

Pride and prejudice

Addressing bias in the
analysis of political violence

Kars de Bruijne
Erwin van Veen

Clingendael Report



Clingendael

Netherlands Institute of International Relations



Clingendael

Netherlands Institute of International Relations

Pride and prejudice

Addressing bias in the analysis of political violence

Kars de Bruijne
Erwin van Veen

Clingendael Report
December 2017

Disclaimer: This report was commissioned by the Netherlands' ministries of Foreign Affairs and Defence within the PROGRESS framework agreement, lot 5, 2017. Responsibility for the contents and for the opinions expressed rests solely with the authors. Publication does not constitute an endorsement by the Netherlands' ministries of Foreign Affairs and Defence.

This report is part of Clingendael's political violence program on information gathering, analysis and recognition of (data) bias by policy-makers.

December 2017

© Netherlands Institute of International Relations 'Clingendael'.

Cover photo: Journalist documenting events at the Independence square. Clashes in Ukraine, Kyiv. Events of February 18, 2014. © Wikimedia Commons / Mstyslav Chernov

Unauthorized use of any materials violates copyright, trademark and / or other laws. Should a user download material from the website or any other source related to the Netherlands Institute of International Relations 'Clingendael', or the Clingendael Institute, for personal or non-commercial use, the user must retain all copyright, trademark or other similar notices contained in the original material or on any copies of this material.

Material on the website of the Clingendael Institute may be reproduced or publicly displayed, distributed or used for any public and non-commercial purposes, but only by mentioning the Clingendael Institute as its source. Permission is required to use the logo of the Clingendael Institute. This can be obtained by contacting the Communication desk of the Clingendael Institute (press@clingendael.org).

The following web link activities are prohibited by the Clingendael Institute and may present trademark and copyright infringement issues: links that involve unauthorized use of our logo, framing, inline links, or metatags, as well as hyperlinks or a form of link disguising the URL.

About the authors

Kars de Bruijne is a Research Fellow at Clingendael and a Senior Researcher in the Armed Conflict Location Event Dataset Project (ACLED). His research focuses on intrastate conflict, fragile states, trend analyses and expert forecasting.

Erwin van Veen is a Senior Research Fellow with Clingendael's Conflict Research Unit. A political scientist by training, Erwin applies this lens to relations between political order, security and justice in conflict-prone environments. Extensive travel in the Middle East also generated Erwin's lasting interest in the region's conflicts.

The Clingendael Institute
P.O. Box 93080
2509 AB The Hague
The Netherlands

Follow us on social media

 [@clingendaelorg](https://twitter.com/clingendaelorg)
 The Clingendael Institute
 The Clingendael Institute

Email: info@clingendael.org
Website: www.clingendael.org

Contents

1	Introduction	1
2	Mind the gap: bias at the data-gathering stage	3
	Bias #1: Unknown amounts of missing data	3
	Bias #2: Short field presences result in limited data-gathering	5
	Bias #3: The consequences of missing data: witness, reporting and description practices	6
	Bias #4: Data-gathering is limited to urban elites	8
3	Mind the frame: bias at the data-analysis stage	10
	Bias #5: Conflict framing	10
	Bias #6: Insufficient reflection on description biases of new data sources	12
	Bias #7: Single case studies lead to overconfident generalisations	14
	Bias #8: Description problems with local conflict databases	14
4	What can policy-makers and researchers do to mitigate bias?	16

1 Introduction

High-quality conflict analysis is an essential ingredient for sound conflict prevention, mitigation and intervention strategies. Unfortunately, conflict settings are not conducive to producing such an analysis. Conflict areas can be dangerous to visit, information flows tend to be controlled, or at least coloured while both conflict description and framing are subject to distortion.

This report offers a discussion of common bias problems that occur when analysing political violence in intrastate conflict settings. These issues matter both for researchers who actually conduct such analysis, as well as for policy-makers that base their thinking and initiatives on such an analysis (at least in part) and sometimes engage in data collection themselves.¹ We define bias as the difference between actual reality and reality as reported or perceived.² ‘Actual reality’ represents an individual or collective act of political violence that has occurred. Subsequently, these acts are ‘constructed’, for example in how they are reported, analysed or framed by perpetrators, researchers, journalists or policy-makers.³ Bias, therefore, occurs when there is a difference between ‘actual acts of political violence’ and ‘constructed acts of political violence’.

Examples of bias in an analysis of political violence are inevitable and occur regularly. For example, there may be an absence of reporting on significant events that have occurred or reporting on events that have not actually happened. Or the nature of incidents become distorted, e.g. a personal feud is framed as a political act. Or actions are incorrectly attributed, e.g. a rebel group is thought to be behind an attack while it is in fact a rogue commander. Or an inappropriate framework is used to describe violence, e.g. a local conflict is regarded as manifestation of a regional conflict. We do not claim that the difference between ‘actual reality’ and ‘constructed reality’ can be overcome. But we do claim that bias can be reduced by recognising and mitigating the ways in which we describe and frame political violence.

-
- 1 We note that researchers can be based within or outside of governmental bodies and intragovernmental organisations. I.e. this term refers to academic researchers, think-tank organisations, analysts within the intelligence community and strategic analysis units in ministries of foreign affairs, to name but a few.
 - 2 The brief is part of a broader research package conducted by Clingendael and partners under the Progress tender of the Dutch Ministries of Foreign Affairs and Defence. The authors would like to thank Joost van Elk (Dutch Foreign Ministry) for his support with this brief in particular. They are also grateful to Ursual Daxecker (Associate Professor, University of Amsterdam) and Fransje Molenaar (Clingendael) for a helpful peer review and Lauriane Héau for research assistance. The brief is based on the personal experience of the authors, a literature review and a half-day workshop involving Dutch government departments.
 - 3 Throughout the brief, bias is used in a statistical sense and does not denote any judgement.

