

RESPONSIBLE DATA REFLECTION STORIES 7

A collection of real-life examples of the risks that are faced when using data in advocacy work, along with mitigation strategies to overcome these challenges.

Recognising uncertainty in statistics

Quantitative data is the result of numerous subjective human decisions.

CONTEXT

*More often than not, it is not the writer that is twisting the numbers but the numbers themselves twisting up the writer; manipulation of the facts, or of the reader, is usually not intentional. The exploration of the use and misuse of numbers is at the base of a large, and growing, body of academic and popular work on quantitative literacy.—p3, *Numbers are Only Human*, Brian Root*

Understanding and using statistics responsibly in human rights advocacy can be incredibly difficult. As Brian Root excellently outlines in his article, “*Numbers are Only Human*”¹, understanding that quantitative data is the result of numerous subjective human decisions, can make a big difference to how an organisation chooses to use certain statistics to support their work.

One concrete example of how difficult these decisions can be, though, can be seen if we look at data on killings due to the ongoing Syrian conflict.

1

<https://global.oup.com/academic/product/the-transformation-of-human-rights-fact-finding-9780190239497?cc=us&lang=en&>



BACKGROUND

There are several organisations who are monitoring casualties in Syria—including, but not limited to²:

- › March 15 Group
- › Syrian government
- › **Syrian Center for Statistics and Research**³
- › **Syrian Network for Human Rights**⁴
- › **Syrian Observatory for Human Rights**⁵
- › Syrian Revolution General Council
- › Syria Shuhada Website
- › Violations Documentation Centre, the documentation arm of the Local Coordination Committees
- › Damascus Center for Human Rights Studies

Figures released by all of the above parties (some of whom are no longer collecting up to date data) differ greatly. As an example, in 2013 the Syrian Observatory for Human Rights (SOHR) enumerated estimated that the **TOTAL NUMBER OF PEOPLE KILLED WAS 110,371 PEOPLE**.

² Selection taken from HRDAG report, <https://hrdag.org/wp-content/uploads/2013/06/HRDAG-Updated-SY-report.pdf> with additions

³ <http://csr-sy.org/>

⁴ <http://sn4hr.org/blog/category/victims/death-toll/>

⁵ <http://www.syria-hr.com/en/>

ESTIMATE VS ENUMERATE

Using a convention from the Human Rights Data Analysis Group (HRDAG), the word “estimate” is used to describe an estimation of a figure that includes a measure of uncertainty. Other non-statistical ‘guesses’ are referred to throughout this piece as “enumerations”.

They even released more disaggregated figures along with this estimation:

CIVILIANS KILLED	40146 OF WHICH:
WOMEN	NEARLY 4000
CHILDREN	MORE THAN 5800
REBEL FIGHTERS	21,850
REGIME ARMY SOLDIERS	27654
PRO-REGIME MILITIA	17824
HEZBOLLAH	171
UNIDENTIFIED	2726

THE PROBLEM OF SUBJECTIVITY

As Root identifies, the disaggregated categories within the data are also subject to a lot of human decision making:

Imagine the decision that might have to be made to categorize a typical citizen with no military training, who has picked up a gun shortly before his death. Perhaps the coder might have a bias to continue calling this person a civilian. But this person took up arms against the government, did they not? How would you code a Syrian army defector now fighting with an opposition group? ... Without some sort of standard protocol, rigorously followed, the coding of affiliation allows for a degree of subjectivity...-p6

As he rightly identifies—there are lots of human decisions that go into creating these statistics, and without knowing how these deaths have been coded, it’s difficult to trust in the figures. But this nuance can be difficult to convey without using long-winded explanations, and sadly, soundbites of short, snappy figures, often get much more traction in public debate (see **REFLECTION STORY 6** for more on this).

The messiness of the Syrian conflict adds to this confusion in coding; it can easily be unclear who is responsible for a certain attack, especially those which nobody wants to take responsibility for, such as chemical weapons attacks.

