4

ATROCITY CRIMES AS A DISEASE

A statistical approach to early detection[1]

Birger Heldt

1. Introduction

One conservative estimate puts the number of civilian fatalities in intra-state armed conflicts over the period 1989–2015 at around 828,000, which includes the 1994 genocide of Tutsis in Rwanda, estimated conservatively at 500,000 fatalities.[2] Another estimated 305,000 fatalities were recorded as of "unknown status" in terms of being civilians, rebels, or government personnel. The same source conservatively estimates the number of deliberate killings of civilians – as opposed to collateral damage fatalities – in armed conflicts as close to 772,000, including the Rwanda genocide but also interstate conflicts that are linked to fewer than 5,000 civilian fatalities. Removing the 500,000 victims of the Rwanda Genocide, as well as the civilian fatalities linked to interstate conflicts, leaves approximately 323,000 civilian fatalities in intrastate conflicts, of which 267,000 were intentionally killed by rebels (181,000, or 69%) and by governments (82,000, or 31%).

Not counting the 1994 genocide in Rwanda, these estimates show that civilian killings in the context of civil wars are more than 80% intentional killings instead of "collateral damage", and that rebel groups were responsible for around two-thirds of the intentional killings. Whereas genocidal killings are intentional as well as systemic, intentional killings of civilians are not necessarily systematic. The latter occur when civilians are killed not as a consequence of a policy or a strategy, but foremost because of lack of discipline and decisions of individual combatants. In line with common vocabulary (e.g., Verdeja 2012), instances of genocides and intentional killings of civilians – whether systematic or not – are labeled "atrocity crimes" throughout this chapter.

Risk models and early warning models are bricks in the international atrocity crime prevention architecture. This chapter proposes an additional brick to that architecture. As a complement to risk models and early warning models, and sharing basic features with standard approaches and concepts in epidemiology and early detection mechanisms for disease outbreaks, this chapter outlines and applies in an illustrative manner a disease-oriented approach to the *early detection* of *systematic* atrocity crimes. The approach involves detecting whether civilian fatalities are unsystematic, in that they are randomly distributed across time or space – and thus statistically independent events though still possibly intentional – or whether civilian fatalities are non-random and interdependent events, and thus likely systematic.

The early detection approach outlined in this chapter can serve two functions. First, and as the name suggests, it is forward-looking or prospective in that it can identify systematic atrocity crimes in their early phases, and thus constitutes a de facto early warning mechanism, not only for genocides, but also for systematic atrocity crimes that are not genocidal in intent. Second, it may also be applied in a backward-looking or retrospective manner in terms of forensic analysis of past civilian fatalities to establish the extent to which the fatalities were not just intentional but also potentially systematic. In the latter capacity, it can complement other pieces of information – such as witness accounts or documentary evidence – used to assess the intentionality of killings of civilians.

Section 2 summarizes the current state of risk assessment and early warning of atrocity crimes. Section 3 briefly describes the basic approach to early detection as practiced by the disease surveillance community. The basic elements and principles of the suggested data analysis are presented in non-technical manner in Section 4, whereas Section 5 illustrates the approach by applying it to an emerging case of genocide and atrocity crime: Darfur in 2003–2004. The final and concluding section discusses the need to analyze extensions and complications of employing statistical approaches, and highlights the need for a thorough analysis of key concepts that may assist in more precisely specifying what to look for in data. This may assist in developing criteria for the statistical detection of atrocity crimes.

2. Risk assessments and early warning of atrocity crimes

Risk models of atrocity crimes – whether systematic or not – address the question of *where* outbreaks may take place, and focus on often slow-moving structural risk factors, such as regime type and level of economic development. In contrast, early warning models of atrocity crimes address the question of *when* atrocity crimes may break out, and focus on immediate or proximate causes in terms of fast-moving factors or triggers.[3] Risk models and early warning models are ideally used in a complementary manner, as they serve different functions. First, risk models are used to identify countries at risk that may then be placed on a "watch list". Second, early warning analysis homes in on when an outbreak may be imminent in the countries at risk.[4]

After the pathbreaking statistical study by Harff (2003), and to varying degrees related to Harff's statistical model, the number of published cross-national statistical studies attempting to predict the onset, occurrence, duration, or magnitude of genocides and politicides has grown, and this applies also to the number of statistical studies attempting to predict the occurrence and magnitude of mass killings (with or without genocidal intent, and whether systematic or not).[5] These studies constitute de facto risk assessments, as they focus almost exclusively on slow-moving factors.[6] There are meanwhile as yet no statistical studies that focus on triggers to predict the onset of genocides and mass killings, i.e., de facto early warning. A possible reason for this asymmetry is that historical data on slow-moving risk factors are easier to collect than data on fast-moving events and triggers, which in turn hampers the development of statistical risk models[7].

In contrast to the mentioned retrospective statistical studies, three projects generate public country-by-country risk forecasts on the basis of statistical models that have been validated by historical data. Barbara Harff and Ted Robert Gurr provide an annually updated list of a subset of countries predicted to be most at risk for genocides and politicides, whereas the Atrocity Forecasting Project at the University of Sydney provides countrywide multi-year forecasts.[8] Another forecasting initiative is the Early Warning Project at the U.S. Holocaust Memorial Museum that provides global risk assessments for all countries of the world and covers the broader category of mass atrocities, whether genocidal or not.[9]

One difference among these three projects is that, whereas the two former ones provide forecasts solely on the basis of statistical models trained on historical data, the latter approach complements forecasts from a statistical model with input from an expert opinion pool. Another difference is that the Atrocity Forecasting Project includes not only slow-moving risk factors, but also fast-moving potential triggers, such as election periods and political assassinations. As such, it is a risk assessment–early warning hybrid forecast model, and thus serves to illustrate Verdeja's (2016) observation of no "hard line" between risk assessment and early warning of atrocity crimes. These and other difference aside, as the three projects rely on broadly similar explanatory models, despite differences in how data are analyzed, they generate risk lists that overlap to a considerable extent, even though the relative risk rankings of individual countries may differ.

As outlined by Harff (2009), it is still the case in 2017 that, in contrast to these risk forecasts, purely statistical genocide/politicide early warning forecasts based on empirically validated statistical models, reliant on triggers/events alone, and providing numerical and replicable early warning assessments do not exist. The core reason is the absence of empirically validated statistical models on the subject matter. As a consequence, the genocide prevention community is currently better equipped for forecasting *where* rather than *when* state-led genocides, politicides, or mass killings may break out. Early warning is instead currently a practice based on analytical frameworks and expert judgements. For instance, the *Framework of Analysis for Atrocity Crimes* of the United Nations Office of the Special Advisor on the Prevention of Genocide Crimes lists 143 underlying and proximate factors to assess the risk of genocide, crimes against humanity, and war crimes.[10] In practice, illustrating Verdeja's (2016) observation of no "hard line" between risk assessment and early warning of atrocity crimes, the *Framework* combines fast-moving early warning factors with slow-moving risk factors, and appears to cover not just states but also nonstate actors as potential perpetrators. It meanwhile poses a daunting task for any analyst, as it requires the simultaneous assessments of 143 factors, or moving parts.

