

# Views to a war: systematic differences in media and military reporting of the war in Iraq

Karsten Donnay<sup>1\*</sup> and Vladimir Filimonov<sup>2</sup>

\*Correspondence:

kdonnay@ethz.ch

<sup>1</sup> Department of Humanities, Social and Political Science, Chair of Sociology, Modeling and Simulation, ETH Zürich, Clausiusstrasse 50, Zürich, 8092, Switzerland

Full list of author information is available at the end of the article

## Abstract

The quantitative study of violent conflict and its mechanisms has in recent years greatly benefited from the availability of detailed event data. With a number of highly visible studies both in the natural sciences and in political science using such data to shed light on the complex mechanisms underlying violent conflict, researchers have recently raised issues of systematic (reporting) biases. While many sources of bias are qualitatively known, biases in event data are usually not studied with quantitative methods. In this study we focus on a unique case - the conflict in Iraq - that is covered by two independently collected datasets: Iraq Body Count (IBC) reports of civilian casualties and Significant Action (SIGACT) military data. We systematically identify a number of key quantitative differences between the event reporting in the two datasets and demonstrate that even for subsets where both datasets are most consistent at an aggregate level, the daily time series and timing signatures of events differ significantly. This suggests that at any level of analysis the choice of dataset may substantially affect any inferences drawn, with attendant consequences for a number of recent studies of the conflict in Iraq. We further outline how the insights gained from our analysis of conflict event data have broader implications for studies using similar data on other social processes.

**Keywords:** event data; reporting bias; conflict; Iraq

## 1 Introduction

In recent years the increasing availability of detailed data on conflict events has led to a number of highly visible studies that explore the dynamics of violent conflict [1–4]. Taking a natural science or complex systems perspective, these studies complement a quickly growing quantitative literature in political science that heavily relies on detailed empirical records to systematically study the micro-dynamics of conflict, in particular how individual- or group-level interactions lead to the larger conflict dynamics we observe [5–9].

The conflict event datasets used in these studies primarily draw on media reports and rely to varying degrees on automatic coding as well as the expertise of country or subject experts for coding decisions and quality control [10, 11]. In specific cases - for example in studies focusing on single countries, cities or regions - data may also be based on records collected through Non-Governmental Organizations (NGOs), local newspapers

or researchers' own field work [7, 9, 12]. These conflict event data, however, have been found to be prone to bias [13–16]. Even for otherwise unbiased and flawless research designs this may strongly affect any inferences with regard to conflict dynamics and mechanisms. Data biases do not only arise from variations in data quality and coding across different datasets but also from systematic uncertainties associated with the data collection efforts themselves. Unfortunately, such issues are notoriously hard to identify and difficult to eliminate in the process of data collection, even within institutionalized large-scale collection efforts. Furthermore, identification of potential biases in existing datasets is complicated by the fact that usually not more than one independently generated dataset exists, making it very difficult to infer any biases *post hoc*.

In this study, we focus specifically on a unique empirical case - the conflict in Iraq - that is covered by two independently collected datasets, one of them based on media sources (Iraq Body Count or 'IBC'), the other collected 'on the ground' by the U.S. military (Significant Action or 'SIGACT' data). We use these data to quantitatively test agreement of the event reporting in the two datasets at different temporal resolution and thus systematically identify relative biases. In particular, we find that even for subsets where both datasets are most consistent at an aggregate level the daily time series of events are significantly different. This suggests that whether analyses are based on IBC or SIGACT data may substantially affect the inferences drawn. Our findings are thus highly relevant to a number of recent studies that investigate detailed event dynamics of the war in Iraq using both IBC [2, 3, 17–19] and SIGACT data [8, 20] and contribute to the ongoing debate on issues and implications of data quality in conflict event data.

More broadly, our study speaks to a quickly growing literature that systematically analyzes highly resolved data on social processes. This includes work that uses news media articles to detect international tensions [21] or analyzes Twitter messages to detect mood changes [22]. In fact, much of 'Big Data' derived from artifacts of human interactions corresponds to time-stamped information about social processes. Studies analyzing such data, however, only very rarely consider the potentially substantive biases arising from how they are generated. In fact, these data are subject to much of the same structural limitations as conflict event data (see Section 2.2), with resulting biases that are just as hard to identify and difficult to infer from data *post hoc*. Similarly, inferences based on such data may thus also be substantially affected by the choice of dataset, its characteristics and limitations.

This study is structured as follows. Section 2 introduces the empirical case and the datasets used: IBC data and the U.S. military (SIGACT) dataset made available by *The Guardian*. In Section 3 we systematically compare the reporting of events in both datasets, starting with an aggregate comparison before turning to an in depth analysis of the time series of number of events and event severity. We further analyze the timing signatures in each dataset separately. Section 4 discusses implications of our findings for quantitative analyses of conflict and, more broadly, for studies of social processes that rely on similar data.