We define political violence as the actual use of force by an organised armed actor for the purpose of a political goal (e.g. such goals include gaining public authority/resources and reconfiguring public authority structures).⁴ Such violence can be carried out against governmental institutions, communities, groups or individuals. Political violence is a common indicator of conflict.

For policy-makers and researchers alike, the issue of bias is important in this context especially as conflict parties are likely to intentionally 'distort' reality to their advantage. In other words, they will seek to use conflict reporting to shape, warp, diminish or augment reality to serve their purposes.⁵ A number of parallel realities will exist that are supported by the same set of 'facts'. The task for parties which are external to the conflict is to produce the fullest and most trenchant analysis possible of the drivers, state and consequences of political violence. This means, among other things, that the different components of each narrative will need to be scrutinised and that a search for evidence will need to take place (e.g. views of minority groups or marginalised sectors of society). With the growing amount of data available, as well as purposeful covert and overt attempts to influence public opinion, the problem of bias is an increasing cause for concern.

Section 2 of this paper discusses common biases that can occur during the data-gathering stage. Section 3 does the same for the data-analysis stage. Section 4 offers a number of mitigation strategies.

4 See for definitions on political violence Nicholas Sambanis, "What Is Civil War?: Conceptual and Empirical Complexities of an Operational Definition," *Journal of Conflict Resolution* 48, no. 6 (December 1, 2004): 814–58, <https://doi.org/10.1177/0022002704269355>; Timothy Besley and Torsten Persson, "The Logic of Political Violence," *The Quarterly Journal of Economics* 126, no. 3 (August 1, 2011): 1411–45, <https://doi.org/10.1093/qje/qjr025> on violence as a vehicle for acquiring rights, recognition and more representative / equitable power structures. Kaldor, M., *New and old wars: Organized violence in a global era*, 3rd edition, Stanford: SUP, 2012 and Tilly, C., *The politics of collective violence*, Cambridge: CUP, 2003. We exclude interstate violence, isolated lone-wolf acts and organised crime, but include terrorist acts carried out by an organised armed actor (such as the Islamic State).

5 A subtle example of successful conflict framing is how Israel has persuaded the BBC to take into account its view on Israeli settlements in Palestine. Most BBC reporting will include a sentence like this: 'These [i.e. the settlements] are considered illegal under international law, although Israel disputes this.' See for example: <http://www.bbc.com/news/world-middle-east-42246564> (accessed 18 December 2017). The subliminal effect of this sentence is that the reader is left with the feeling that these settlements are, perhaps, not so illegal after all. Although this is not actually the case if one takes into account United Nations Security Council Resolution nos 242 (22 November 1967), 338 (22 October 1973) and 2334 (23 December 2016); as well as: International Court of Justice, *Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory*, The Hague: ICJ, Summary 2004/2, 9 July 2004.

2 Mind the gap: bias at the data-gathering stage

Today's political violence data landscape is characterised by a great many sources. In part this is due to the data information revolution, and in part to greater analysis of civil war after the Cold War (note that the incidence of civil war has not increased). Unfortunately, the availability of more sources has not necessarily resulted in better analysis. Bias continues to exist both in situations where there is insufficient information – e.g. when parts of the population or transnational actors are hard to access – and in situations where there is an abundance of evidence – e.g. due to data absorption or framing issues. While bias in low-data availability situations is likely to have a more distortive effect and less likely to be challenged than is the case in high-data availability situations, four common biases can interfere in either situation with the data-gathering process. These are briefly discussed below and are drawn from existing qualitative and quantitative research.

Bias #1: Unknown amounts of missing data

Raw conflict data is produced by conflict parties and non-conflict parties. Such raw data is subsequently often recorded in some form or another before it is accessible for research and policy purposes (e.g. military records, news sources, archives, testimony). The salient observation here is that every recording is based on data selection choices. Such choices are informed by criteria like data collection safety, area accessibility, political significance and organisational priorities. Each of these choices can result in limitations and omissions in initial recordings (intentional or unintentional). In other words, every article, report and database provides a snapshot based on a 'convenience sample'.

While these limitations are well known, the consequences are not sufficiently appreciated. For many recordings, it is not known how much 'actual reality' is contained therein. For example, reports from renowned journals, military records and think-tanks may be based on nearly all of the reality or only a fraction thereof. The issue is illustrated in **Figure 1**. Its two circles represent 'the real number of events' and the 'number of events as recorded'. If the circles overlap we know that the data is complete. If they do not overlap the bias is large. Hence, it is not known – neither by policy-makers nor researchers – how much we in fact do not know.

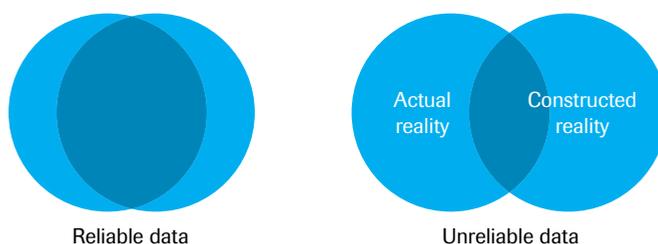


Figure 1 We don't know how much 'reality' our information gathering efforts capture

For policy-makers and researchers the problem of missing data is important since it may lead to incorrect conclusions about trends. For example, a recent analysis of the conflict in Syria compared four different organisations monitoring civilian killings in Tartus (a Syrian province) from 2011 to 2013.⁶ As **Figure 2** shows, each organisation reported a different trend in the number of civilians being killed. Three organisations reported a sudden peak while a fourth organisation – the Violence Documentation Centre – showed a slow build-up in civilian killings. If the data had been used for early warning of civilian killings, we would have seen – and hopefully acted – only on data from the Violence Documentation Centre. Yet if we had relied on data from either one of the three other organisations we would not have seen a warning in advance.