Current risk assessment models and forecasting projects, as well as early warning analysis of atrocity crimes, display some limitations. First, only a couple of statistical studies have assessed the risk of mass killings by rebel groups and other nonstate actors, despite the fact that targeting of civilians by rebel groups is common, as noted in the introduction to this chapter. Second, as pointed out by Ulfelder (2012), since genocides and politicides are – thankfully – relatively rare events, the empirical basis for developing accurate models of genocide onset is constrained: there are relatively few cases from which we can learn with confidence, and on which we can "train" statistical risk and early warning models. That is, the fewer the cases, the less the data, and by extension the less certain the conclusions from statistical studies. This observation raises, in turn, the question of how much more accurate statistical genocide/politicide risk assessment models and forecasts – and future genocide early warning models and forecasts – can become: even if methods, theories, and models become more refined, and the trigger data challenge for early warning models is addressed, the inherent challenge of predicting the onset of rare events will remain.

The larger number of instances of atrocity crimes – whether systematic or not – that are not genocidal or politicidal in intent means that predictive models that focus on this sub-category should in theory be able to attain a higher predictive accuracy than models that focus on the less-common genocides and politicides. However, whereas genocides and politicides may conceptually be regarded as processes with a distinct and discernible onset, the same does not apply to non-genocidal/-politicidal atrocity crimes. For instance, does 100, 1,000, or 10,000 civilians intentionally killed indicate that a larger amount of atrocity crime is to follow, and within what time period? This suggests that statistical models that focus on the onset of non-genocidal/-politicidal atrocity crimes face conceptual challenges.

Third, analytical early warning frameworks are labor-intensive as they contain many moving parts. Moreover, since they rely on expert judgements, they are also not systematized and replicable like the three aforementioned risk assessments, and they do not generate numerically precise, replicable, and empirically validated early warning forecasts. As such, they do not live up to standard early warning model criteria in terms of providing "timely, accurate, valid, reliable and verifiable" warnings (Verdeja 2016: 14).

This state of affairs raises the question of whether there exists a complementary approach to current early warning approaches that (1) sidesteps data challenges as well as the time-consuming and costly process of developing empirically validated early warning models, (2) avoids the labor intensiveness, non-replicability, and accuracy issues of analytical frameworks, (3) can cover state and non-state actors as perpetrators, (4) focuses on *systematic* atrocity crimes, (5) is simple and relies on standardized and well-known statistical diagnostics that do not require extensive statistical expertise, and (6) generates timely, replicable, and precise numerical scores.

One approach that meets these requirements focuses on early detection. In practice, the approach involves reconceptualizing atrocity crimes as a disease, and adapting standard tools from the disease surveillance community to the particularities of *systematic* atrocity crimes. Moreover, in contrast to possible future early warning models that will attempt to statistically forecast the timing of outbreaks by analyzing a series of key trigger variables, and early warning frameworks that may contain more than 100 variables, the systematic atrocity crime detection approach to be presented turns previous early warning analysis on its head by not including any predictor variables, and by using only a single moving part or variable: civilian fatality data.

3. Early detection in epidemiology

The complexity of the early detection of disease outbreaks is indicated by the multitude of data analysis techniques that have been developed and refined over the past 20 years (see Unkel et al. 2012). While employing different techniques, the shared goal is to detect statistically unusual or conspicuous clusters of disease cases in time or in space. As expressed by Wong et al. (2005), "The basic question asked by all detection systems is whether anything strange has occurred in recent events". By "strange" is meant that event rates are statistically unusual or unlikely in that they are beyond a certain threshold, whether in time, space, or both. Related to this is the concept of "clusters" that refers to "aggregations of relatively uncommon events or diseases in space and/or time in amounts that are believed or perceived to be greater than could be expected by chance" (Porta 2014). Similar to the early warning criteria cited by Verdeja (2016), disease surveillance systems are typically judged on a number of criteria, including sensitivity (proportion of accurate predictions of "onset", or true positives), specificity (proportion of accurate predictions of "no onset", or true negatives), timeliness, validity, and simplicity (Center for Disease Control 2007).

Automated disease surveillance systems depart from a baseline, continuously add data, and strive to detect whether the added data deviate too much from the baseline – that is, whether "anything strange" has occurred. Epidemiologists rely on a historically established normal incidence rate as the baseline, against which observed incidence rates are statistically assessed as normal or deviant, given some pre-established threshold beyond the baseline. An incidence rate over a certain period of time that exceeds a certain threshold *may* indicate that an outbreak is imminent. Some diseases have non-stationary rates that vary across time, and this means that the baseline varies across time. For instance, the normal or baseline rate of influenza infections varies across years. The wide applicability of this general approach for identifying anomalies is demonstrated by its application to a series of other non-disease related phenomena, including early detection of terrorism wave outbreaks and clustering of car crashes (Gao et al. 2013; Sparks , Okugami, and Bolt 2012). Since so-called count data are studied, reliance on the Poisson distribution is at the core of disease surveillance systems.

Typically, surveillance systems generate an "alarm" when observed rates deviate too much from the baseline, after which experts assess whether the alarm may be due to data and/or model errors (false alarm), whether the observed deviation may have natural explanations such as seasonality of data (false alarm), or whether a disease outbreak may be imminent (true alarm) (see, for example, Hulth et al. 2010). Surveillance system alarms are thus not accepted at face value, and are only acted upon after further expert assessments. Thus, the early warning systems are in practice regarded as, and used as a decision support for, experts. It may in this connection be noted that the Early Warning Project at the U.S. Holocaust Memorial Museum adheres to this principle in that it applies expert assessments to the country-specific statistical risk forecasts before the final risk assessment is issued. Meanwhile, expert assessments of the forecasts generated by Harff and Gurr, and the Atrocity Forecasting Project, are in practice left to potential end users of the forecasts.

Whereas epidemiology early detection approaches focus on deviations from a normal incidence rate, such a volume- or rate-centered approach is difficult to apply to the early detection of *systematic* atrocity crimes in civil conflicts. First, there are no "normal" or "natural" rates or volumes of civilian fatalities that can be used as baselines. Second, the goal is to detect outbreaks (forecasting) or the past occurrence (forensic analysis) of *systematic* atrocity crimes, not whether a systematic atrocity crime is ahistorically large, or whether atrocity crimes in general are taking place.