## 2 The case of Iraq

The Iraq conflict ranks among the most violent conflicts of the early 21st century and is characterized by excessive violence against civilians with fatality estimates exceeding at least 130,000 by mid-2014 [23].<sup>a</sup> In mid-2003 the conflict began as an insurgency directed

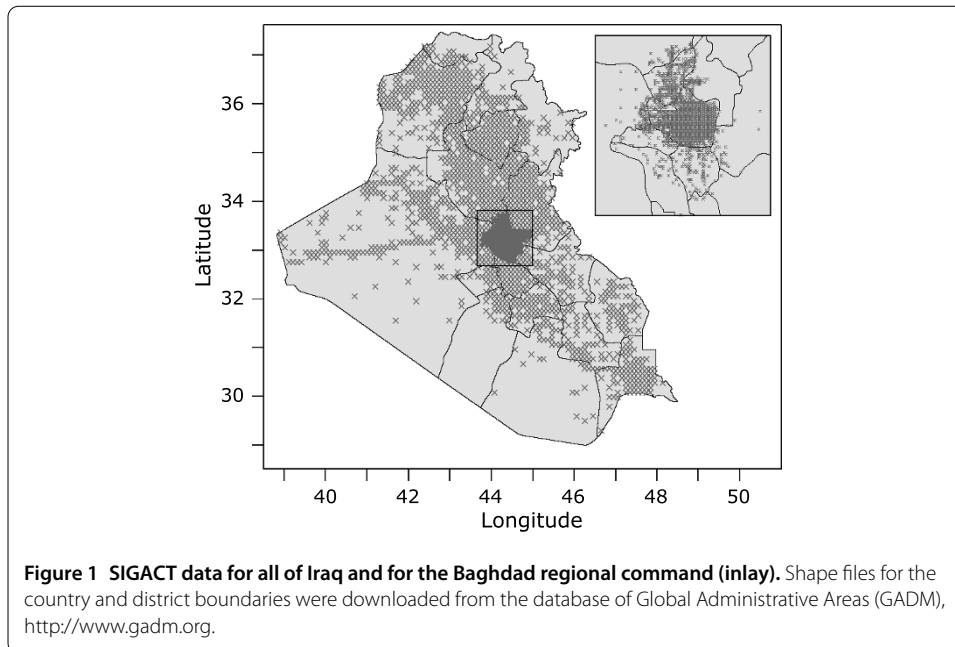
at the U.S. military, its allies and the Iraqi central government. Attacks were initially largely carried out by forces loyal to Saddam Hussein, but by early 2004 radical religious groups and Iraqis opposed to the foreign occupation were responsible for the majority of attacks. The insurgency subsequently intensified throughout 2004 and 2005. Increasingly marked by excessive sectarian violence between the Sunni minority and Shia majority the conflict rapidly escalated in 2006 and 2007. Following the U.S.-led troop 'surge' in 2007, a massive increase of U.S. boots on the ground accompanied by a major shift in counter-insurgency tactics [24–26], the conflict eventually de-escalated significantly throughout 2008. After the U.S. withdrawal from Iraq in 2011 the country continues to experience acts of violence on a (close to) daily basis, both as a result of the continued insurgency against the central government but also increasingly again as a consequence of a renewed escalation of sectarian violence. The recent take-over of the north-western (Sunni) provinces by the Islamic State of Iraq and the Levant (ISIL), an Al-Qaeda affiliate, now even threatens the very existence of a multi-ethnic Iraq.

## 2.1 Data sources

In our analysis we draw on data from the two most commonly used Iraq-specific datasets: Iraq Body Count (IBC), a web-based data collection effort administered by Conflict Casualties Monitor Limited (London) [23], and U.S. military (SIGACT) data available through *The Guardian* [27]. We are very mindful of the sensitivity of the SIGACT data and the debate surrounding their use in academic studies.<sup>b</sup> While this debate continues studies are making use of these data, most notably a recent political science publication on Iraq [8] and an analysis published in the Proceedings of the National Academy of Science (PNAS) using data on Afghanistan [4]. Note that subsets of the SIGACT Iraq data had previously been made accessible to selected researchers and institutions [6, 17, 28] making SIGACT one of the two leading sources of data on the war in Iraq.

The IBC dataset covers violent events resulting in civilian deaths from January 1, 2003 onward until present day and is being updated continuously. We rely here on the publicly available version of the IBC records that does not disaggregate by perpetrator group [23]. The data made available through *The Guardian* contains information on all 'significant actions' (SIGACTs) reported by units of the U.S. military in Iraq that resulted in at least one casualty. The dataset covers the period January 1, 2004 until December 31, 2009 but is missing 2 intervals of 1 month each (from April 30, 2004 to June 1, 2004 and from February 28, 2009 to April 1, 2009) [27]. In order to be consistent in our dataset comparison we have selected our study period as ranging from June 1, 2004 to February 28, 2009 - a period covered by both datasets without any gaps. This period covers the main phases of the conflict described above.<sup>c</sup>

