

I. Progress in the collection of quantitative data on collective violence

MICHAEL BRZOSKA

Over the past decade the number and quality of data sets on violence have increased substantially.¹ There are now a good number of ambitious and comprehensive data collection projects, and a multitude of data sets available on specific aspects of violence. This growth in the availability and diversity of quantitative data has been driven by academic study of the causes and consequences of various forms of violence, and a wider recognition of its costs in terms of human security and human development.

There has also been a broadening in the scope of the data. A few decades ago, the focus of quantitative data sets was on interstate wars. Following recognition that intrastate wars had become the dominant form of warfare, a number of data sets offering comprehensive coverage of additional forms of armed conflict became available. Most of the more recent data sets do not stick to what might be called the 'standard definition' of armed conflicts, in which organized armed forces, of which at least one must be a government, fight each other for political objectives. While there are differences in how far data collection efforts extend their coverage beyond wars and armed conflicts, however, they are all informed by a broader perception of human suffering through the intentional infliction of violence. This section describes and discusses select additions to the pool of freely available data sets.

Major questions remain unanswered about the scope of and trends in violence inflicted by one group of people against another. Perhaps the most important is the question of whether there has been a general, progressive decline in the level of human-inflicted violence over recent decades. Prominent proponents of this thesis, such as Andrew Mack, the lead author of the Human Security Report series which covers collective violence with political objectives, and Steven Pinker, who looks at violence of various types, have produced impressive evidence to support their claim.² However, scepticism remains strong and has increased in the wake of recent wars in Africa,

¹ The focus of this section is physical violence against humans exerted by groups for political purposes, in the form of wars, armed conflicts, genocide and terrorist acts. However, other forms of physical violence have also attracted attention, as discussed below. Gleditsch, K. S., Metternich, N. W. and Ruggeri, A., 'Data and progress in peace and conflict research', *Journal of Peace Research*, vol. 51, no. 2 (Mar. 2014), pp. 301–14. For earlier assessments of data sets on collective violence in SIPRI Yearbooks see Seybolt, T. B., 'Measuring violence: An introduction to conflict data sets', *SIPRI Yearbook 2001*, pp. 81–96; and Brzoska, M., 'Collective violence beyond the standard definition of armed conflict', *SIPRI Yearbook 2006*, pp. 94–106.

² See the website of the Human Security Report Project, <<http://www.hsrgroup.org/>>; and Pinker, S., *The Better Angels of our Nature* (New York, NY: Viking, 2011). For the most recent annual update on his findings see Pinker, S., 'Now for the good news: Things really are getting better', *The Guardian*, 11 Sep. 2015.

Table 6.1. Violent death: differing concepts, different sources

Type of violence	Estimated no. of victims	Source
Direct conflict deaths: deaths as a result of armed conflicts, political violence and terrorism	70 000	GBAV
Intentional homicides: deaths as a result of interpersonal violence, gang violence and economically motivated crimes	377 000	GBAV
Victims of legal interventions: deaths of civilians caused by law enforcement and state security forces during legal interventions	19 000	GBAV
Unintentional homicides: deaths as a result of 'accidental killings'	42 000	GBAV
Battle-related deaths	60 000	UCDP
Victims of intentional attacks on civilians by governments and formally organized armed groups	7 000	UCDP
Victims of communal and organized armed conflict where none of the parties is the government of a state	6 000	UCDP
Victims of terrorist attacks	20 000	GTD

GBAV = Global Burden of Armed Violence; GTD = Global Terrorism Database; UCDP = Uppsala Conflict Data Programme.

Note: Estimates of violent death, rounded to thousands, annual average data (2007–12 for GBAV; 2009–14 for UCDP and GTD).

Sources: Author's calculations from UCDP and GTD data sets, and Geneva Declaration, *Global Burden of Armed Violence 2015: Every Body Counts*, chapter 2, <<http://www.genevadeclaration.org/measurability/global-burden-of-armed-violence/global-burden-of-armed-violence-2015.html>>.