As will be developed in the next section of this chapter, instead of focusing on volumes of killings, an alternative early detection approach focuses on the pattern of killings. Such an approach has a number of benefits. First, it requires only minute amounts of data. Second, it is easy to apply, and the numerical findings are easily replicable. Third, it does not require an understanding of the causes or drivers of genocides and other atrocity crimes. Hence, there is no need for a costly and time-consuming development of causal models. Fourth, since the focus is not on the volume of violence but its distribution, systematic atrocity crimes can be detected in their early stages (forecasts) when data and other circumstantial evidence – such as documents or witness accounts – are limited and the violence is at a low level. Early detection can thus serve an early warning function. Fifth, it only requires one piece of data – fatality data – and there is no need for "big data". Finally, the approach is related to tried and tested approaches within epidemiology and disease outbreak surveillance systems. The latter means also that the further development of the approach may gain traction from already-existing disease surveillance detection practices and concepts.

4. Randomness in count data

The distribution of count data, such as the number of persons killed in war for a given temporal and/or spatial unit, may be in line with the Poisson distribution in which the variance is in theory equal to the mean (King 1989a, 1989b). This distribution is based on two core assumptions (*ibid.*): independence, in that events are independent of one another, and homogeneity, in that the rate of occurrence of an event is in theory constant across time. Translated to violence against civilians, the former assumption means that civilian fatality events are independent of one another across time or space, in that they are isolated – but still possibly common – events. That is, the violence is not systematic. The latter assumption means that the underlying risk of civilian fatality events is constant across time or space, and means in practice that there is no so-called seasonality in the data.

The distribution of events may in reality be characterized by so-called under-dispersion or over-dispersion, which indicate either non-randomness in terms of events not occurring independently of one another other (*ibid.*), or seasonality in the data. The latter means that the assumption of an underlying constant probability of an event across time is violated. Under-dispersion means that the variance of the data is smaller than the mean, resulting in a pointier probability distribution; over-dispersion means that the variance is larger than the mean, resulting in a flatter probability distribution. In turn, over-dispersion may indicate positive contagion; that is, violence begets violence, in that violence occurs in clusters and is somehow linked rather than consisting of isolated events (*ibid.*; Richardson 1944). Meanwhile, under-dispersion may indicate negative contagion (e.g., violence begets non-violence) (*ibid.*). Again, these interpretations apply provided that the homogeneity assumption holds, in that there is no material seasonality in the data.

Incidentally, perhaps the earliest scientific study of war relied on the Poisson distribution, in this case to examine whether war onset was a random process (Levy 1982, Mansfield 1988, Richardson 1944). Richardson (1944) reported that the observed distribution of war onset from the year 1500 to 1931 did not deviate from the predictions of the Poisson distribution, thereby indicating that war onset during one time period was not contingent on war onset during the previous time period, but instead random: war onsets did not show clustering across time. Thus, a war diffusion process was not evident. Richardson's finding does not mean that war onset cannot be predicted, but only that onsets of individual wars were overall independent of one another: they were isolated events.

A study often cited in statistics textbooks or university statistics courses as a classical case and illustration of an application of the Poisson distribution is Clarke's (1946) famous single-page study that examined whether German V1 and V2 flying bombs fell in clusters over London, and thus hit London in a non-random manner.[11] Clarke's study focused on south London, which he divided into 576 squares measuring 500 meters. Clarke reported that 299 squares were not hit at all, that 211 squares were hit once, 93 squares were hit twice, 35 squares were hit three times, 7 squares were hit 4 times, and 1 square was hit 5 or more times. As it turned out, the observed number of hits per square was perfectly in line with the number predicted by the Poisson distribution.[12] As the observed numbers did not deviate from the Poisson-predicted numbers, the findings meant that the bomb hits were randomly distributed across the 579 squares, and did not fall in non-random clusters. Clarke's original table is reproduced as Table 1 below.

5. Early detection of systematic atrocity crimes in civil conflicts

As suggested by numerous scholars, and as follows from the definition of genocide in terms of being intentional, *systematic*, coordinated, and sustained, genocides should reveal themselves in the patterns of killings. To borrow from, among others,

TABLE 4.1 Distributions of German V1 and V2 bombs hitting south London

No. of flying bombs per square	*Expected no. of squares (Poisson)*	*Actual no. of squares*
0	226.74	229
1	211.39	211
2	98.54	93
3	30.62	35
4	7.14	7
5 and over	1.57	1
Sum	576	576

Source: Clarke (1946)

Rosenberg (2012), genocide is a process, not an event. Consistent with this conceptualization, Park (2011) discusses four direct and circumstantial pieces of evidence that can be applied on any of the five dimensions of the Genocide Convention to assess intent: statements indicating genocidal intent; the scale of the atrocities committed; *systematic* targeting of the protected group; and evidence suggesting that the acts were consciously planned. Related to this, Verdeja (2012) focuses on circumstantial evidence, and divides behavior into three dimensions that lend themselves to further analysis: level of lethality; degree of coordination (how *systematic*, coordinated, and sustained does lethal violence appear to be, such as the use of similar destructive tactics in a wide area); and scope (the extent to which coordinated lethal violence was applied against all or a substantial part of a victim group).[13]

Rwanda 1993 constitutes an illustrative example of the approach of assessing whether violence against Tutsis was systematic. The genocide in 1994 was preceded by violence against Tutsi civilians during 1991–1993. The exhibition at the Genocide Memorial Museum in Kigali describes it as a government dress rehearsal, as well as an attempt to assess the reaction of the international community to a planned mass killing of Tutsi civilians. A UN inquiry into human rights violations was launched and issued its report in April 1993 (United Nations 1993). According to team leader Mr. Bacre Waly Ndiaye, UN Special Rapporteur on the Commission on Human Rights, the team looked into the location and timing of violence, and found among other things that violence was located along major roads, and often preceded by local electricity blackouts.[14] Based on data covering October 1990 until December 1992, the team concluded that this pattern of seemingly interdependent events meant that the violence was not random, but systematic and co-ordinated.

The above insights on the systematic character of genocidal violence applies to the intentional killings of civilians that may not be genocidal or politicidal in intent but are still *systematic* instead of a consequence of random acts by government soldiers or rebels: if killings are systematic, then the pattern should be different as compared to when the killings are not systematic or random, though still potentially intentional. Speaking more broadly of systematic atrocity crimes instead of only genocides and politicides, and thinking in statistical terms: if they are systematic rather than essentially random – whether intentional or not – civilian killings, then the distribution of civilians killed across time and/or space should not evince randomness. This means in turn that the distribution – rather than the volume – of fatalities becomes of interest for assessments of whether systematic atrocity crimes have taken place.

The goal then becomes to detect whether killings are non-random and thus form a pattern or process of interdependent or clustered fatalities, rather than random (that is, non-organized) isolated fatalities that would be the case if decisions to kill civilians are taken by individuals without any co-ordination with one another. Similar to the approaches used by, e.g., Clarke (1946), Richardson (1944), and, for many years, in standard diagnostic tests of different types of count data, a basic formal test of whether fatalities are non-random is whether the distribution of fatalities across time and/or space deviates from the Poisson distribution in terms of showing over-dispersion. Over-dispersion indicates in turn that fatalities are not isolated occurrences, but instead interdependent and non-random. This means that a certain probability distribution (Poisson) rather than historical volumes of fatalities is used as baseline. A routine statistical package diagnostic procedure to assess whether an observed count data distribution adheres to the Poisson distribution involves a Pearson Chi-Square goodness-of-fit test that compares expected (Poisson) rates with rates across all observations, and assesses whether the fit is "good": if the observed data deviate too much from the Poisson predicted data, then it may be concluded that observed data are not randomly distributed.