The two datasets differ significantly with regard to the geocoding of conflict events. IBC provides 'human description' of the location (such as 'near Birtilla, east of Mosul' or 'behind al-Faiha'a hospital, central Basra') which implies limited spatial accuracy. In comparison, SIGACT data entries are categorized by U.S. military regional command but more importantly geo-tagged with latitude and longitude coordinates. These coordinates are truncated at a tenth of a degree (about 10 km) for Iraq outside of Baghdad (Figure 1) and at a hundredth of a degree (about 1 km) for the military zone of Baghdad (Figure 1, inlay). The two datasets further differ with regard to their temporal resolution. SIGACT events carry timestamps with a resolution of minutes while IBC events are generally coded to daily precision only. Finally, in contrast to SIGACT data which reports the number of individuals



killed (KIA) and wounded (WIA) for both military actors and civilians, the IBC dataset exclusively covers deadly violence against civilians.<sup>d</sup> In order to compare the two datasets we thus restricted the SIGACT data to entries pertaining to deadly violence directed at civilians. Note that focusing on civilian casualties exclusively rather than including incidents that wounded civilians may, in fact, lead to a biased view of the violence dynamics in Iraq - simply because whether an attack lead to casualties or not may dependent more on chance than intent [29]. To control for this, we performed robustness checks where we additionally included the number of wounded civilians reported in SIGACT; these results are included in Section 3 of Additional file 1.

## 2.2 Structural differences in reporting

There are a number of significant differences between the reporting underlying the IBC and SIGACT datasets that may introduce systematic biases in their respective coverage of violent events. An important source of data bias in geo-referenced event datasets arises directly from the 'spatial' nature of the data, i.e., the location of where a violent event occurs may already strongly influence both its chance of reporting and how it is reported [13, 15]. Such biases may simply be structural, for example, due to the fact that newspapers and their local sources - NGOs, development agencies etc. - often only maintain a constant presence in cities or certain regions of a country. Consequently, reporting likely has a specific *urban or regional bias*, i.e., a more complete coverage of events in those areas compared to others with only limited access [15]. This is often aligned with or equivalent to a *center-periphery* bias since the access and coverage of the media and its sources generally tend to be much lower in remote, peripheral regions compared to the capital or population centers [15]. The same may apply for government or military reporting, simply because administrative infrastructures and a permanent government presence (offices, police and military installations etc.) are often much less developed in the periphery. In volatile states a central government might even effectively not have any control over large parts of the country.

In Iraq the media-based reporting of IBC is quite likely affected by issues arising from limited coverage, especially for locations outside of the main population centers. SIGACT data may also be prone to spatial bias since the U.S. military or coalition forces did not maintain a constant presence everywhere in the country [29]. This limitation, however, should be minimal in a highly patrolled region such as Baghdad. For our quantitative analyses we have thus chosen to focus exclusively on the greater Baghdad area, by far the most violent region during the entire conflict. This choice guarantees that our analysis is not systematically affected by geographic reporting bias since within Baghdad both media-based data and SIGACT's field report-based reporting are least likely to be systematically constrained in their coverage.<sup>e</sup> Focusing on a comparably small and coherent spatial region also avoids the fallacy of studying time series of potentially unrelated or only weakly related incidents that are geographically far apart. The violence dynamics in Kirkuk in the predominantly Kurdish north, for example, are very different from the dynamics in Baghdad. In fact, we contend that since Baghdad was the main locus of violence during the conflict but least prone to geographically biased coverage, it represents the 'best case' scenario for the reporting of violent events in Iraq and any systematic differences in reporting we uncover should also apply to the full datasets.

Notice that even when focusing exclusively on the Baghdad area, IBC's reporting may be prone to additional biases that arise from its reliance on the quality and accuracy of the media coverage. There is ample evidence that newspaper reports of incidents are subject to a number of biases including selective reporting of certain types of events [30, 31], as well as better coverage of types of events that have occurred before and of larger events compared to smaller events [32]. Such size bias should be especially pronounced in situations with a high density of incidents and only limited reporting capacity - in Iraq this would have been most relevant during the escalation of the conflict in 2006-2007. SIGACT data on the other hand is directly based on military reports from the field and should therefore, as long as military presence is high as in the case of Baghdad, cover more incidents regardless of size. Based on these structural differences in the reporting we can therefore expect that:

(I) IBC should cover systematically fewer low casualty events than SIGACT, but also that

(II) Differences in reporting, in particular of events with few casualties, should be greater the more intense the conflict.