There is ample research showing that incomplete data is a common phenomenon. For example, recent research on Afghanistan compared a database that contained almost all events (it was a leaked US military file) with a public conflict database that is often used by policy-makers.⁷ The comparison highlighted that the public conflict database covered only a third of all events. Another example from Sierra Leone showed a similar picture (only 20 to 40% of conflict incidents was covered in public conflict databases).⁸

6 Megan Price and Anita Gohdes, "Searching for Trends: Analyzing Patterns in Conflict Violence Data," *Political Violence at a Glance* (blog), April 2, 2014, <http://politicalviolenceataglance.org/2014/04/02/searching-for-trends-analyzing-patterns-in-conflict-violence-data/>.

7 SIGACTS, a database on 'significant activities' in Afghanistan and Iraq collected by the United States military and Coalition partners to monitor the conflict. Data of this nature is usually not public and is very likely to be almost complete. Nils B. Weidmann, "A Closer Look at Reporting Bias in Conflict Event Data," *American Journal of Political Science* 60, no. 1 (January 1, 2016): 206–18, <https://doi.org/10.1111/ajps.12196>.

8 Kars de Bruijne, "Introducing the Sierra Leone Local – Location Event Dataset (SLL-LED)", *Armed Conflict Location Event Dataset* (2014), <http://www.acleddata.com/local-data/>.

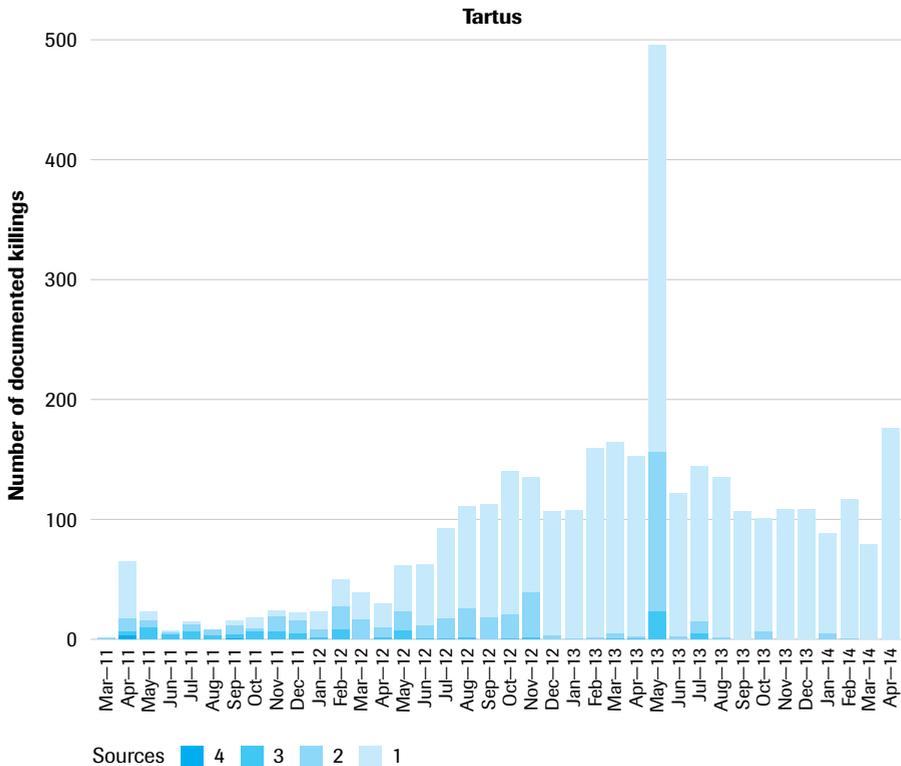


Figure 2 Intensity of civilians targeted in Tarsus as reported by four organisations (2011-2013)

Bias #2: Short field presences result in limited data-gathering

Data is often generated in contexts that are unsafe and, in part for this reason, over short periods. As a consequence, only limited data can be gathered. This feature reinforces other biases, such as a focus on urban elites and easy reconfirmation of existing conflict frames (both are discussed below in greater detail). A personal anecdote may serve here. When one of the authors was in Baghdad researching relations between key Shi’a political parties, he interviewed representatives of Al-Sadr’s parliamentary bloc (a Shi’a party). Yet he did not manage to obtain access into Sadr city that is the social base of this party because of a mix of safety concerns, a lack of connections/access and a short stay. Considering this, how well could the positions of this party be understood? One possible problem is that researcher heuristics and biases are insufficiently challenged as a researcher is not exposed to the unusual or to the owners of radically different frames than the mainstream (see Section 3 for a more extensive discussion). Such challenges become more evident – and more resolvable –

when one benefits from longer or deeper exposure to the history, legacies and intricacies of a particular conflict.⁹

Bias #3: The consequences of missing data: witness, reporting and description practices

An act of political violence ends up as conflict data when three criteria are satisfied. First, the act needs to be experienced and reported by a victim or bystander that remains alive. Second, the observer then needs to reach an outlet (media, human rights organisations, intelligence unit, document). Finally, the outlet will have to make the act public. Each stage generates a number of biases that have an impact on whether and how the act is reported. We present the three most important biases that are known to exist in most event-databases:

- *Conflict events with many casualties are reported more often than conflict events with no casualties.* A consistent research finding is that acts with many casualties are overrepresented. For example, in 2009 research on Iraq compared US military records with the Iraq Body Count project. They found that events where more than 20 people had been killed were covered by both databases (in 94% of cases) while events with only one casualty had little overlap (17%).¹⁰ The reason why conflict events with higher casualties are more often reported, is that they have more witnesses, that testimonies reach outlets more easily and the outlets have incentives to report on high-impact events. One problem here is that trends concerning civilian killings and the culpability of conflict parties therein are hard to establish. Group A could, for instance, engage in a few high-impact events while group B is involved in smaller incidents with fewer casualties each but a larger total.