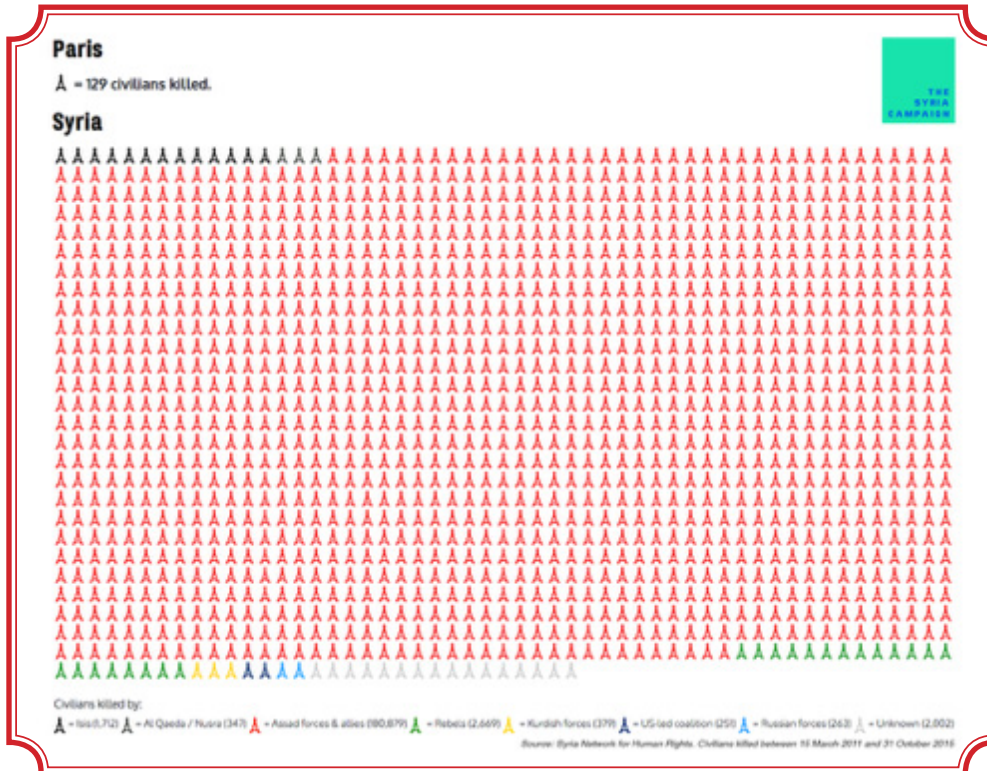


Image from the Syria Campaign: <https://thesyriacampaign.org/>

As Anita Gohdes, researcher at the Human Rights Data Analysis Group, highlights: in addition to the problem of subjectivity and human judgement, this level of disaggregation can convey a false sense of accuracy. Breaking down figures to such precise levels—“21,850 rebel fighters” rather than “Approximately 22,000 rebel fighters” removes the sense of uncertainty that is undoubtedly there.

A FALSE SENSE OF ACCURACY

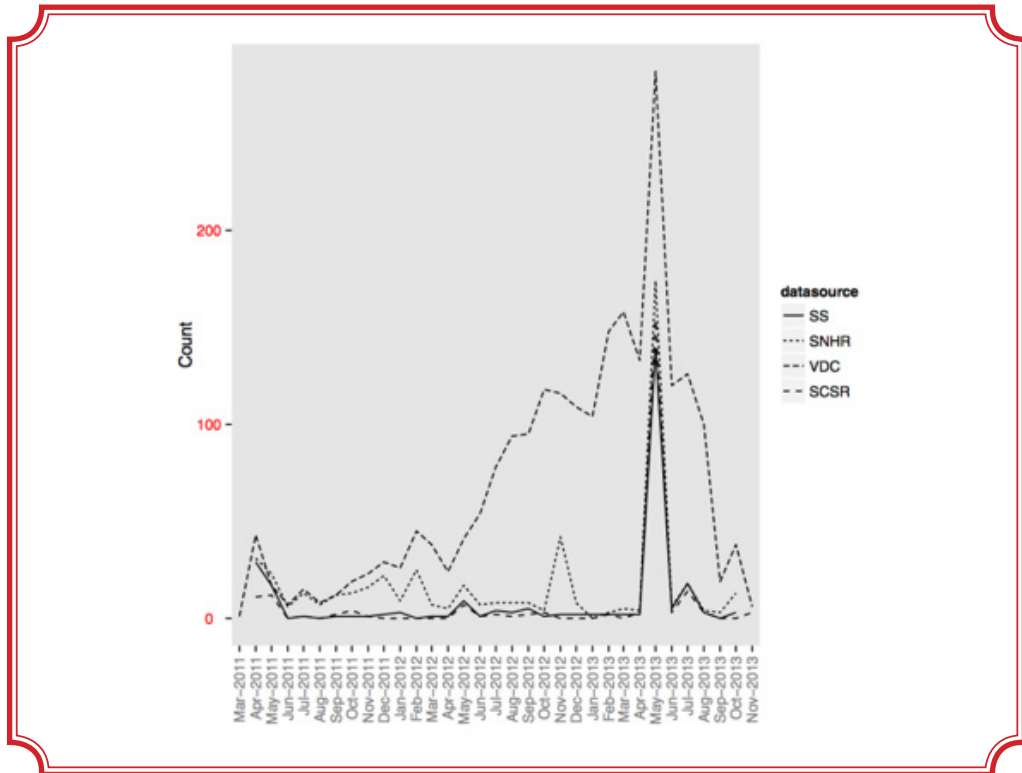
It’s true though, that images such as the visualisation above draw attention to some important issues. Though they state their data source (the Syria Network for Human Rights) what we’ve explored here so far makes it clear that **THIS DATA HAS FLAWS**. We can’t know for sure the extent of those flaws, though, and some might argue that as long as the main message is transmitted, the details don’t matter so much.