Note that (II) also extends beyond mere coverage - i.e., whether an incident is reported at all - to the quality of reporting. The more intense the fighting the less accurately field reports are able to reflect casualty counts, simply because soldiers may not always be able to reliably account for all casualties in such situations [29]. Similarly, media reports may also not always precisely reflect 'true' casualty counts - in fact, IBC explicitly codes for lower and upper bounds of casualty estimates.<sup>f</sup>

In the case of events with larger casualty counts, the reliance of SIGACT on field reports may negatively affect reporting accuracy. One key reason is that longer and intense confrontations involving multiple units may be falsely reported as several separate incidents by each unit instead of being coded as one large episode. This may lead to over-reporting of the number of incidents and under-reporting of the number of casualties per incident. Note further that the categorization of incidents and identification of victims, in particular, may sometimes be ambiguous [29]. In fact, prior quantitative research confirms that the

interest of the observer tends to affect how incidents are reported [33]. Ideological biases in media reporting - such as government-directed negative reporting on the opposition or simply general limitations to press freedom - result in an inaccurate representation of the situation in a country/region and may thus bias how events are reported [15].

In Iraq, we would further generally expect coalition troops' reporting of civilian casualties to be comparably more conservative than the news media. Modern counterinsurgency doctrines emphasize the importance of 'population-centric' warfare, favoring tactics and rules of engagement that minimize collateral civilian casualties [34]. In turn, this implies strong incentives for U.S. troops to keep civilian fatality reports of operations as low as possible. These incentives are strongest for comparably larger incidents with significant unintentional ('collateral') civilian casualties. Note, too, that especially during the escalation of violence in 2006-2007 the conflict in Iraq became highly politicized along the Sunni/Shia divide. This provided strong incentives for newspapers from either side to emphasize the atrocities of the other, i.e., to provide less conservative casualty estimates, especially for large incidents. Overall we can thus expect that

(III) IBC should report comparably more events with many casualties than SIGACT.

Note that in general the timing (and location) of attacks can be expected to be more accurate when derived from field reports compared to IBC, whose coverage is fundamentally constrained here since newspaper articles usually only report approximate times and locations. However, it is also known that SIGACT reporting in Iraq did not adhere to homogenous reporting standards throughout the entire conflict, including the integration of reports (or initial lack thereof) from Iraqi military units [29]. There is also a known issue of field reports being entered with midnight timestamps if the exact reporting time is unknown. These differences should not systematically affect aggregate agreement between the two datasets but may be important when analyzing the microstructure of the data and when matching entries day-by-day. It is important to also mention that both IBC and SIGACT improved their overall reporting throughout the conflict. Taking into account that additional biases may arise from reporting during intense conflict periods as discussed before, we would therefore expect that:

(IV) The most accurate day-by-day agreement between the two datasets should be found in the later, less violent stages of the war.

We will return to these four theoretical expectations when analyzing and interpreting the results of our quantitative data comparisons.

Before turning to our analysis of the data on Iraq we would like to emphasize that issues of data bias are, of course, not unique to conflict event data. Researchers, for example, increasingly rely on social media data - such as Twitter messages - to analyze social dynamics [22]. Similar to conflict event data, these messages are time-stamped and carry location information. The same is true for data on human mobility derived from mobile phone traces that provide detailed time-resolved information about the location of users [35]. In both cases, data may be subject to biases that arise from non-uniform geographic coverage: globally Twitter is known to be heavily biased towards users from North America, Europe and Asia [36] but it also tends to be biased towards urban populations in each country [37]. Mobile phone traces rely on data released by phone companies. Since customer base and coverage of companies tend to vary across regions, they may also have a distinct geographic bias.<sup>§</sup> As in the case of conflict event data the character of the data source may also lead to bias. Twitter, for example, only represents a small, non-representative sam-

**Table 1 Datasets**

Codename	Number of events		Number of casualties	
	KIA	KIA + WIA	KIA	KIA + WIA
IBC Baghdad	9,068		29,359-31,128	
SIGACT Baghdad	18,157	18,504	33,688	59,276
SIGACT 20 km	17,533	17,854	32,522	57,151
SIGACT 30 km	18,548	18,919	34,450	60,465
SIGACT 40 km	19,369	19,782	36,061	63,215

ple of the overall population [37]. And a recent study of the web presence of scientists on Wikipedia found that influential academic scholars are poorly represented [38]. This suggests that any scientometric analyses based on Wikipedia entries would have a strong relative bias compared to studies based on Facebook and Twitter, which tend to be much more consistent with citation-based metrics of academic impact [39]. The similarities in the sources of bias thus suggest that analyzing the implications of systematic bias in conflict event data also has broader implications for analyses using similar data on other social processes.