For illustrative purposes, this early detection data analysis approach is applied in a forensic manner on the genocide in Darfur. The analysis is confined to July 2003 until December 2004, when evidence of a systematic character of atrocities against civilians was mounting but still contested in some quarters. The purpose of the application is to illustrate the early detection approach and demonstrate insights that can be gained from analyzing only fatality data. The purpose is not to enter into a debate on fatality figures.[15] Whereas the application is retrospective and thus forensic in nature, it could be used in a prospective manner by applying it to real-time data from ongoing conflicts to provide for real-time early detection and, in effect, early warning of systematic atrocity crimes. For instance, analysis could be carried out on continuously updated weekly, bi-weekly, or monthly data.

In June and July 2004, the United States carried out an investigation of the violence in Darfur through the Atrocities Documentation Project (ADP), the purpose of which was to assess whether genocide had been and/or continued to be carried out in Darfur (Totten 2006). During July and August 2004, the project collected interview data that covered the period from late 2003 until August 2004. In September 2004, and "based on a consistent and widespread pattern of atrocities (killings, rapes, the burning of villages) committed by the Janjaweed and government forces against non-Arab villagers", the United States Secretary of State Colin Powell declared that genocide had taken place in Darfur (*ibid.*). Similarly, in late 2004, the UN launched a Commission to "investigate reports of violations of human rights law and international humanitarian law" in Darfur, with a report issued in January 2005. But in contrast to the declaration of Secretary of State Colin Powell, and while the Commission reported that serious human rights violations had taken place and was "alarmed" by "attacks on villages, killing of civilians, rape, pillaging and forced displacement", it did not consider that a genocide had taken place:

> [...] the crucial element of genocidal intent appears to be missing, at least as far as the central Government authorities are concerned. Generally speaking the policy of attacking, killing and forcibly displacing members of some tribes does not evince a specific intent to annihilate, in whole or in part, a group distinguished on racial, ethnic, national or religious grounds. Rather, it would seem that those who planned and organized attacks on villages pursued the intent to drive the victims from their homes, primarily for purposes of counter-insurgency warfare.
>
> *(United Nations 2005: 3)*

The question is whether the outlined statistical early detection approach for systematic atrocity crimes generates findings that corroborate the ADT's conclusions, or the UN report's conclusions. In this particular case, the outlined approach for assessing data could have been used as part of the UN's as well as the ADT's effort to analyze fatality data. For the analysis, the Darfur civilian fatality data can be arranged spatially with regard to some geographical grid system, or in accordance with, e.g., "natural" geographic areas.[16] In this case, civilian fatality data are arranged temporally, across time, and the analysis examines fatality patterns for the entire Darfur region.

The analysis covers a long period, 18 months, which entails risks of variations in fatalities as caused by seasonality – here with regard to effects of weather patterns in terms of the rainy season from June to September that may reduce the mobility of government troops and the Janjaweed in Darfur. Seasonality may in turn undermine the analysis and provide for faulty conclusions, as it would violate the theoretical assumption of an underlying constant probability of violent events across time. Hence, any observed over-dispersion in the Darfur fatality data could be due to systematic violence, but could also in whole or in part be due to seasonality in the data. Another factor that may possibly influence the fatality data is the ceasefire agreement of April 2004 and the deployment of the African Union Mission to Sudan (AMIS) in July 2004, which by October consisted of 465 personnel and by January 2005 of around 1,300 personnel (Human Rights Watch 2006). While this long period is selected to align broadly the analysis with the period covered by the UN report and the ADP project, the issue of potential seasonality will be returned to in the data analysis.

Over the period of July 2003–December 2004, the number of civilians in Darfur killed by the government or the Janjaweed is estimated at around 4,696.[17] For the analysis, data are arranged in half-month periods, generating 36 half-monthly observations, which entails an average of 130.4 civilian fatalities per half-month period.[18] In the graph in Figure 4.1, it is difficult to discern any obvious within-year seasonality in the data as being due to the rainy season. In 2003, civilian fatalities are high in July and August, which is the larger part of the rainy season, and the number of civilian fatalities in September (last month of the rainy season) are not materially different from the number of fatalities in October or November. In 2004, fatalities are actually higher during the rainy season than during many other months of the year. To the extent that seasonality exists in the data, it is difficult to link it to the rainy season. Finally, fatality levels for 2004 are lower than in 2003, but it is difficult to discern any straightforward link to the deployment as well as enlargement of AMIS. This issue of potential seasonality will be returned to in the analysis.

A visual inspection of the graph in Figure 4.1 does not yield any obvious conclusions as to whether civilian killings are systematic or random, even though the fatality figures are high. To compare whether expected (Poisson) – or baseline – and observed fatality counts are different, and thus whether civilian fatalities may form a systematic pattern, the data are categorized into segments of 25 persons killed per half-month period. Table 4.2 includes the observed distribution across these segments together with the random distribution predicted by the Poisson distribution.

It is clear from data in the table, which is also illustrated in Figure 4.2, that the observed pattern is inconsistent with the Poisson predicted (random) pattern or baseline. The fit between the observed and predicted distributions is not good. Moreover, the variance of the observed distribution is much larger than expected, which means in this case that extreme over-dispersion is evident. Given a half-monthly average of 130.4 civilian fatalities, the Poisson distribution predicts that virtually all half-monthly civilian fatality rates should fall between 100 and 150: if events are isolated and thus independent of one another, and given that the underlying risk of civilian fatalities is constant across

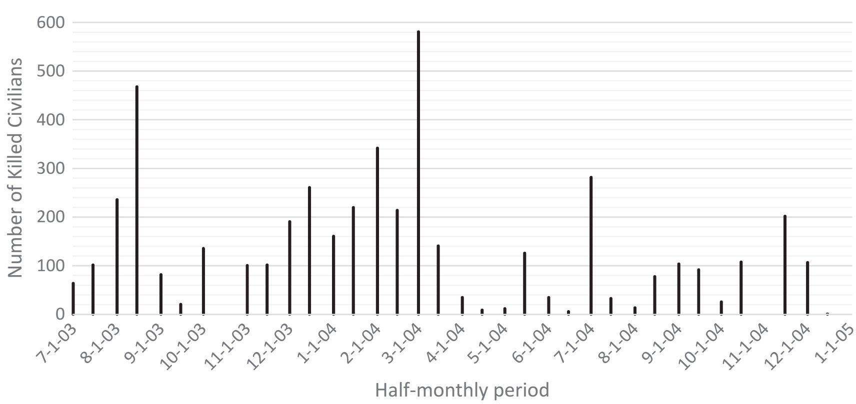

FIGURE 4.1 Civilian conflict-related fatalities in Darfur, July 2003–December 2004