9 For an extensive discussion of the matter: Kahneman, D., P. Slovic and A. Tversky, *Judgment under uncertainty: Heuristics and biases*, Cambridge: CUP, 2008.

10 Dustin Carpenter, Tova Fuller, and Les 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," *Prehospital and Disaster Medicine* 28, no. 03 (June 2013): 223–29, <https://doi.org/10.1017/S1049023X13000113>. Another example on Afghanistan found that military records from WikiLeaks and public databases overlapped when the number of casualties was higher: Nils B. Weidmann, "The Higher the Better? The Limits of Analytical Resolution in Conflict Event Datasets," *Cooperation and Conflict* 48, no. 4 (2013): 567–576; Weidmann, "A Closer Look at Reporting Bias in Conflict Event Data"; Megan Price and Patrick Ball, "Selection Bias and the Statistical Patterns of Mortality in Conflict," *Statistical Journal of the IAOS* 31, no. 2 (January 1, 2015): 263–72, <https://doi.org/10.3233/sj-150899>.

- *Events in rural areas with fewer people at a greater distance from the capital are underreported.* Decades of research show that violence is more often reported from densely-populated places, in urban areas and close to the capital. This has been shown in research on racial conflict,¹¹ demonstrations in Switzerland,¹² violence in Afghanistan¹³ and state violence in Guatemala.¹⁴ This means that conflict data is better suited to trends in conventional war than guerilla conflicts as the latter tend to take place in rural areas.¹⁵ Also, conflict data may reinforce urban biases and middle-class concerns because conflict incidents in wealthier, urban settings will be reported on more often.¹⁶
- *Some sources are biased towards smaller incidents with fewer casualties.* It is known that local press,¹⁷ social media and phone coverage,¹⁸ blogs,¹⁹ newswires like

-
- 11 M. Herbert Danzger, "Validating Conflict Data," *American Sociological Review* 40, no. 5 (1975): 570–84, <https://doi.org/10.2307/2094196>; Jennifer Earl et al., "The Use of Newspaper Data in the Study of Collective Action," *Annual Review of Sociology* 30 (2004): 65–80, <https://doi.org/10.1146/annurev.soc.30.012703.110603>; 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): 427–450; Robert Justin Goldstein, "The Limitations of Using Quantitative Data in Studying Human Rights Abuses," *Human Rights Quarterly* 8, no. 4 (1986): 607–27, <https://doi.org/10.2307/762195>.
 - 12 José Barranco and Dominique Wisler, "Validity and Systematicity of Newspaper Data in Event Analysis," *European Sociological Review* 15, no. 3 (September 1, 1999): 301–22, <https://doi.org/10.1093/oxfordjournals.esr.a018265>; Johan Galtung and Mari Holmboe Ruge, "The Structure of Foreign News," *Journal of Peace Research* 2, no. 1 (1965): 64–91; Tuchman G., *Making News: A Study in the Construction of Reality*, New York: Free Press, 1978.
 - 13 Weidmann, "A Closer Look at Reporting Bias in Conflict Event Data."
 - 14 Davenport and Ball, "Views to a Kill," 445.
 - 15 Although some analysis points to the likelihood of conflict, including guerilla-style strife, shifting to urban settings as a result of global megatrends: Kilcullen, D., *Out of the mountains: The coming of age of the urban guerilla*, Oxford: OUP, 2013.
 - 16 Stathis N. Kalyvas, "The Urban Bias in Research on Civil Wars," *Security Studies* 13, no. 3 (March 2004): 160–90, <https://doi.org/10.1080/09636410490914022>.
 - 17 Barranco and Wisler, "Validity and Systematicity of Newspaper Data in Event Analysis."
 - 18 Allan Dafoe and Jason Lyall, "From Cell Phones to Conflict? Reflections on the Emerging ICT-political Conflict Research Agenda," *Journal of Peace Research* 52, no. 3 (May 1, 2015): 401–13, <https://doi.org/10.1177/0022343314563653>; Mihai Croicu and Joakim Kreutz, "Communication Technology and Reports on Political Violence Cross-National Evidence Using African Events Data," *Political Research Quarterly* (September 29, 2016), <https://doi.org/10.1177/1065912916670272>; Weidmann, "A Closer Look at Reporting Bias in Conflict Event Data."
 - 19 P.J. Shoemaker and Akiba Cohen, "News Around the World: Content, Practitioners, and the Public," *News Around the World: Content, Practitioners, and the Public* (January 1, 2005), 1–409, <https://doi.org/10.4324/9780203959091>.

Reuters and Xinhua and international organisations such as the UN and the OSCE²⁰ report on smaller incidents. Using them will – to some extent – alleviate biases that exists with other sources. However, using these sources is no panacea to mitigate the bias of overreporting of events with large numbers of casualties. For example, reporting by international organisations is curtailed by mandates and battle lines and will be incomplete. Blogs – such as the *Long War Journal* – need to grab readers’ attention and select their stories accordingly. Social media sources require uninterrupted phone coverage. Phone coverage is uneven in conflict zones and regimes curtail access during active military campaigns.²¹ Hence, using sources with a greater ability to report on smaller incidents will lead to the inclusion of other biases in the data.

These biases illustrate the need to know how key sources of data are produced and what its source-reporting characteristics are. Without this knowledge, it is entirely possible to base policies on an incorrect analysis. For example, it is a consistent finding that democratic regimes experience more ‘terrorist’ attacks than autocratic regimes. Without knowing the sources, this could lead to policies to assist (semi-)democratic regimes in overcoming terrorist threats. However, research shows that the difference between both regimes is not real but is a product of press freedom.²² Democratic regimes tend to be more open about the reporting of terrorist incidents than autocratic regimes. Hence, a policy directed towards democratic regimes would be misguided.²³ The whole point of this example is that the awareness of data requires specialised knowledge at governmental departments.