Europe and particularly the Middle East.³ A related hotly debated question is that of the human costs of violence beyond those which are immediately recognizable as such.⁴ Finally, although there is generally speaking considerably more data available, progress has been uneven. In particular, the collection of data on violence by small groups of people or that results in few or no fatalities continues to be challenging and problematic.

New data sets provide an improved basis for the study of many forms of violence but authors continue to come to differing conclusions about the trends in, and causes and consequences of collective violence.⁵ This is partly

³ For detailed assessments questioning such sweeping conclusions see sections III and IV in this chapter.

⁴ See Krüger, J. et al., 'It doesn't add up: Methodological and policy implications of conflicting casualty data', eds T. B. Seybolt, J. D. Aronson and B. Fischhoff, *Counting Civilian Casualties: An Introduction to Recording and Estimating Non-military Deaths in Conflict* (Oxford University Press: Oxford, 2013), pp. 53–77; and Pérouse de Montclos, M., Minor, E. and Sinha, S. (eds), *Violence, Statistics and the Politics of Accounting for the Dead* (Springer: Heidelberg, 2016).

⁵ Collective violence has been defined by the World Health Organisation (WHO) as the 'instrumental use of violence by people who identify themselves as members of a group—whether this

due to differences and gaps in the available data, but generally more so to the choice of one among several possible data sets that reflect differences in the scope and form of violence. The increase in the number of data sets on human-inflicted violence, and their use of different definitions and different methods of data collection, allow for a more diverse and thus less unified view of the important factors in driving and preventing violence.

Major advances in the collection and availability of data are reviewed below, with a particular focus on the widening of conceptual ambitions, increased precision in the recording of the occurrence of violence and innovations in source-mining techniques. The section concludes by discussing major ongoing problems, such as remaining data gaps and issues of data collection.

Major progress in the availability of data on armed violence

Conceptual ambition

Conflicts between armed forces that fight for political causes continue to be at the centre of academic, political and humanitarian interest. Major wars such as the current war in Syria, which began in 2011, lead to large-scale devastation and mass suffering. The trend in the number of armed conflicts—about which there are a number of sources, all of them with partially differing definitions—as well as the number of victims, which as shown below can be defined and estimated in various ways, are probably the most often quoted indicators of the trend in violence.⁶

However, in terms of human suffering, conflicts between armed forces may not rank as the worst. When considering deaths from human-inflicted violence, a broader perspective on armed violence can be of major interest. As can be seen from the data in table 6.1, some forms of armed violence have claimed more lives in recent years than directly died in battle. For the most recent six-year period for which data is available from the two sources used in table 6.1, the estimates of the average annual number of battle deaths are 70 000 and 60 000 respectively. The number of indirect victims of war, for instance through a deterioration in health systems or starvation, has been much larger in a number of cases, but there are no comprehensive global

group is transitory or has a more permanent identity—against another group or set of individuals, in order to achieve political, economic or social objectives’, see WHO, ‘Collective violence’, Fact sheet, 2002.

⁶ Institutions that collect data on armed conflicts in addition to the ones reported in table 2 include the Correlates of War Project; the Center for Systemic Peace; the Working Group for Research on the Causes of War, University of Hamburg; and the Heidelberg Institute for International Conflict Research. For a larger listing of sources on armed conflicts and other forms of conflict see Paul Hensel’s listing of data sources for the International Studies Association, <<http://www.paulhensel.org/compendium.html>>.

statistics. Some conflict deaths are also counted by data sources that collect data on terrorist attacks. The most prominent of these is the Global Terrorism Database (GTD) maintained by the Study of Terrorism and Responses to Terrorism (START) project at the University of Maryland.⁷ Not all terrorist attacks occur in armed conflicts, however, and in some definitions of terrorism, attacks directed against organized forces are even excluded.