TABLE 4.2 Predicted versus observed number of half-monthly periods for a certain number of fatalities, Darfur July 2003–December 2004

Half-monthly number of civilian fatalities	*Predicted (Poisson) number of half-monthly periods*	*Observed number of half-monthly periods*
0–24	0,0	8
25–49	0,0	4
50–74	0,0	1
75–99	0,1	3
100–124	10,9	6
125–149	23,2	3
150–174	1,8	1
175–199	0,0	1
200–224	0,0	3
225–249	0,0	1
250–274	0,0	1
275–299	0,0	1
300–324	0,0	0
325–349	0,0	1
350–374	0,0	0
375–399	0,0	0
400–	0,0	2

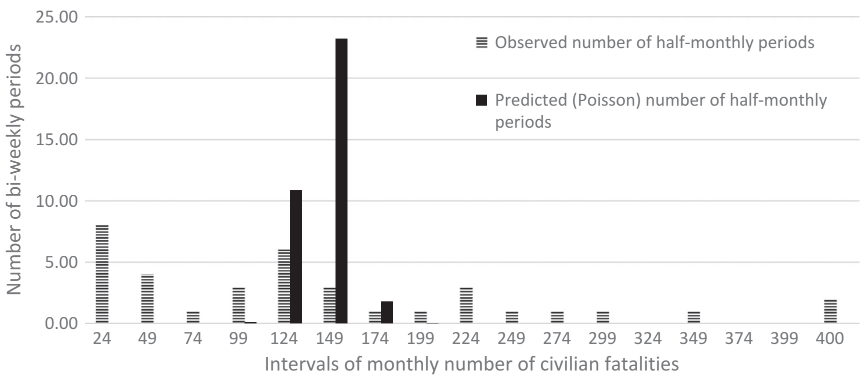

FIGURE 4.2 Predicted versus observed number of half-monthly periods for a certain number of civilian fatalities, Darfur July 2003–December 2004

time, there should ideally not exist half-monthly periods with lower fatality rates, and only one period with a higher fatality rate. However, the observed pattern is very different: there are nine half-monthly periods with the expected number of fatalities, and the remaining periods are distributed across the entire X-axis.

Applying the standard Pearson Chi-square goodness-of-fit test to assess whether the observed data are Poisson-distributed – that is, whether the observed data fit the predicted data – generates a probability of less than 1 in 1,000,000 that the observed data are drawn from the Poisson distribution. This means that the observed fatality data are not randomly distributed, and, in turn, claims that the civilian fatalities were overall random and isolated instances instead of overall systematic can be rejected with a high degree of certainty. This conclusion applies given that the underlying risk of civilian fatalities is constant across time – that is, there is no material underlying seasonality in the data.

In this case, seasonality may be at hand because of at least two conspicuous factors: the rain period, and the AMIS operation. To reduce the risk for seasonality to undermine inferences, the length of the analysed period was shortened to include only 2004, but the results were the same. The analysis was also rerun after aggregating the data into 18 monthly observations, but the result was the same. Finally, the analysed period was limited to after the 2003 rain period, and up to the ceasefire agreement in April 2004, but the findings were the same: the observed data was heavily over-dispersed, and it deviated from the Poisson predicted observations.

While these results do not prove intent to commit genocide, they constitute strong circumstantial evidence that refutes the UN report's claim that violence against civilians was caused "*primarily*" *by* "*counter-insurgency warfare*" and thus did not consist of systematic or inter-connected events. Had the UN report's conclusions been valid, then the observed data of fatalities across time would have been similar to the data predicted by the Poisson distribution. The results meanwhile are supportive of – but strictly speaking do not prove – the United States Secretary of State Colin Powell's claim that a genocide had taken place, although they show beyond reasonable doubt that systematic atrocity crimes had indeed been carried out.

6. Conclusions and the way forward

The outlined approach that focuses on *whether* instead of *where* and *when* involves a narrow and simple task, yet one that is complementary to current early warning and risk assessment methods. This illustrative application may be seen as a first step towards assessing strengths and limitations of the approach that may serve retrospective atrocity crime forensic purposes, as well as prospective atrocity crime early detection purposes. An analysis of data patterns is useful in conjunction with other pieces of direct and indirect evidence (statements, orders, surveys, witness accounts, anecdotes, intelligence, etc.) that in some cases exist in abundance for past as well as ongoing atrocity crimes. The approach can also, for instance, be applied to incident data for attacks of civilians (e.g., destruction of villages), even when the number of civilians killed is unknown. In short, it can be applied to any of the Genocide Convention's dimensions to assess systematic intent, which in conjunction with other pieces of information offers a clearer picture of the character of events. There is thus no inherent need to focus statistical analysis on fatality data alone to unearth systematic features of atrocity crimes. The approach can also be applied to other conflict-related issues, such as to assess whether ceasefire violations in inter- or intra-state conflicts are isolated, accidental, and local events, or whether the events appear to be inter-connected and systematic.

Sometimes, only incident data are available, especially in emerging cases of atrocity crimes. In such emerging cases, this data-driven approach may constitute the only analysis tool available. As stressed earlier, non-randomness of fatality data does not prove intent or genocide, but it does reject claims that patterns of killings are random or accidental as caused by, e.g., crossfire. This, in and by itself, is a useful piece of information, and may in individual cases assist in inducing action by the international community to halt violence, but also to more carefully monitor and collect data on specific cases. Nevertheless, the confounding effects of potential seasonality of data, as well as data quality, need to be carefully considered when carrying out analysis of this type. Hence, as for disease surveillance systems, careful expert assessments of data are crucial to assesses whether and to what extent findings may have been influenced by faulty data, seasonality, or other error sources. That is, the outlined early detection approach should be regarded as a decision support tool.

One feature of the outlined approach is that, in a strict sense, it is not early warning or forecasting, but rather "nowcasting", which means "predicting the present" (Tetlock 2010). Nowcasting is a feature shared with disease surveillance systems that in practice detect outbreaks once they have taken place. However, if systematic atrocity crimes are identified at an early stage, there is still room and time for early action, since historically – with the exception of Rwanda in 1994 – genocides and other systematic atrocity crimes typically develop slowly. Moreover, since the outlined early detection approach builds on actual data and involves nowcasting rather than predictions of atrocity crimes, the findings cannot be dismissed as predictions that may be inaccurate, for the simple reason that the underlying predictive model has insufficient precision and a non-perfect track record.

Another feature is that since the approach does not entail assumptions on triggers or causes – or the *whys* – but only an analysis of fatality data in terms of *whether*, the approach does not offer guidance to the international community on how to prevent, or how to manage, atrocity crimes. A focus on detection sidesteps these important questions. A third feature is that the approach sidesteps the controversial issue of the magnitude of atrocity crimes by examining only one moving part – the pattern of violence – and this means in turn that only very limited amounts of data are required for meaningful analysis. Yet another feature is that the approach is easy to apply, consistent, replicable, and assists in detecting patterns that escape the human eye. As such, the approach appears to be well-aligned with the previously mentioned early warning model criteria in terms of providing "timely, accurate, valid, reliable and verifiable" warnings.