20 Daniel Wigmore-Shepherd, “Urban Bias in Media Reporting – Aceddata – Crisis,” accessed October 4, 2017, <http://www.crisis.aceddata.com/urban-bias-in-media-reporting/>.

21 Anita R. Gohdes, “Pulling the Plug: Network Disruptions and Violence in Civil Conflict,” *Journal of Peace Research* 52, no. 3 (May 1, 2015): 352–67, <https://doi.org/10.1177/0022343314551398>.

22 Konstantinos Drakos and Andreas Gofas, “The Devil You Know but Are Afraid to Face: Underreporting Bias and Its Distorting Effects on the Study of Terrorism,” *Journal of Conflict Resolution* 50, no. 5 (October 1, 2006): 714–35, <https://doi.org/10.1177/0022002706291051>; Matthew A Baum and Yuri M Zhukov, “Filtering Revolution: Reporting Bias in International Newspaper Coverage of the Libyan Civil War,” *Journal of Peace Research* 52, no. 3 (May 2015): 384–400, <https://doi.org/10.1177/0022343314554791>; Moran Yarchi et al., “The Impact of Political Context on News Coverage: Covering Qatar in the Israeli Press,” *Media, War & Conflict* (July 13, 2017), <https://doi.org/10.1177/1750635217711202>.

23 See for a similar example the idea that phone coverage increases violence, while the relationship is in fact that phone coverage increases the reporting on violence, Dafoe and Lyall, “From Cell Phones to Conflict?” Weidmann, “A Closer Look at Reporting Bias in Conflict Event Data.”

Bias #4: Data-gathering is limited to urban elites

A fair amount of analyses of political violence are conducted by means of case studies. Typically, such case studies focus on key informant interviews as a primary data source, in addition to using literature reviews, (social) media analysis and events. Key informants are often members of a country's urban elites who are accessible. I.e. they are known to the outside world and they preferably speak English. This introduces a significant scope for bias in conflict analysis because the views, frames and behaviours of non-accessible urban elites, rural elites and ordinary citizens – in either rural or urban settings – are often not adequately taken into account. The extent to which this is problematic depends on the objective and focus of the analysis. In any case, it easily creates top-down, overly-ideologised, binary and overly intellectualised interpretations of conflict.²⁴

24 Kalyvas, "The urban bias in research on civil wars".

3 Mind the frame: bias at the data-analysis stage

Once research data on political violence in conflict settings has been gathered, the mindset of policy-makers and researchers becomes, in a sense, their greatest enemy. It is a well-known fact that human analysis, judgment and decision-making – the staples of any purposeful human activity – often take place on the basis of heuristics that risk introducing serious bias. Only a part of the human mind is wired to produce rigorous thinking without training, procedure or peer review but this part is not commonly used.²⁵ For a large part, this is due to the limitations of human mental processes of assessing representativeness, availability and probabilities.²⁶ Fortunately, good research is characterised by a number of safeguards to mitigate such processes from taking hold of research findings, but a few common biases are nevertheless worth enumerating.

Bias #5: Conflict framing

Researching conflict is particularly complex once a significant level of violence has been reached. This is because the occurrence of violence accelerates a partisan and partial framing of the conflict on the basis of political interests. Parties need to justify the use of violence for reputational, moral and practical reasons. For example, the Israeli government generally refers to Palestinian stabbing attacks as ‘acts of terrorism’, which suggests a powerful ‘security-first, anti-terrorism’ frame. This narrative is generally replicated in the international media and can easily seep into research efforts as well.²⁷ The problem of conflict frames makes it important to understand the research hypotheses and questions underlying any research work. What are the conceptual and normative frames from which research efforts depart? Producers of the type of conflict analysis under scrutiny in this brief – such as the media, think-tanks and perhaps also embassies – do not typically reflect on this question to a great extent. Whether the

25 Kahneman, D., *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 2011; Amos Tversky and Daniel Kahneman, “Judgment under Uncertainty: Heuristics and Biases,” *Science* 185, no. 4157 (September 27, 1974): 1124–31, <https://doi.org/10.1126/science.185.4157.1124>.

26 Kahneman et al. (2008), *op.cit.*

27 Yet, the evidence does not suggest a terrorist intent or an organisation behind many of these attacks, but rather the humiliation and desperation that are the result of prolonged Israeli occupation of the Palestinian territories. See for example: Chabon, M. and A. Waldman (eds.), *Kingdom of olives and ashes: Writers confront the occupation*, London: 4th Estate, 2017.

consumers of such conflict analysis do so, like ministries of foreign affairs, is an open question. Time and capacity are typically in short supply.²⁸

One part of such conflict framing bias is the uncritical acceptance of global frames for local conflicts. Many conflicts acquire a global framing that may or may not be related to the local conflict frame. This happens when the local conflict obtains significance in a broader global clash or because local conflict parties have foreign patrons or sponsors with their own interests. Most global conflict frames resonate easily, are highly simplified and are rather partial. For example, the US and Saudi Arabia tend to portray a range of conflicts in the Middle East as a Shi'a-Sunni affair while the utility of this frame for actually understanding what happens is rather doubtful. It underappreciates, for example, the intra-Shi'a and Christian dimensions of the conflict in Lebanon, the anti-Iranian stance of many Iraqi Shi'a and the close Sunni-Shi'a-Kurdish alliance that used to support the regime of President Assad. Such frames bias the interpretation of events by misdiagnosing the driving forces and the motivations of actors.