Two institutions that collect data on broader forms of violence merit particular attention. The Uppsala Conflict Data Program (UCDP) collects data on two additional forms of violent death: one-sided violence and non-state violence. It is by far the most used data set in academic work on armed violence.⁸ The best estimate for the total number of victims from these three categories of violence together in 2014 is reported as 126 059, well above the average for the six-year period 2009–14 of about 73 000, mainly due to the conflict in Syria.⁹

An even broader approach to violence is taken by the Geneva Declaration on Armed Violence and Development, an intergovernmental programme supported by more than 100 governments, which has the objective of promoting tailored interventions and programmes to reduce the number of victims of armed violence.¹⁰ Based on a human security perspective, its Global Burden of Armed Violence (GBAV) data collection effort adds victims of intentional and ‘accidental’ homicide, as well as victims of legal state interventions to those of armed conflict, genocide and terrorism.¹¹ In total, the GBAV estimates an average of 508 000 annual victims of violent death per year for the period 2007–12. By far the largest number of these are linked to intentional homicide, which covers various forms of collective violence such as gang violence, communal violence and illegal forms of state violence, but also homicides by individuals. The way in which the data is collected, primarily from homicide statistics (see below), does not allow for the separation of collective intentional violence from individual violence. Nor is a distinction made with respect to the objectives of the violence, such as political aims or economic gain.

⁷ National Consortium for the Study of Terrorism and Responses to Terrorism (START), <<https://www.start.umd.edu/>>.

⁸ UCDP data was used as the primary data source on violence in almost a quarter of the more than 350 incidences of the use of databases on violence in four leading academic journals focused on quantitative research on conflict and violence (*Journal of Peace Research*, *Journal of Conflict Resolution*, *International Interactions* and *Journal of Conflict Management and Peace Science*) in 2010–15. Other data sources, such as the Correlates of War Project, each made up below 10% of all uses. Annika Maretzki assisted the author with collecting this information.

⁹ Melander, E., ‘Organized violence in the world, 2015: An assessment by the Uppsala Conflict Data Program’, UCDP Paper no. 9 (Uppsala University, Department for Peace and Conflict Research: Uppsala, 2015).

¹⁰ See the Geneva Declaration on Armed Violence and Development, <<http://www.genevadeclaration.org/the-geneva-declaration.html>>.

¹¹ Geneva Declaration, *Global Burden of Armed Violence 2015: Every Body Counts*, chapter 2 (Geneva Declaration: Geneva, May 2015).

Data sets on a variety of forms of violence allow for a more comprehensive as well as a more nuanced analysis of organized violence. If, for instance, the analytical purpose is to study the use of violence for political purposes, it makes sense to expand the scope to one-sided and non-state violence, as well as terrorism. Additional data sets allow for a broader, but also more detailed, assessment of the human cost of violence, identifying for instance whether it is by governments, non-state armed groups with political objectives or other violent actors.

Georeferenced events data

A number of collection efforts go beyond those described above by focusing on the reporting of violent events in detail and locating them geographically.¹² They also extend beyond violence in armed conflicts. While these do not necessarily tell us much more about trends in violence, event data allows more detailed analysis of the incidence, causal conditions and effects of armed violence, making it possible to analyse, for instance, in which parts of a country armed conflict can be found.

The basis for disaggregated data is the collection of detailed information on violent events, such as fighting among armed groups or a terrorist attack in a particular location. Event data has a long tradition in conflict research. A prominent example is the Militarized Interstate Dispute (MID) data set, an element of the larger Correlates of War data collection programme. MID records messages and activities between states that involve the threat, display or use of military force short of war by one state against another, and classifies them on a scale of intensity.¹³ Other examples are the various data sets on terrorism.¹⁴ Data on events are also the foundation for traditional data sets on wars and armed conflicts. For these, however, it is sufficient to know that there was one battle among armed forces within a year in a particular country. Event data sets aim to collect all the relevant incidences.