The outlined early detection approach relies on a widely applied and understood statistical distribution and shares features with standard disease surveillance systems. Because of the minimal data requirements, it is a low-cost approach that can in theory be automated in terms of prospective early detection surveillance systems of systematic atrocity crimes at regional or sub-regional levels and rely on real-time and continuously updated fatality data. Such systems could resemble current government-run disease surveillance early warning systems, by adding a layer of expert analysis to assess the veracity of any warnings generated. In that regard, and even though the focus is on fatality distributions instead of fatality volumes, surveillance system templates and lessons learned from the disease surveillance community can be consulted for further technical guidance.

If potential perpetrators are aware of the ability to easily detect systematic patterns of atrocities against civilians, the launch of countrywide early detection systems for countries at risk of genocides or involved in intrastate armed conflicts may serve as a powerful deterrent and restraint against systematic violence against civilians. However, in theory and in a rather far-fetched scenario, a perpetrator of atrocities may then try to be "smart", in that it understands the Poisson distribution and attempts to generate civilian fatalities in a seemingly random pattern across time or space so that the systematic pattern cannot be detected. In practice, such attempts to simulate randomness would require a huge amount of planning, coordination, and communication, which would not only be very difficult or impossible in practice in an organizational context, but also leave other forensic traces behind, such as orders, communication, documents, eyewitness accounts, and so on. It is thus virtually impossible for perpetrators to circumvent a detection system of the outlined character.

Moving from prospective or early detection usage, to retrospective or forensic usage on past instances of atrocity crimes, in some cases there may not exist any other reasonable conclusion from a data analysis than that non-randomness does indeed signal genocide or systematic killings. Whether, as suggested by some scholars, such circumstantial evidence is admissible in court is another question. Yet, given the strength of the outlined method and the few assumptions it entails, there is little reason why a statistical analysis of patterns of killings could not be used in conjunction with other pieces of evidence. Materially, the probabilities and thus indirect evidence obtained from the illustrated analysis of fatality data shares features with probabilities and indirect evidence obtained from DNA analyses of crime scenes, and the latter piece of evidence is admissible in courts.

This chapter has scratched the surface of the complexities of inferring intent and the potential of using detection statistics for such purposes. More research and method development in this direction may be fruitful to identify extensions, limitations, and avenues for further applications of early detection approaches. For instance, potential statistical implications in terms of non-stationary of the underlying risk – as due to, e.g., seasonality in data – for atrocity crimes need to be further considered. However, such concerns can in practice be reduced in forward-looking or prospective analysis if the continuously updated time-periods of interest are short, and also for retrospective or forensic analysis, again if the time-periods are short.

On a final note, the way forward is not confined only to questions of method, as there is also a need for greater clarity of key concepts such as "intent" and "genocide", and how they may manifest themselves in terms of observables. Such an analysis and development of key concepts may assist in more precisely determining what to look for in data patterns, and thus assist in formulating statistical criteria for falsification and rejection of claims of intent and systematic killings. It is likely that a greater interest in data patterns and analysis will spur further development and operationalization of key concepts.

Notes

1 This chapter has benefitted from comments from participants at the Genocide Prevention Advisory Group annual meeting, Israel Academy of Science and Humanities, Jerusalem, 8–10 January, 2017, the editors of this book, and James Finkel.

2 Data are from *UCDP Georeferenced Event Dataset* (GED) Global version 5.0 (2015) and *UCDP One-sided Violence Dataset* version 1.4 (2016). The data cover all countries of the world apart from the ongoing civil war in Syria, and incorporate 128,264 violent episodes of varying length, 749 of which are interstate conflicts. See also Sikkink and Kim (2013) and Kulkarni (2016) for historical data on prosecution of human rights violations, and Fjelde, Hultman, and Sollenberg (2016) for an aggregate descriptive analysis of patterns of violence against civilians in civil wars. Estimates of civilian fatalities are controversial and difficult; for discussion, see Lacina and Gleditsch (2005), Lacina, Gleditsch, and Russett (2006), Spagat et al. (2009), Human Security Report 2010 (2010), Muggah (2011) and Seybolt, Aronson, and Fischoff (2013). As noted by Muggah (2011), there is a distinction between estimates building on incident/event reports, and those building on surveys that are applied as a last resort when incident reports are unavailable. Estimates obtained through the two methods can differ greatly. For a historical overview of efforts and methods to count fatalities in wars, see Jewell, Spagat, and Jewell (2017).

3 This distinction stems from Schmeidl and Jenkins (1998) and Harff (2006, 2009, 2013). See also Butcher et al. (2012). A detailed overview of early warning mechanisms and projects in general can be found in OECD (2009), and an overview of early warning systems within the United Nations can be found in Zenko and Friedman (2011). Neukirch (2015) provides an overview of the early warning system of the Organisation for Security and Co-operation in Europe (OSCE).

4 Harff (2006, 2013), Butcher et al. (2012), and Goldsmith and Butcher (2016). Whereas "genocide" refers to sustained policies and actions intended to destroy groups that share ethnic, religious, or racial traits, "politicide" refers to situations where targeted groups are defined by actual or imagined political beliefs (Harff 2003).

5 See Anderton and Carter (2015) for a comprehensive inventory of cross-national statistical studies and findings. The study also lists, but does not analyze, country or case-specific statistical studies, the findings of which may not be applicable to other countries or cases. Forthcoming cross-national statistical studies focusing on violence against civilians include Raleigh and Choi (2017) and Pospieszna and DeRouen (2017). See also Verdeja (2016), Butcher et al. (2012), and Goldsmith et al. (2013) for overviews of previous statistical studies.

6 Examples of statistical studies that complement slow-moving risk factors with trigger type variables include, *inter alia*, Goldsmith et al. (2013) (variables for political assassination and elections), Wayman and Togo (2010) (variable for military coups), and Wood, Kathman, and Gent (2012) (variable for civilian fatalities incurred by the opposing side in a civil conflict).

7 See Heldt (2012) and Verdeja (2016).

8 See Harff and Gurr (various years) and Goldsmith and Butcher (2016).

9 See Ulfelder (2015) and www.earlywarningproject.com.

10 See Dieng and Welsh (2016) and United Nations (2014).

11 Another study that is often used to illustrate the Poisson distribution was published in 1898 and assessed whether fatalities in the Prussian cavalry caused by horse kicks were

random events. A simple Google search that combines "horse kick", "data", and "Poisson" will point to many instances of statistics courses at universities in which the classical horse kick study data is used to illustrate the Poisson distribution.

12 Like Richardson, Clarke carried out a standard statistical test to ascertain that the observed data distribution differed in a statistically significant way from the expected (Poisson) data distribution.