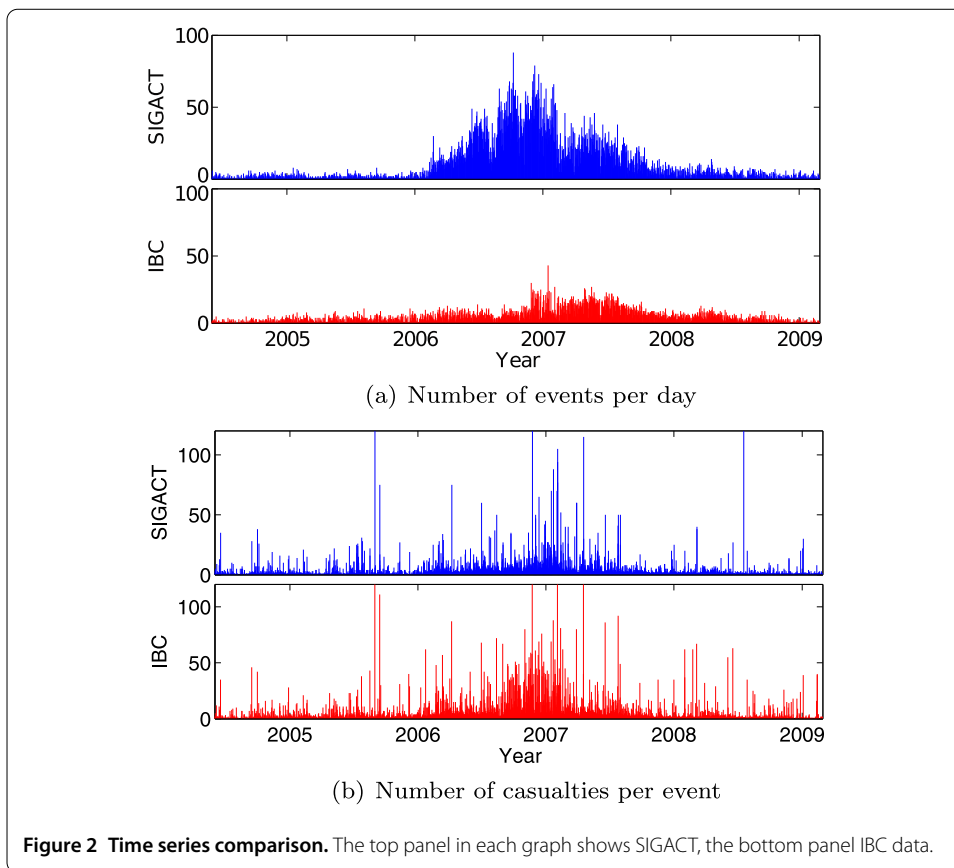
### 2.3 Baghdad data

The IBC Baghdad subset we analyze comprises events location-coded as ‘Baghdad’ but also those that carry more precise location tags such as ‘Sadr City’ or ‘Hurriya’. In the SIGACT dataset we rely on the U.S. military’s definition of the greater Baghdad area and the corresponding regional command ‘MND-BAGHDAD’. As a robustness check we then perform each of our analyses for subdatasets generated by selecting all events in SIGACT that fall within a radius of 20 km, 30 km and 40 km from the city center. These analyses confirm that the choice of dataset does not affect our substantive findings - whenever not directly reported in the manuscript the results can be found in Section 3 of Additional file 1.

Table 1 shows comparative statistics of the five Baghdad subdatasets used in our analysis: (a) IBC data filtered for events in the greater Baghdad area, (b) SIGACT data filtered by Baghdad regional command and by geo-coordinates for a radius of (c) 20 km, (d) 30 km and (e) 40 km from the city center. In the aggregate it appears as if IBC reports a much smaller number of events (approximately 2-3 times smaller than in the SIGACT data). The total number of deaths over the period of analysis also differs but is comparably more consistent. Figure 2(a) and (b) show time series of events per day and casualties per event for both datasets. Visual comparison already suggests that at a disaggregate level the datasets differ substantially with regard to the number of events per day and casualties per event reported. Note further that while both datasets capture the escalation of violence in 2006-2007, not only the number of events and casualty counts differ but also the timing of when violence escalated most.

## 3 Results

In recent quantitative studies casualty distributions in Iraq have been analyzed in aggregate form [1, 2], but studies mostly focus on time series of events - monthly, bi-weekly or most often daily [2, 3, 8, 17, 20]. In line with these different levels of analysis we will compare the reporting of IBC and SIGACT at different levels of disaggregation. We start



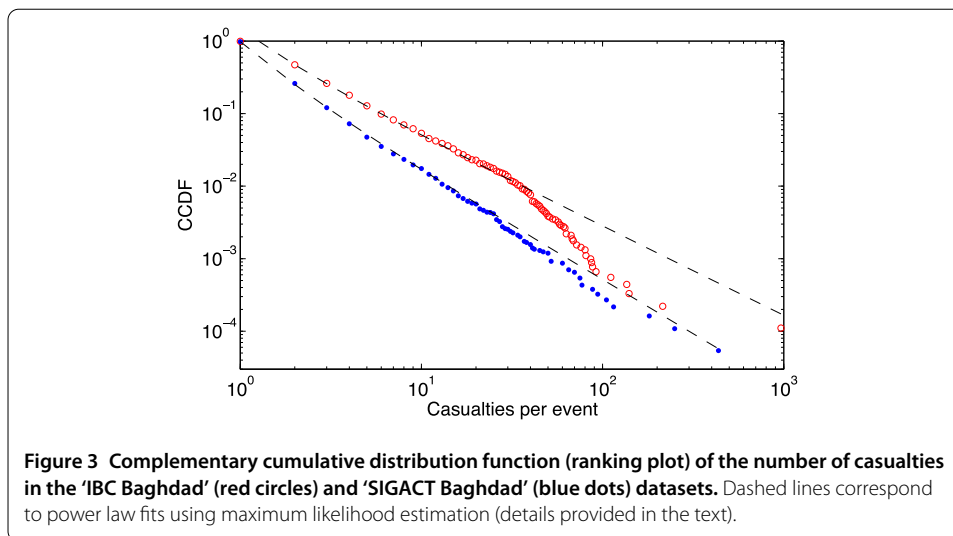
with aggregate data and then compare the datasets at increasingly smaller temporal resolutions. The (relative) biases we identify at each level of disaggregation can then be related to our theoretical expectations on structural differences in reporting.