Producing comprehensive event data sets is resource-intensive and time-consuming. Recently, however, several geo-referenced data sets on aspects of armed violence with a wide geographical coverage have become available (see table 6.2). These data sets have different starting points. The UCDP Georeferenced Event Data (GED) data set starts out from the existing UCDP data sets on armed conflict, non-state and one-sided armed violence, but extends its coverage to all forms of armed violence above the threshold

¹² Bernauer, T. and Gleditsch, N. P., 'New event data in conflict research', *International Interactions*, vol. 38, no. 4 (Sep. 2012), pp. 375–81; and Salehyan, I., 'Best practices in the collection of conflict data', *Journal of Peace Research*, vol. 52, no. 1 (Jan. 2015), pp. 105–09.

¹³ Correlates of War Project, 'Militarized interstate disputes, 1816–2010', Version 4.1, .

¹⁴ See e.g. Sheehan, I. S., 'Assessing and comparing data sources for terrorism research', eds C. Lum and L. Kennedy, *Evidence-based Counterterrorism Policy* (Springer: Wiesbaden, 2013), pp. 13–40.

Table 6.2. Forms of armed violence and their coverage in selected event databases

Data set source and regional coverage	Type of information	Data on armed conflicts	Coverage of additional forms of violence
UCDP GED Global	Instances of political violence by date, location, actors and intensity of violence (fatalities)	Above threshold of 1 fatality	Covers all forms of armed violence by an organized actor with political objectives above threshold of 1 fatality
ACLED Africa, South and South East Asia	Events by dates, locations, groups involved, event types and changes in territorial control	Battles and other uses of force without fatality threshold	All 'politically violent events' including violent and non-violent (e.g. troop movements), uses of force by a group with a political purpose or motivation as well as violent and non-violent protests
SCAD Africa, Mexico, Central America and the Caribbean	Events by location, actors, intensity	Limited to social conflict events within armed conflicts	Protests, riots, strikes, inter-communal conflict, government violence against civilians and other forms of social conflict
Worldwide Atrocities Dataset	Events by location, actor, type of violence, perpetrators, victims	Limited to use of acts of violence against minimum of 5 civilians	All forms of intentional violence against non-combatants (civilians)
SPEED Global	Civil unrest, state repression and coup d'état events by location, actors, type and intensity	Limited to use of low-level violence within armed conflicts	Broad coverage of politically motivated violent and non-violent events (e.g. demonstrations, bombings and political prosecutions/arrests)

ACLED = Armed Conflict Location and Event Dataset; SCAD = Social Conflict Analysis Database; SPEED = Social, Political and Economic Event Database; UCDP GED = Uppsala Conflict Data Programme Georeferenced Event Data.

Sources: UCDP GED, <<http://www.ucdp.uu.se/ged/>>; ACLED, <<http://www.acleddata.com/>>; SCAD, <<https://www.strausscenter.org/scad.html>>; Worldwide Atrocities Dataset, <<http://eventdata.parusanalytics.com/data.dir/atrocities.html>>; and SPEED, <<http://www.clinecenter.illinois.edu/data/speed/event/>>.

of one death. Other data sets, sponsored by a variety of research and governmental institutions, come from a broad tradition of the quantitative study of specific forms of political instability, protest and government repression.¹⁵

¹⁵ The latter includes the US Central Intelligence Agency, which, through its sponsorship of the Political Instability Task Force, supports a number of data sets, including the Worldwide Atrocities Dataset (see table 6.2).

The data sets listed in table 6.2 differ from those in table 6.1 with respect to their coverage of the types and scope of violent events. They also contain additional information. There are differences too in the reporting of events and the quality of the data. A comparative analysis of the Armed Conflict Location and Event Data (ACLED) project and the UCDP GED data sets—the two that have received most scholarly attention—found a substantially larger number of errors in the ACLED. However, it covered considerably broader classes of events, including non-violent ones, in a greater number of countries.¹⁶ There seems to be a trade-off, therefore, between coverage and accuracy, due to the resources needed to produce the data.