13 For detailed discussion of the Genocide Convention, see Ambos (2009), Bauer (2009), Human Rights Watch (2010), and Verdeja (2012, 2013).

14 Information provided by Mr. Ndiaye during a dinner conversation in Budapest, September 2010.

15 The application relies on Uppsala Conflict Data Project data that were originally compiled from incident reports, either directly from the reports themselves (e.g., news reports, radio reports), or indirectly from summary reports from, e.g., NGOs. Numerous reports have been published on the case of Darfur that document atrocities, offer first-hand accounts, and provide a large amount of supplemental information and evidence.

16 For instance, the PRIO GRID covers the entire world and divides it into quadratic grids of roughly 50 x 50 kilometres. See Tollefsen, Strand, and Buhaug (2012) and www.prio.no/CSCW/Datasets/PRIO-Grid/.

17 Data are from UCDP GED version 5.0, 1989–2015, and include battle-related and non-battle-related (e.g., murders and massacres) civilian fatalities. Fatality data used in this application include instances where Government of Sudan and Janjaweed were coded as "side A" and Justice and Equality Movement (JEM) or civilians were coded as "side B" in the dataset. Data for killings of civilians may have errors in terms of size, some violent events may be missing, low-level violent events may be under-reported, and the fatality estimates are conservative. Nevertheless, it is unlikely that large (conspicuous) instances of killings of civilians have been missed. Note that these figures do not include fatalities caused by disease, dislocation, starvation, lack of access to health facilities, or even fatalities caused by injuries sometime after an attack or battle, all of which together are many times higher than the direct fatalities analysed in this chapter. For estimates that take these types of fatalities into account, see Reeves (various years).

18 In a few instances violent events stretched across the half-month periods. In these few instances, the number of civilian fatalities were divided equally across the half-month periods in question.

References

Ambos, K. (2009). "What Does 'Intent to Destroy' in Genocide Mean?". *International Review of the Red Cross*, Vol. 91 (no. 876): 833–858.

Anderton, C. H. and J. R. Carter (2015). "A New Look at Weak State Conditions and Genocide Risk". *Peace Economics, Peace Science and Public Policy*, Vol. 21 (no. 1): 1–36.

Bauer, Y. (2009). "Genocide Prevention in Historical Perspective". *Politorbis*, (no. 47): 25–32.

Butcher, C. E., B. E. Goldsmith, D. Semenovich, and A. Sowmya (2012). *Understanding and Forecasting Political Instability and Genocide for Early Warning*. Sydney: University of Sydney.

Center for Disease Control (2007). Updated Guidelines for Evaluating Public Health Surveillance Systems. Available at www.cdc.gov/mmwr/preview/mmwrhtml/rr5013a1.htm.

Clarke, R. D. (1946). "An Application of the Poisson Distribution". *Journal of the Institute of Actuaries*, Vol. 4 (no. 81): 32.

Dieng, A. and J. Welsh (2016). "Assessing the Risk of Atrocity Crimes". *Genocide Studies and Prevention: An International Journal*, Vol. 9 (no. 3): 4–12.

Fjelde, H., L. Hultman, and M. Sollenberg (2016). "Violence Against Civilians in Civil Wars", pp. 42–49 in D. A. Backer, R. Bhavnani, and P. K. Huth (eds.). *Peace and Conflict 2016*. New York: Routledge.

Gao, P., D. Guo, K. Liao, J. Webb, and S. L. Cutter (2013). "Early Detection of Terrorism Outbreaks Using Prospective Space–Time Scan Statistics". *The Professional Geographer*, Vol. 65 (no. 4): 676–691.

Goldsmith, B. E. and C. Butcher (2016). New Forecasts for 2016–2020 & Evaluation of our Forecasts for 2011–2015. University of Sydney, Atrocity Forecasting Project. Available at http://sydney.edu.au/arts/research/atrocity_forecasting/downloads/docs/AFP_2016_2020.pdf.

Goldsmith, B. E., C. Butcher, D. Semenovich, and A. Sowmya (2013). "Forecasting the Onset of Genocide and Politicide: Annual Out-Of-Sample Forecasts on a Global Dataset, 1988–2003". *Journal of Peace Research*, Vol. 50 (no. 4): 437–452.

Harff, B. (2003). "No Lessons Learned from the Holocaust? Assessing Risks of Genocide and Political Mass Murder since 1955". *American Political Science Review*, Vol. 97 (no. 1): 57–73.

Harff, B. (2006). Risk Assessments and Early Warning of Genocide: Some Guidelines for the Office of the Special Adviser to the UN Secretary-General on the Prevention of Genocide. Unpublished manuscript. Available for download at www.un-ausa.org/atf/cf/%7B49C555AC-20C8-4B43-8483-A2D4C1808E4E%7D/Harff%20January%202006.pdf.

Harff, B. (2009). "How to Use Risk Assessment and Early Warning in the Prevention and De-Escalation of Genocide and Other Mass Atrocities". *Global Responsibility to Protect*, Vol. 1 (no. 4): 506–531.

Harff, B. (2013). "Detection: The History and Politics of Early Warning", pp. 85–110 in A. Lupel and E. Verdeja (eds.), *Responding to Genocide: The Politics of International Action*. Boulder, CO: Lynne Rienner.

Harff, B. and T. R. Gurr (various years). Global Risks. Available at www.gpanet.org/content/global-risks.

Heldt, B. (2012). "Mass Atrocities Early Warning Systems: Data Gathering, Data Verification and Other Challenges", pp. 13–32 in *Guiding Principles of the Emerging Architecture Aiming At the Prevention of Genocide, War Crimes, and Crimes Against Humanity*. Genocide Prevention Advisory Network Conference Report.

Hulth, A., N. Andrews, S. Ethelberg, J. Dreesman, D. Faensen, W. van Pelt, and J. Schnitzler (2010). "Practical Usage of Computer-Supported Outbreak Detection in Five European Countries". *Eurosurveillance*, Vol. 15 (no. 36): 1–6.

Human Rights Watch (2010). Genocide, War Crimes and Crimes Against Humanity: A Digest of the Case Law of the International Criminal Tribunal for Rwanda. Available at www.hrw.org/sites/default/files/reports/ictr0110webwcover.pdf.

Human Security Report Project (2010). "Shrinking Costs of War", Part II in *Human Security Report 2009*. Vancouver: HSRP.

Jewell, N., M. Spagat, and B. Jewell (2017). "Accounting for Civilian Casualties: From the Past to the Future". *Social Science History*, (March).

King, G. (1989a). "Event Count Models for International Relations: Generalizations and Applications". *International Studies Quarterly*, Vol. 33 (no. 2): 123–147.

King, G. (1989b). "Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator". *American Journal of Political Science*, Vol. 33 (no. 3): 762–784.

Kulkarni, A. (2016). "Criminal Justice for Conflict-Related Violations – Developments during 2014", pp. 192–209 in D. A. Backer, R. Bhavnani and P. K. Huth (eds.), *Peace and Conflict 2016*. New York: Routledge.