### 3.1 Aggregate comparison

The two Baghdad datasets are relatively consistent in the total number of casualties reported: 29,441-31,222 in IBC and 32,531-36,213 in SIGACT (see also Table 1). They do, however, differ noticeably in the numbers of casualties reported per event (see Figure 2(b)). These differences in overall casualty counts can be best quantified by analyzing aggregate casualty size distributions. Figure 3 shows the complementary cumulative distribution function (ccdf) of the number of casualties in the datasets ‘IBC Baghdad’ and ‘SIGACT Baghdad’ on a log-log scale. The distributions for IBC and SIGACT both appear to follow a power law distribution but differ noticeably in their slopes and their tail behavior. Note that the distributions for the geo-filtered datasets (‘SIGACT 20 km’, ‘SIGACT 30 km’ and ‘SIGACT 40 km’) only differ slightly from ‘SIGACT Baghdad’ and are therefore not discussed separately here. In the case of discrete data, such as the casualty counts analyzed here, the ccdf of a power law distribution is given by:

$$P(x) = \frac{\zeta(\alpha, x)}{\zeta(\alpha, x_0)}, \quad x \geq x_0, \tag{1}$$





where  $P(x) = \Pr(X \geq x)$  is a probability of finding event with no less than  $x$  casualties,  $\zeta$  is a generalized Hurwitz zeta function [40],  $\alpha$  is the exponent of the power law distribution and  $x_0$  is the lower bound of the power law behavior.

To verify formally whether or not the distributions do indeed exhibit power law behavior we performed a maximum likelihood fit for a power law distribution using the methodology developed by Clauset et al. for analyzing power law behavior in empirical data [41]. The SIGACT data exhibits clear power law scaling (with exponent 2.57) starting at  $x_0 = 2$ , which is valid for almost 2.5 decades. In the IBC data, however, the presence of power law behavior is highly doubtful from a statistical point of view: the power law fit returns an exponent of 2.23, but the scaling is observed for only one decade and the tail clearly deviates from a power law distribution. Note that the power law shape of casualty event size statistics is a well-known empirical fact. It has been studied historically in the context of inter-state wars [42, 43] and more recently for terrorism [1] and intra-state conflict [2, 3]. We here do not intend to discuss the scaling relation of the distribution of event sizes and their possible origins but rather take these as 'stylized facts' and good quantitative indicators for marked differences between the two datasets. We would, however, like to note that in complex social or socio-economic systems deviations from power law may be indicative of incomplete data - see, for example, the discussion in [44] with respect to cyber-risk applications.

The significant upward shift of the IBC ccdf with respect to the SIGACT ccdf indicates the presence of much less small events (1-2 casualties) in the IBC data compared to SIGACT.<sup>h</sup> In order to quantify this difference we used a two-sample Anderson-Darling test [45, 46]. The test is a modification of the Kolmogorov-Smirnov (KS) test that gives more weight to the tail of the distribution and is thus a much better choice in the case of fat-tailed data [47]. Specifically, we use it to find the minimal threshold of casualty numbers for which the hypothesis of equal distribution of the two datasets can not be rejected. For this we proceeded as follows: For a given threshold, we select from both datasets only events with casualty counts greater or equal than a given threshold. We then apply a two-sample Anderson-Darling test (adjusted for ties) to test if both datasets were chosen from the same distribution. Varying the threshold value finally allows us to identify the minimal threshold for which the two datasets are statistically not distinguishable.

**Table 2 Results of the pairwise comparison of the distributions of casualties**

Threshold	Number of events					$A^2$ statistic			
	(i)	(ii)	(iii)	(iv)	(v)	(i)-(ii)	(i)-(iii)	(i)-(iv)	(i)-(v)
1	9,004	18,157	17,533	18,548	19,369	1,098.13	1,103.76	1,095.52	1,088.41
2	4,273	4,813	4,611	4,940	5,201	84.11	81.82	87.03	85.04
5	1163	876	851	901	952	6.60	7.11	7.81	7.92
10	484	323	310	325	340	7.25	6.72	6.63	6.82
15	296	159	154	161	169	<b>1.93</b>	<b>1.86</b>	<b>1.58</b>	<b>1.94</b>
20	206	105	100	105	108	<b>2.29</b>	<b>1.77</b>	<b>1.54</b>	<b>1.47</b>
25	159	77	75	79	82	3.09	<b>2.82</b>	<b>2.49</b>	<b>2.57</b>
30	123	47	47	51	52	<b>1.43</b>	<b>1.43</b>	<b>1.28</b>	<b>1.30</b>
40	69	29	29	31	32	<b>1.85</b>	<b>1.85</b>	<b>2.02</b>	<b>1.87</b>