Another, related part of such conflict framing bias is that local interests are insufficiently examined. As conflicts are highly partisan affairs, it is obvious that conflict parties seek to create a narrative that is conducive to their cause. Yet, the possibilities for doing so have radically increased with the advent of global mass media, online media and social media. The ability of researchers to really gain an understanding of the differences in what parties say, think and do, remains limited and is further obscured by the amount of information available. At times, conflict parties even go as far as to develop their own 'evidence base' for their own narrative.²⁹ The risk here is that such local interests are insufficiently scrutinised or that shifts are not detected in time. For example, is Turkey still a state suffering from an insurrection or an autocracy repressing a minority? Are the Hashd al-Sha'abi largely Iranian proxies or Iraqi entities with significant autonomous agency and ties to Iran?

28 The speed of the analysis – and the greater relevance this creates – is generally traded for accuracy and depth.

29 Consider, for example, the 'NGO Monitor' (<http://www.ngo-monitor.org/>) that many researchers consider to be highly partial in the Israeli-Palestinian conflict but designates itself as a neutral NGO watchdog. For an anecdotal discussion of the role of the media in Russia in 'creating narratives': Pomerantsev, P., *Nothing is true and everything is possible*, London: Faber & Faber, 2016. One of the counter-responses is citizen investigative journalism and fact-checking sites like Bellingcat. See: <https://www.bellingcat.com/> (accessed 16 October 2017). For research problems related to social media analysis: Zeitzoff, T., 'How social media is changing conflict', *Journal of Conflict Resolution*, Vol. 61 no 9 (2017): 1970-91.

Bias #6: Insufficient reflection on description biases of new data sources

The information revolution has created at least three new sources of information on political-violence trends: a) satellite data; b) 'real-time' information; and c) text mining. All three 'innovations' have the potential to increase the quality of the analysis of political violence. Yet, there are also a few biases associated with these 'new information carriers':

- *Satellite data*: Satellite data is a form of remote sensing that originally allowed for the identification of bomb damage, the monitoring of troop movements and the mapping of military installations as well as minefields.³⁰ Recent non-military applications are counting refugees (through tents) or assessing the quality of public service provision (through light emission). A clear advantage of satellite information is that reporting does not require an observer, so that events of political violence can be captured without the risk of witness or urban elite bias. Most problematic is that it does not allow for the identification of actors and can only be used for analysing political violence such as shelling and (conventional) military presence. Moreover, satellite imagery is available at time intervals that may not be suitable for analytical purposes. Satellites are therefore most appropriate as an additional source to control for bias, rather than a stand-alone source to assess overall trends.
- *Real-time information*: A great deal of data on political violence is made public with significant delays. For example, the well-known Uppsala Data Conflict Program (UDCP) releases conflict data on a yearly basis, meaning that data for January 2014 is only released in July 2015. The real-time provision of data on political violence is, however, becoming more prevalent. The main disadvantage is that the quality of information declines with the speed of its availability. Collecting credible information takes time. For this reason, Syrian human rights organisations document individual killings with a delay (they need to locate family members, for example).³¹ One new application is promising, however. A recent experiment in Congo set up a messaging

30 Dave Donaldson and Adam Storeygard, "The View from Above: Applications of Satellite Data in Economics," *The Journal of Economic Perspectives* 30, no. 4 (2016): 171–98; Frank D. W. Witmer and John O'Loughlin, "Satellite Data Methods and Application in the Evaluation of War Outcomes: Abandoned Agricultural Land in Bosnia-Herzegovina After the 1992–1995 Conflict," *Annals of the Association of American Geographers* 99, no. 5 (October 30, 2009): 1033–44, <https://doi.org/10.1080/00045600903260697>; Nicholai Lidow, "Rebel Predation: Remotely Sensed Evidence from Liberia," *Working Paper*, 2011.

31 Very reliable conflict data could be collected through well-designed randomised surveys. L. Alison Smith et al., *Conflict Mapping in Sierra Leone : Violations of International Humanitarian Law from 1991 to 2002: Preliminary Edition for the Opening of the SCSL Courthouse*, Freetown, Sierra Leone: No Peace Without Justice, 2004.

system for very remote villages in Northern Kivu where villagers received a cell phone with credit to report incidents of violence.³² The data proved to be highly credible (because researchers can select a representative sample of villages) and came in real time. Similar cell phone-based systems could be used by researchers or funded by policy-makers in other conflict contexts.

- *Text mining*: Another data ‘innovation’ is text mining, which usually takes two forms. One group is machine-coded data through software that ‘reads’ reports and codes them into data.³³ A second group of text mining tools is designed to carry out specific tasks such as identifying key concepts, topics or recognising sentiments. The main concern is that at the moment a great deal of automated data is not sufficiently credible. For example, it is known that databases are biased in recognising places, often cannot deal with conflicting reports and are unable to deal with complex analytical tasks.³⁴ Moreover, many databases focusing on text mining are insufficiently validated – meaning that original documents are compared to the actual output of the automated tool.³⁵ Hence, at this moment in time caution is warranted when making use of text mining tools – although technological development is rapidly catching up.

32 Peter Van der Windt and Macartan Humphreys, “Crowdseeding in Eastern Congo Using Cell Phones to Collect Conflict Events Data in Real Time,” *Journal of Conflict Resolution* 60, no. 4 (June 1, 2016): 748–81, <https://doi.org/10.1177/0022002714553104>.

33 Automated or machine-coded event-data has been used since the 1970s (WEIS, COPDAB, KEDS, TABARI, PETRARCH) but the internet revolution and self-learning algorithms have rapidly improved the number of applications and the quality of coding. Prominent examples include databases such as GDELT, ICEWS and PHOENIX. They all rely on Natural Language Processing.

34 Jesse Hammond and Nils B. Weidmann, “Using Machine-Coded Event Data for the Micro-Level Study of Political Violence,” *Research & Politics* 1, no. 2 (2014); Idean Salehyan, “Best Practices in the Collection of Conflict Data,” *Journal of Peace Research* 52, no. 1 (January 1, 2015): 105–9, <https://doi.org/10.1177/0022343314551563>; Haewoon Kwak and Jisun An, “A First Look at Global News Coverage of Disasters by Using the GDELT Dataset,” in *Social Informatics*, Lecture Notes in Computer Science (International Conference on Social Informatics, Springer, Cham, 2014), 300–308, https://doi.org/10.1007/978-3-319-13734-6_22.