Lacina, B. and N. P. Gleditsch (2005). "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths". *European Journal of Population*, Vol. 21 (no. 2–3): 145–166.

Lacina, B., N. P. Gleditsch, and B. Russett (2006). "The Declining Risk of Death in Battle". *International Studies Quarterly*, Vol. 50 (no. 3): 673–680.

Levy, J. S. (1982). "The Contagion of Great Power War Behavior, 1495–1975". *American Journal of Political Science*, Vol. 26 (no. 3): 562–584.

Mansfield, E. D. (1988). "The Distribution of Wars Over Time". *World Politics*, Vol. 41 (no. 1): 21–51.

Muggah, R. (2011). "Measuring the True Costs of War: Consensus and Controversy". *PLoS Med* Vol. 8 (no. 2): doi:1000417.

Neukirch, C. (2015). "Early Warning and Early Action – Current Developments in OSCE Conflict Prevention Activities" pp. 123–133 in *OSCE Yearbook 2013: Yearbook on the Organization for Security and Co-operation in Europe*. London: Bloomsbury Academic.

OECD (2009). *Preventing Violence, War and State Collapse: The Future of Early Warning and Response*. Paris: OECD.

Park, R. (2011). "Proving Genocidal Intent: International Precedent and ECCC Case 002". *Rutgers Law Review*, Vol. 63 (no. 1): 130–191.

Porta, M. (2014). *A Dictionary of Epidemiology* (6th Edition). Oxford: Oxford University Press.

Pospieszna, P. and K. DeRouen (2017). "Civil War Mediation and Rebel Use of Violence Against Civilians". *Armed Forces & Society*. Online first publication. DOI: doi:10.1177/0095327X16647538.

Raleigh, C. and H. J. Choi (2017). "Conflict Dynamics and Feedback: Explaining Change in Violence against Civilians with Conflicts". *International Interactions*. Online first publication. Doi: doi:10.1080/03050629.2017.1235271.

Reeves, E. (various years). Quantifying Genocide: Dafur Mortality Update. Available at www.sudanreeves.org/.

Richardson, L. F. (1944). "The Distribution of Wars in Time". *Journal of the Royal Statistical Society*, Vol. 107 (no. 3/4): 242–250.

Rosenberg, S. R. (2012). "Genocide Is a Process, Not an Event". *Genocide Studies and Prevention*, Vol. 7 (no. 1): 16–23.

Schmeidl, S. and J. C. Jenkins (1998). "The Early Warning of Humanitarian Disasters". *International Migration Review*, Vol. 32 (no. 2): 471–486.

Seybolt, T. B., J. D. Aronson, and B. Fischoff (eds.) (2013). *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict*. New York: Oxford University Press.

Sikkink, K. and H. J. Kim (2013). "The Justice Cascade: the Origins and Effectiveness of Prosecutions of Human Rights Violations". *Annual Review of Law and Social Science*, Vol. 9: 269–285.

Spagat, M., A. Mack, T. Cooper, and J. Kreutz (2009). "Estimating War Deaths: An Arena of Contestation". *Journal of Conflict Resolution*, Vol. 53 (no. 6): 934–950.

Sparks, R., C. Okugami, and S. Bolt (2012). "Outbreak Detection of Spatio-Temporally Smoothed Crashes". *Open Journal of Safety Science and Technology*, Vol. 2 (no. 3): 98–107.

Strauss, S. (2016). *Fundamentals of Genocide and Mass Atrocity Prevention*. Washington D.C.: United States Holocaust Memorial Museum.

Tetlock, P. E. (2010). "Second Thoughts About Expert Political Judgment: Reply to The Symposium". *Critical Review: A Journal of Politics and Society*, Vol. 22 (no. 4): 467–488.

Tollefsen, A. F., H. Strand, and H. Buhaug (2012). "PRIO-GRID: A Unified Spatial Data Structure". *Journal of Peace Research*, Vol. 49 (no. 2): 363–374.

Totten, S. (2006). "The US Investigation into the Darfur Crisis and the US Government's Determination of Genocide". *Genocide Studies and Prevention: An International Journal*, Vol. 1 (no. 1): 57–77.

UCDP Georeferenced Event Dataset (GED). Global version 5.0 (2015). Department of Peace and Conflict Research, Uppsala University. Available for download at http://ucdp.uu.se/downloads/.

UCDP One-sided Violence Dataset. Version 1.4 (2016). Department of Peace and Conflict Research, Uppsala University. Available for download at http://ucdp.uu.se/downloads/

Ulfelder, J. (2012). Forecasting Onsets of Mass Killing. Paper prepared for presentation at the Annual Northeast Political Methodology Meeting, a.k.a. NEMP, New York University, May 4, 2012.

Ulfelder, J. (2015). *Promising Initial Results from a New Mass-Atrocities Early Warning System*. Unpublished manuscript.

United Nations (1993). Report by Mr. B.W. Ndiaye. Special Rapporteur, on his mission to Rwanda from 8 to 17 April 1993. UN document E/CN.4/1994/7/Add.1.

United Nations (2005). *Report of the International Commission of Inquiry on Darfur to the United Nations Secretary-General*. New York: United Nations.

United Nations (2014). *Framework of Analysis for Atrocity Crimes*. New York: United Nations.

Unkel, S., F. C. Paddy, P. H. Garthwaite, C. Robertson, and N. Andrews (2012). "Statistical Methods for the Prospective Detection of Infectious Disease Outbreaks: A Review". *Journal of the Royal Statistical Society: Series A*, Vol. 175 (no. 1): 49–82.

Verdeja, E. (2012). "The Political Science of Genocide: Outlines of an Emerging Research Agenda". *Perspectives on Politics*, Vol. 10 (no. 2): 307–321.

Verdeja, E. (2013). "Genocide: Debating Definitions" pp. 21–46 in A. Lupel and E. Verdeja (eds.), *Responding to Genocide: The Politics of International Action*. Boulder, CO: Lynne Rienner.

Verdeja, E. (2016). "Predicting Genocide and Mass Atrocities". *Genocide Studies and Prevention: An International Journal*, Vol. 9 (no. 3): 13–32.

Wayman, F. W. and A. Tago (2010). "Explaining the Onset of Mass Killing, 1949–1987". *Journal of Peace Research*, Vol. 47 (no. 1): 3–13.

Wong, W-K., A. Moore, G. Cooper, and M. Wagner (2005). "What's Strange About Recent Events (WSARE): An Algorithm for the Early Detection of Disease Outbreaks". *Journal of Machine Learning Research*, Vol. 6 (Dec.): 1961–1998.

Wood, R., J. Kathman, and S. Gent (2012). "Armed Intervention and Civilian Victimization in Intrastate Conflict". *Journal of Peace Research*, Vol. 49 (no. 5): 647–660.

Zenko, M. and R. R. Friedman (2011). "UN Early Warning for Preventing Conflict". *International Peacekeeping*, Vol. 18 (no. 1): 21–37.