The datasets are (i) 'IBC Baghdad', (ii) 'SIGACT Baghdad', (iii) 'SIGACT 20 km', (iv) 'SIGACT 30 km' and (v) 'SIGACT 40 km'. We used a two-sample Anderson-Darling tests (adjusted for ties) for comparison. Bold font marks cases where the value of the Anderson-Darling statistic  $A^2$  is smaller than the critical level  $A_{0.05}^2 = 2.492$  (large-sample approximation) and the hypothesis of two datasets being sampled from the same distribution can not be rejected at a 5% significance level.

The results are shown in Table 2. The relative comparison of IBC data (i) and SIGACT data (ii)-(v) clearly shows that IBC under-reports small events and over-reports larger events compared to SIGACT. While the total number of events in the IBC dataset is almost two times smaller than in SIGACT, the number of events with 2 or more casualties in both datasets are almost equal. For larger casualty sizes IBC even reports almost twice as many events with 25 casualties and more compared to SIGACT. Note that this, of course, also implies a considerably larger absolute fraction of events with 2 and more casualties in IBC which is clearly reflected in the flatter slope of the IBC cdf compared to SIGACT. Overall, this points to very significant differences in the aggregate casualty statistics between the two datasets.

These differences are also confirmed by our statistical tests. The hypothesis that the casualty distribution in IBC and SIGACT were sampled from the same distribution can be easily rejected for small thresholds (1-10 casualties per event, see Table 2 columns 7-10). The Anderson-Darling  $A^2$  statistic reaches the critical value for a significance level of 0.05 and stays below it only for thresholds starting at 15 and more casualties. The hypothesis of agreement can again be rejected for threshold values between 22-28 where the value of the  $A^2$  statistic stays slightly higher than critical level. Note, however, that a threshold of 15 casualties already selects only a very small subset of events from the whole dataset - less than 300 in IBC and less than 160 in SIGACT for the whole 5 years of data, i.e., less than 3% and 0.8% correspondingly. For thresholds greater than 25 casualties, subsets of the SIGACT datasets are even smaller (less than 100 events). In the quantitative comparisons of the two datasets in the following sections we therefore focus only on reasonably small thresholds of 1-10 casualties.

At an aggregate level, our analysis overall quantitatively confirms that IBC both reports systematically less events with few casualties (I) and more events with many casualties (III) compared to SIGACT - we can not test expectation (II) or (IV) here since these require a disaggregated comparison. It is important to point out that the differences in the casualty reporting we observe extend to the four most violent incidents in the period analyzed. In fact, their casualty counts in IBC and SIGACT disagree significantly, with IBC reporting more casualties in all four cases (Table 3).

**Table 3 Most violent events and number of casualties reported by IBC and SIGACT**

Date	Event	IBC report	SIGACT report
August 31, 2005	Baghdad bridge stampede <sup>*</sup>	965-1,005	436
November 23, 2006	Sadr City car and mortar bombings <sup>†</sup>	215	181
April 18, 2007	Baghdad car bombings <sup>‡</sup>	140	115
February 3, 2007	Baghdad market bombing <sup>§</sup>	136-137	105

<sup>\*</sup>A cry of suicide bomber, and 700 perish in Iraq stampede, The Guardian, <http://www.guardian.co.uk/world/2005/sep/01/iraq.rorycarroll1> (accessed: 08/07/2013)

<sup>†</sup>Iraq, Nov 23, 2006: A Day in Hell, Spiegel Online, <http://www.spiegel.de/international/world/iraq-nov-23-2006-a-day-in-hell-a-722544.html> (accessed: 08/07/2013)

<sup>‡</sup>Up to 200 killed in Baghdad bombs, BBC News, [http://news.bbc.co.uk/2/hi/middle\\_east/6567329.stm](http://news.bbc.co.uk/2/hi/middle_east/6567329.stm) (accessed: 08/07/2013)

<sup>§</sup>Terror takes toll on market, vendors, The Washington Times, <http://www.washingtontimes.com/news/2007/feb/6/20070206-115808-3925r/> (accessed: 08/07/2013)