Data mining methods

Until fairly recently, event data sets almost exclusively relied on major news-producing organizations in Western countries, or publicly available collections of news reporting. ACLED, for instance, continues to collect its information from the Associated Press and Agence-France Presse. There has been a long debate about potential selection bias through relying on such sources.¹⁷ The Internet has greatly expanded the number and scope of easily accessible sources from many regions of the world. Data from social media, such as Twitter, add further information on violent conflict. The growth in Big Data has stimulated the development of sophisticated automated methods of data mining. Large amounts of data are time-consuming to scan using simple search tools. Advanced automated methods offer the prospect of extracting data of interest quickly using relatively few resources. Considerable efforts have therefore been committed to developing advanced automated data processing tools. One such is the Global Data on Events, Location and Tone (GDELT) database, supported by Google Ideas. Information from print, broadcast and web news media in over 100 languages is collected, translated, classified and categorized, including into conflictive and cooperative events.¹⁸

However, there continue to be major problems with automated data mining.¹⁹ Searches for key words or phrases tend to miss ‘latent content’, which only becomes obvious in longer passages of text or through use of context.

¹⁶ Eck, K., ‘In data we trust? A comparison of UCDP GED and ACLED conflict events datasets’, *Cooperation and Conflict*, vol. 47, no. 1 (Mar. 2012), pp. 124–41.

¹⁷ See e.g. Öberg, M. and Sollenberg, M., ‘Gathering conflict information using news resources’, eds. K. Hoglund and M. Öberg, *Understanding Peace Research: Methods and Challenges*, (Routledge: Abingdon, New York, 2011), pp. 47–73; and Hendrix, C. S. and Salhyan, I., ‘No news is good news: Marc and recapture for event data when reporting probabilities are less than one’, *International Interactions*, vol. 41, no. 2 (Apr. 2015), pp. 392–406.

¹⁸ For more information see the GDELT Project website, <<http://gdeltproject.org/>>.

¹⁹ Grimmer, J. and Stewart, B. M., ‘Text as data: The promise and pitfalls of automatic content analysis for political texts’, *Political Analysis*, vol. 3, no. 21 (Summer 2013), pp. 267–97; and Nardulli, P., Althaus, S. L. and Hayes, M., ‘A progressive supervised-learning approach to generating rich civil strife data’, *Sociological Methodology*, vol. 45, no. 1 (Aug. 2015), pp. 148–83.

The same problems are also prone to produce 'false positives'. For instance, automated data mining tools have difficulty in distinguishing reports of actual events from text that comments on such events.

More promising, for the time being, than fully automated systems are hybrid systems that combine machine-based and human-control approaches to data mining. One such project is the Social, Political and Economic Event Database (SPEED), mentioned in table 6.2. Its procedures combine automated machine searches and manual codification of information in a feedback loop, in order to improve the selection reliability of the automated part of the process through 'progressive learning'.²⁰

Gaps and challenges

The increased availability of data on major forms of violence has the potential to improve the basis for decision making on political and humanitarian action as well as quantitative academic research. However, major gaps remain and new issues have arisen, in addition to the permanent challenge of mustering the necessary personnel and material resources to maintain large and complex databases.

Intensity of armed violence

There is a continuing lack of comprehensive, valid indicators of the level of collective violence capable of informing debates such as the one on the long-term trend in violence mentioned above. The growing number of indicators that are now available do not necessarily point in the same direction. While this diversity can contribute to a better understanding of the complexities of collective violence, the only currently available indicator for comparing and unifying the data is direct violent death. Violent death, however, is only one indicator of the costs of collective violence, albeit an important one. It does not for instance include deaths indirectly linked to the use of force, such as those arising from damaged health care or sanitation systems. Nor does it include the psychological effects of the use or threat of armed violence, such as in bombing campaigns. Some data is useful for measuring the humanitarian costs of violence, such as on refugees and internally displaced persons, but this is not easy to aggregate or to compare with data on violence in order to provide a more comprehensive picture.²¹ Interesting survey-based methods have been proposed to better measure the intensity of armed violence, but these are difficult to implement.²² Going beyond victims, other potential

²⁰ Nardulli et al. (note 19).

²¹ A useful point of reference for such data is the United Nations High Commissioner for Refugees' Statistical Online Population Database, see <<http://www.unhcr.org/45c06c662.html>>.