35 Salehyan, “Best Practices in the Collection of Conflict Data.”

Bias #7: Single case studies lead to overconfident generalisations

A seventh bias is that many case studies of political violence in conflict settings are qualitative one-off affairs that neither benefit from longitudinal nor comparative repetitions.³⁶ This makes them of limited utility in terms of how far they can be generalised across the whole scope of a conflict, time or a variety of conflicts. If they are also overly based on one particular data source – e.g. document analysis, semi-structured interviews with key informants or social media analysis – instead of mixing several sources of evidence, the risk of overgeneralisation increases. An additional problem is that many such qualitative single case studies provide little clarity on their methodology, sourcing and research limitations. This while many consumers of such analysis do not systematically assess the methodological rigour of analysis when using it in the process of policy formulation. The bias this introduces is a significant scope for personal opinions, frameworks and error.

Bias #8: Description problems with local conflict databases

Local databases tracking political violence are increasingly common reference points for conflict analysis for policy-makers and researchers alike. Policy-makers specialising in particular conflicts may be tempted to rely strongly on some of these databases. The advantages are clear: these databases are often generated with local networks, granular and context specific. However, there are problems with their quality as well as in the ways they describe violence.³⁷

For example, they often do not feature definitions of what exactly constitutes a 'battle', a 'territorial takeover' or a 'civilian killing'. Hence, data entered on one day may not be comparable with data entered on another day. In addition, there are many that are insufficiently transparent about how they gather data (for example, how algorithms scrape social media work). Finally, organisations have been known to change definitions in the process of data-gathering without explaining these changes. An example from Syria illustrates the problems that can subsequently arise. **Figure 3** presents five maps of territorial control in Syria. Each map is produced by a different organisation for the same date (October 16, 2017) and presents the 'main actors' based on 'data'. As is evident from these maps, there are clear differences. ISW and the Carter Center present

36 This is because conflict is a contemporary, real-world event. Conflict analysis focuses on 'why' and 'how' questions.

37 Quality is often a function of organisational design and maturity. When local organisations are newly established or lack experience in systematically collecting information while starting that they report on 'data', the risks are highest. For instance, over 75% of the 30 Syrian organisations reporting on the conflict were set up after 2014.

a vast no man's land in the south of the country while the other three suggest control. Moreover, the alleged territorial control of the three maps that are fully filled still feature major differences concerning the specific locations controlled by the Islamic State as well as which opposition controls the Syrian province of Idlib.

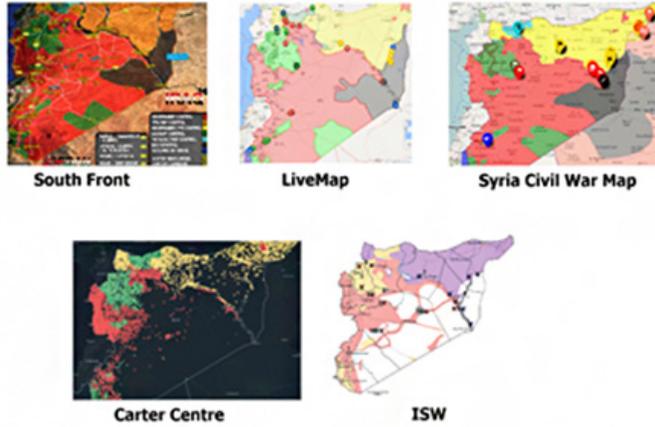


Figure 3 Maps of territorial control in Syria as of October 16, 2017

4 What can policy-makers and researchers do to mitigate bias?

Researchers face bias problems during both research design and research implementation. Policy-makers potentially face bias problems twice, namely when acting as researchers themselves, and when commissioning research. An example helps illustrate why both of these tasks are impacted by bias. A 2017 OECD study of the donor understanding of local political legitimacy and inclusiveness in relation to conflict found that this was generally incomplete and inadequate.³⁸ Donor understanding of the typically fragmented and highly contested politics of fragile societies did not often go beyond the formal representatives of their governments and administrations. This was in part the result of poor conflict analysis that featured a number of biases noted in this brief. Nevertheless, donors continued to work on the basis of standardised templates that, unfortunately, made little sense in the intervention context. This includes issues such as pushing for the introduction of a highly centralised state in Afghanistan that has a history of centuries of decentralised governance, and pursuing national elections as a critical pathway for enhancing legitimacy despite the overwhelming electoral advantages for those with weapons, socio-ethnic networks and money. It is unlikely that such initiatives will be successful. Moreover, the flood of quantitative data which is now available ‘can convey a false sense of accuracy of the precision of the data if not considered carefully’.³⁹ This can lead to an overly confident analysis of conflict, or even unwarranted policy interventions.⁴⁰

38 The study is based on four case studies: Afghanistan, Somalia, South Sudan and Timor-Leste. See: Van Veen, E. and V. Dudouet, *Hitting the target but missing the point? Assessing donor support for inclusive and legitimate politics in fragile societies*, Paris: OECD INCAF, 2017.

39 Kristine Eck, “In Data We Trust? A Comparison of UCDP GED and ACLED Conflict Events Datasets,” *Cooperation and Conflict* 47, no. 1 (March 1, 2012): 124–41, <https://doi.org/10.1177/0010836711434463>.