A data-focused group, the Human Rights Data Analysis Group, (HRDAG), published in mid 2013 a statistical analysis of documentation of killings in Syria⁶, commissioned by the Office of the UN High Commissioner for Human Rights. In it, they integrate findings from eight different databases; seven built by Syrian human rights monitors, and one from the Syrian government.

==

6 <https://hrdag.org/wp-content/uploads/2013/06/HRDAG-Updated-SY-report.pdf>

Image taken from
 'Searching for Trends:
 Analyzing Patterns in
 Conflict Violence Data'
 post by Megan Price and
 Anita Gohdes, copied
 here with permission



In it, they qualify carefully their findings; that the final figure they came up with, 92,901 unique killings of both combatants and non-combatants, is an enumeration and not the complete number of conflict-related killings. They identify potential problems like undetected duplicate deaths among the different databases; inaccurate records within any of those databases; victims presumed dead who may have later been found alive; or, undocumented killings, those that don't appear in any of the the eight databases for any number of reasons, to name just a few.

To highlight the problems that they discuss, take a look at the graph above. It shows **THE DAILY COUNT OF UNIQUELY REPORTED KILLINGS** in the area of Tartus, as collected by four well-known data sources.

As you can see, there are some big discrepancies in the data, notably leading up to May 2013.

All four sources depict a marked increase in violence in May 2013⁷, which corresponds to an alleged massacre in that governorate. Three of the remaining sources observed relatively few victims outside this single spike in violence. The fourth source, VDC, describes the observed peak in May 2013 as the culmination of steadily increasing reports of violence throughout the preceding year. If we did not have access to the VDC data, we would erroneously conclude that there is consensus among data sources that relatively little violence is occurring in Tartus, and that May 2013 was a relatively isolated event. –‘Searching for Trends: Analyzing Patterns in Conflict Violence Data’⁸ post by Megan Price and Anita Gohdes, April 2, 2014.

The challenge: admitting weaknesses in the data, while pushing a strong message

As HRDAG and Brian Root at Human Rights Watch have identified, getting “accurate” data on killings in the Syrian conflict is incredibly challenging. Any and all of the major data sources face uncertainties in getting the data, in coding it accurately according to type of death or killing, and in comprehensive data coverage of hard to reach areas. In short; none of the data is certain accurate enough to be cited as “truth” or fact.

It’s true, though, that data on the topic is very much needed for a number of reasons; in order to get an idea of the scale of the conflict, to understand what humanitarian needs there are; and on the advocacy side, to get people’s attention to a tragic situation, and garner public support.

There are many ways of counting deaths in conflicts that have already ended, as outlined [in this Guardian article](#)⁹—and even when they have ended, there is still a great deal of uncertainty around the accuracy and subjectivity of that data. This challenge is even greater in the case of Syria, where the conflict is ongoing and increasingly messy between different stakeholders.

The main challenge in this case is knowing where to draw the line between using quantitative data to strengthen advocacy, or where to admit uncertainty and potentially weaken the key advocacy message.



7 www.bbc.com/news/world-middle-east-22410392
 8 politicalviolenceatagance.org/2014/04/02/searching-for-trends-analyzing-patterns-in-conflict-violence-data/
 9 www.theguardian.com/global-development-professionals-network/2015/sep/08/from-syria-to-sudan-how-do-you-count-the-dead

Ways of dealing with uncertainty

The Human Rights Data Analysis Group are pioneering the way in collecting and analysing figures of killings in conflict in a responsible way, using what's called '**multiple systems estimation**'.¹⁰ But it's true that the statistical skills required to use these kinds of techniques may well lie far beyond the reach of many advocacy groups, and in this case, there are a few other considerations and techniques that can be employed:

- › questioning methodologies of how the data has been gathered and analysed— for example, who decides what code is given to a data point? What levels of verification do they have in place?
- › making proactive decisions around whether to include or focus upon a certain statistic as part of the advocacy, or to include qualifiers about how reliable it may or not be (eg. adding 'estimated as of ____')
- › linking to other sources of deeper enquiry of the reliability or consistency of the data (eg. in this case, the HRDAG study linked above)

Clearly, with an issue as complex yet important is this—and, as with many of the reflection stories— there is no **RIGHT** answer. At the heart of mitigating against misunderstandings of data is increasing the level of data literacy of those who will be looking at the visualisations, but this is easier said than done.



Licensed under Creative Commons Attribution-ShareAlike 4.0 International License. (CC-BY-SA 4.0)



the engine room

This publication is part series found at <https://responsibledata.io>, produced by the engine room's Responsible Data Program, 2016.