²² Smith, R. P., 'The economic costs of military conflict', *Journal of Peace Research*, vol. 51, no. 2 (Mar. 2014), pp. 245–56.

avenues for measuring the intensity of violence might be to collect data on the number of perpetrators involved in acts of violence or the type of weaponry involved.²³

Differences in the data, and more importantly disagreement about which forms of collective violence to consider, seem to have contributed to the lack of an indicator in the recently agreed Sustainable Development Goals (SDGs). SDG 16 has to ‘significantly reduce all forms of violence and related death rates everywhere’ as its first objective, but there is no indication of how violence should be measured.²⁴

Consolidation of data

A related issue is the need to narrow the gaps between data sets arising from different methods of data collection. Differences among data sets that derive from differing concepts and definitions of various forms of violence should be an asset, because they should complement each other. However, different methodologies should not, in principle, lead to different answers about important academic or policy questions. Current differences among data sets continue to be shaped by diverse approaches to data collection. The data on victims of armed violence reported above provides a good example. The data for the UCDP data sets is collected from a large number of media sources, as well as reports from a number of governmental and non-state organizations, and includes information on victims. The data collated by the GBAV project, on the other hand, primarily comes from national and international organizations, such as the World Health Organization and the UN Office on Drugs and Crime, building on mortality data collected from death certificates or from health surveys.²⁵

Conflict processes

Other requirements for additional and new data on armed violence would include the study of: escalation processes in collective violence, prevention of the use of violence, variability in the use of violence and other forms of contestation in conflict and de-escalation from violence. The new data sources with their more fine-grained information on collective violence provide an improved foundation for studying conflict processes but need to

²³ This is already recorded with respect to the use of small arms in the GBAV. New data sources on the use of arms in conflict, such as iTrace, promise to provide more detailed accounts in the future.

²⁴ On the SDGs see chapter 9 in this volume.

²⁵ The UN Office for Drug and Crime Control maintains a homicide database and publishes the Global Studies on Homicide reports, which only covers illegal forms of killing, thereby excluding most of the deaths in wars. The latter data primarily comes from the WHO. See Geneva Declaration (note 11). On other methods of data collection on violent death see Spagat, M., ‘Estimating the human costs of war: The sample survey approach’, eds M. R. Garfinkel and S. Skaperdas, *The Oxford Handbook of the Economics of Peace and Conflict* (Oxford University Press: Oxford, 2012), pp. 318–41, and section V in this chapter.

be more comprehensively linked to other information, for instance on the mobilization of groups, their choice of methods and their external support.²⁶ Research on social movements, such as the Politics of Contention programme, provides some insight into the study of political processes, which might help to identify avenues for useful additional data sets.²⁷ This would also provide a bridge to the study of non-violent political action, a field of research which has rightly received much recent attention.²⁸

Conclusions

In recent years there has been a major growth in the availability and validity of data sets on various forms of violence. This has been accompanied by political and academic interest in such data, in order to better understand and help to reduce the burden of violence. However, as the widely diverging data on the number of victims in various armed conflicts illustrates, the collection of data on violence continues to be fraught with problems.²⁹ The increased heterogeneity of data on violence is both an opportunity for more detailed assessment and a challenge to the consolidation of knowledge on what is happening on the ground. As the lack of indicators for SDG 16 demonstrates, continued efforts are needed to identify how best to capture through data the causes, costs and consequences of violence.

²⁶ The UCDP recently added a data set on external support. On external support in armed conflict see chapter 4 in this volume.

²⁷ Tilly, C., *Contentious Performances* (Cambridge University Press: Cambridge, 2007).

²⁸ Chenoweth, E. and Cunningham, K. G., 'Understanding nonviolent resistance: An introduction', *Journal of Peace Research*, vol. 50, no. 3 (May 2013), pp. 271–76. Chenoweth and colleagues at the University of Denver have launched the Major Episodes of Contention data set, funded by the US Central Intelligence Agency as part of the Political Instability Task Force, which is supposed to cover violent and 'mixed' events.

²⁹ See section V in this chapter.