. Seybolt, J.D. Aronson and B. Fischhoff, ed., Oxford UP, 2013.
- [5] *Op.cit* Davenport and Ball 2002.
- [6] M. Price, A. Gohdes and P. Ball, Updated Statistical Analysis of Documentation of Killings in the Syrian Arab Republic. Human Rights Data Analysis Group, commissioned by the United Nations Office of the High Commissioner for Human Rights (OHCHR), 2014.
M. Price, J. Klingner, A. Qtiesh and P. Ball. Full Updated Statistical Analysis of Documentation of Killings in the Syrian Arab Republic. Human Rights Data Analysis Group, commissioned by the United Nations Office of the High Commissioner for Human Rights (OHCHR), 2013.
M. Price, J. Klingner and P. Ball, Preliminary Statistical Analysis of Documentation of Killings in the Syrian Arab Republic. The Benetech Human Rights Program, commissioned by the United Nations Office of the High Commissioner for Human Rights (OHCHR), 2013.
- [7] <http://www.csr-sy.com>.
- [8] <http://www.syrianhr.org>.
- [9] <http://syriansshuhada.com>.
- [10] <http://www.vdc-sy.info>.
- [11] See report in the LA Times, BBC, and The Independent, among others.
- [12] Hana Salama and Hamit Dardagan. Stolen Futures: The Hidden Toll of Child Casualties in Syria. Oxford Research Group: <http://goo.gl/Hlv14L>. 2014.
- [13] *Op.cit* Price. et al. 2014.
- [14] See <https://hrdag.org/mse-the-basics/> for the first in a series of blog posts describing Multiple Systems Estimation (MSE) or Kristian Lum, Megan Emily Price and David Banks. Applications of Multiple Systems Estimation in Human Rights Research, *The American Statistician* **67**(4) (2013), 191–200.
- [15] <http://www.iraqbodycount.org>.
- [16] D. Carpenter, T. Fuller and L. Roberts, WikiLeaks and Iraq Body Count: the Sum of Parts May Not Add Up to the Whole – A Comparison of Two Tallies of Iraqi Civilian Deaths, *Prehosp Disaster Med* **28**(3) (2013), 1–7.
- [17] http://www.globalsecurity.org/military/ops/iraq_sigacts.htm.
- [18] *Op.cit* Carpenter et al. 2013.
- [19] See this blogpost <https://hrdag.org/event-size-bias-iraq-body-count/> from November 2014 for a more detailed discussion of the two different matching approaches.
- [20] <https://www.iraqbodycount.org/analysis/numbers/warlogs-appendix/>.
- [21] We downloaded the ibc-incidents file on 14 Feb 2014, and processed it using the pandas package in python.
- [22] See this blogpost <https://hrdag.org/event-size-bias-iraq-body-count/> from November 2014 for a more detailed discussion of the challenges of aggregated (or composite) events.
- [23] The top 100 sources include, for example, AFP, AL-SHAR, AP, CNN, DPA, KUNA, LAT, MCCLA, NINA, NYT, REU, VOI, WP, XIN, and US DOD VIA WIKILEAKS.
- [24] M. Price and P. Ball, Big Data, Selection Bias, and the Statistical Patterns of Mortality in Conflict, *SAIS Review of International Affairs*, Volume 34, Issue 1, Winter-Spring 2014, 2014, pp. 9–20.
- [25] These assumptions can be formalized and tested within the framework of 'species richness,' which is a branch of ecology that estimates the number of different types of species within a geographic area and/or time period of interest using models for data organized in a very similar way to the IBC's event records. See Ji-Ping Wang, Estimating species richness by a Poisson-compound gamma model, *Biometrika* **97**(3) (2010), 727–740.
- [26] *Op.cit* Carpenter et al. 2013.
- [27] A research question to address this might be: Do media-

- reported killings in a globally-interesting conflict like Iraq or Syria decline during periods when other stories attract interest? Do reported killings decline during the Olympics?
- [28] *Op.cit* Krüger et al. 2013.
- [29] Bias issues can sometimes be resolved with appropriate statistical models, that is, with better scientific reasoning about the specific kind of data involved. However, we underline that bias is not solvable with better technology. Indeed, some of the most severely biased datasets we have studied are those collected by semi- or fully-automated, highly technological methods. Technology tends to increase analytic confusion because it tends to amplify selection bias.
- [30] For a description of multiple systems estimation, see *op.cit* Lum et al. 2013. For methods on missing data in survey research which might be applicable to the adjustment of raw, non-random data if population-level information is available, see J. Michael Brick and Graham Kalton. Handling Missing Data in Survey Research, *Statistical Methods in Medical Research* **5**(3) (1996), 215–238. For an overview of species richness models which might be used to estimate total populations from data organized like the IBC, see *op. cit* Wang. For an analysis of sampling issues in “elusive” populations, see Lisa G. Johnston and Keith Sabin. Sampling Hard-to-Reach Populations with Respondent Driven Sampling, *Methodological Innovations Online* **5**(2) (2010), 38–48.
- [31] <https://hrdag.org/why-raw-data-doesnt-support-analysis-of-violence/>.
- [32] D. Lazer, R. Kennedy, G. King and A. Vespignani, Google Flu Trends Still Appears Sick: An Evaluation of the 2013–2014 Flu Season (March 13, 2014). Available at SSRN: <http://ssrn.com/abstract=2408560>.
- [33] T. Harford, Big Data: Are We Making a Big Mistake? *Financial Times Magazine*, 2014. <http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html>.
- [34] Formally, HRDAG is a fiscally sponsored project of Community Partners.