40 Price and Ball, “Selection Bias and the Statistical Patterns of Mortality in Conflict,” 270.

Table 1 Mitigating bias on conflict data-gathering

Bias	Mitigation from a policy perspective	Mitigation from a research perspective
Bias #1: Missing data	<ul style="list-style-type: none"> - Allocate targeted funding when/ where important data is not available to mitigate the effect of biased data on conflict analysis and policy responses. 	<ul style="list-style-type: none"> - Rigorously assess the extent to which data and information is missing and report on it; - If data is missing on a key phenomenon, gather additional data.
Bias #2: Short field presences results in limited data-gathering	<ul style="list-style-type: none"> - Establish conflict research efforts that enable longitudinal research; - Identify analysis or sources to complement single case studies; - Require that analyses are peer-reviewed. 	<ul style="list-style-type: none"> - Work with a reliable local partner to enable longer-term research and recurrent shorter visits; - Ensure that a diversity of perspectives is purposefully captured during short visits; - Ensure a mixed-methods approach that includes document/archival analysis, interviews, participant and direct observations, quantitative data as well as study of physical artefacts.
Bias #3: Witness, reporting and description practices	<ul style="list-style-type: none"> - Develop a genuine realisation that today's conflict databases do not bring absolute truths; - Consider creating a specialised desk that vets research by identifying biases, checking methodologies, source quality and reputation of institute and author; - Engage in source classification and qualification; - Build insight into which bias exists, to enable usage for particular trend-analysis. 	<ul style="list-style-type: none"> - Engage in targeted sampling strategies to account for witness reporting and description bias; - Rely on clever combinations of sources to mitigate the effects of various types of biases; - Further develop statistical techniques to account for bias.
Bias #4: Data-gathering is limited to urban elites	<ul style="list-style-type: none"> - Establish long-term conflict research efforts that include urban and rural dimensions and allow time to build corresponding enabling networks; - Commission public opinion surveys data if not available. 	<ul style="list-style-type: none"> - Include focus groups or workshops with the urban poor and rural poor/elites if needed; - Seek out more traditional elites like clerical leaders or customary chiefs; - Use public opinion survey data if available; - Ensure adequate linguistic skills to capture local media and analysis.

Table 2 Mitigating bias in the data-analysis stage of research

Bias	Mitigation from a policy perspective	Mitigation from a research perspective
Bias #5: Conflict framing	<ul style="list-style-type: none"> - Ensure access to, and time for, the study of non-familiar sources and deeper perspective of a particular conflict; - Explicitly include requests for the perception of 'the other' in commissioned research, and budget accordingly; - Organise challenge or 'red team' sessions with some frequency; - Challenge or interrogate broad label-type thinking like 'the Shi'a' or 'the rebels'. 	<ul style="list-style-type: none"> - Ensure adequate historical understanding of the relevant society and conflict to correct for geopolitical hypes; - Seek out unusual interviewees during field work who can offer correctives; - Use several dimensions to examine interests from different perspectives in parallel (e.g. conflict actors, business interests, popular views, victims/perpetrators); - Work with a reliable local research partner.
Bias #6: Insufficient reflection on the pros and cons of new data sources	<ul style="list-style-type: none"> - Develop a genuine realisation that new techniques are no panacea and have own biases; - Commission longitudinal research (e.g. multi-annual research programmes) that allow for testing and calibration of new data sources. 	<ul style="list-style-type: none"> - Develop innovative mixed methods research approaches that combine tried-and-tested methods with new data sources.
Bias #7: Single case studies lead to overconfident generalizations	<ul style="list-style-type: none"> - Commission more longitudinal and comparative research where possible; - Rate analysis according to the extent to which they employ mixed methods; - Pay attention to the reputation of the research institution and author. 	<ul style="list-style-type: none"> - Conduct at least two case studies where possible and appropriate; - Ensure a mixed-methods approach in case research to improve internal validity.
Bias #8: Quality problems with local conflict databases	<ul style="list-style-type: none"> - Consider quality markers when using local databases such as: <ol style="list-style-type: none"> a) explicit discussions about how data is collected; b) transparency about sources employed; c) clarity about definitions; d) clarity about algorithms and machine coding including reliability testing. Note: without such markers it is advisable not to use databases. 	<ul style="list-style-type: none"> - Assess the comprehensiveness of data by comparing with other sources and/or deploying comparison techniques; - Ensure inclusion of specific overviews of sources used, reporting characteristics of sources and identification of known biases; - Develop a reliability classification of sources and databases.

Together, both tables suggest three general conclusions as to how bias in the analysis of political violence can be reduced:

- Source triangulation is a critical matter in each piece of research and especially when the perfect mix of research methods is not possible for reasons of feasibility or financial resources. Bias problems in data-gathering and analysis can partly be

mitigated by triangulating different sources with a diverse nature. Think, for example, of combining available quantitative data like public opinion research, or existing conflict databases, with qualitative data such as (multiple) case studies. This is something both researchers and policy-makers need to keep in mind – and account for – more explicitly when analysing political violence.

- To address information gathering bias, good (and sometimes technical) insight into the sources used is crucial. Only then can policy-makers mitigate information gathering bias for example by commission specific research questions and allocating funding (e.g. to local partners working on areas/topics that are underreported - **Table 1**).
- To address analysis bias, there need to be good structures in place to assess the credibility and comprehensiveness of outside analysis. Moreover, there could be structures in place to develop counter-narratives (e.g. red-teaming), as well as possibilities to fund longer term research initiatives in order to allow longitudinal research efforts (**Table 2**).

On a final note, an analysis of political violence in conflict situations is only one of the factors that typically enters a policy formulation process and/or political decision-making. Ideally, the formulation of policies intended to address conflict would depart from a sound understanding of the local nature and dynamics of conflict. In reality, domestic political considerations and ‘ancillary’ international effects are at least equally powerful ingredients. Yet, the trade-offs and price of prioritising one of these three dimensions over the other two will not be known unless there is high-quality analysis available with biases being recognised as much as possible. This is a vital ingredient of any responsible policy formulation process, irrespective of its outcomes. And – unfortunately – an ingredient that is insufficiently available.