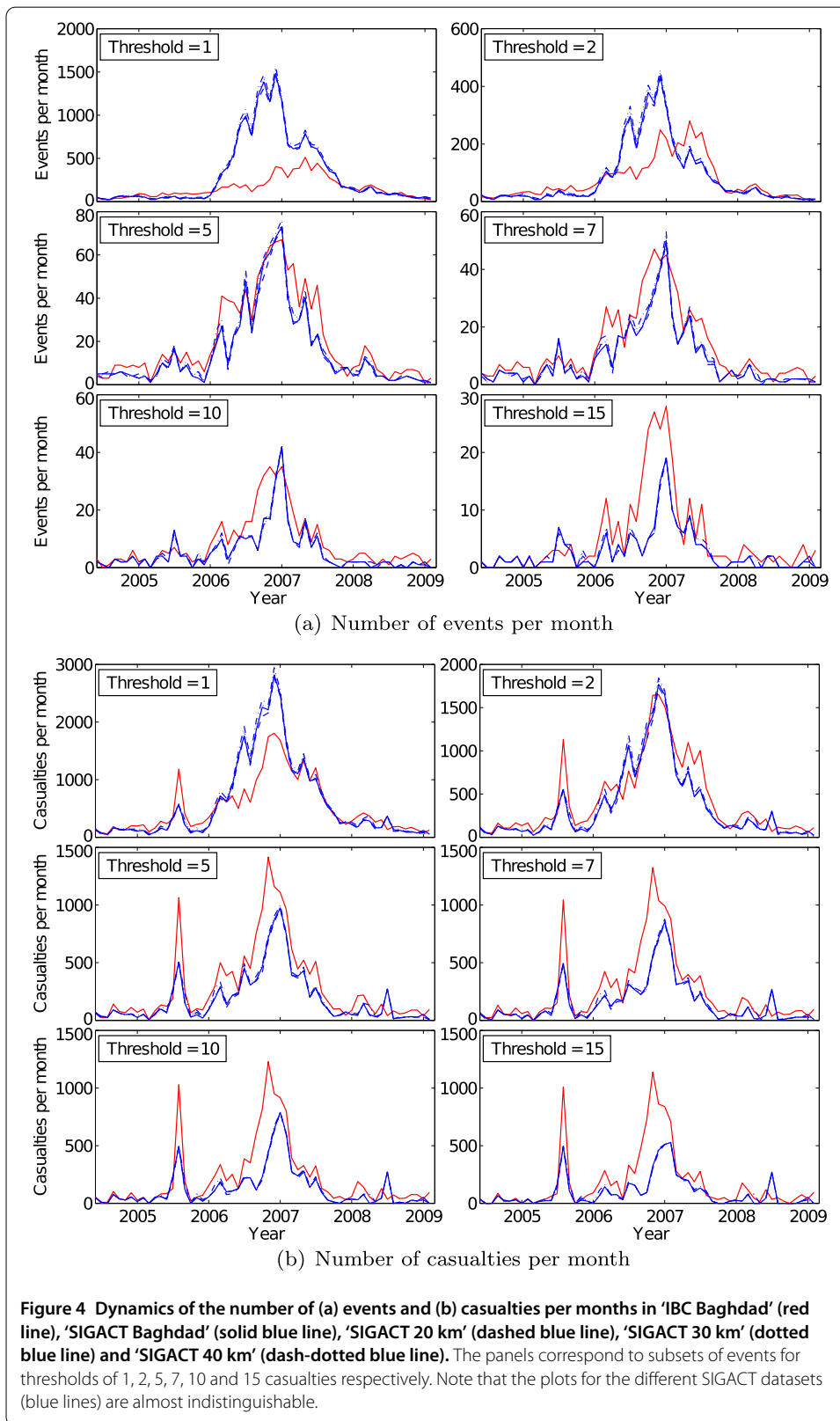
### 3.2 Monthly time series comparison

While aggregate distributional measures of conflict event signatures may already provide unique insights into conflict dynamics [1, 2], the majority of recent studies analyzing conflict mechanisms in Iraq relies on more detailed time series of incidents and their severity [3, 8, 17–20]. In this section we first focus on monthly time series. Note that we again consider a number of subsets with different minimal event sizes to account for the fact that the agreement between the two datasets may vary with the size of the events reported.

Figure 4(a) shows the number of events, Figure 4(b) the number of casualties per month in all five Baghdad datasets (see Table 1) for thresholds of 1, 2, 5, 7, 10 and 15 casualties per event. The panel in the upper left hand corner of each graph depicts the full IBC and SIGACT data (threshold equal to 1). It suggests that at the monthly level the two datasets provide distinctly different accounts of the violence dynamics in Baghdad. These differences in the number of events appear to be most substantial during the escalation of violence in 2006-2007 and for low and high thresholds. If we only exclude events with less than 5 to 10 casualties per event - i.e., intermediate thresholds - the monthly dynamics in the two datasets qualitatively agree much better (Figure 4(a)).

Before turning to a more detailed analysis of the differences in the monthly IBC and SIGACT reporting, we first tested whether at least the overall trends in both the number of events and casualties per month are consistent. A two-step Engle-Granger cointegration test [48] with an augmented Dickey-Fuller test of residuals [49, 50] can reject the null hypothesis of no-cointegration at a 5% significance level for almost all thresholds analyzed here. In other words, the differences in reporting between IBC and SIGACT generally do not affect the agreement of the coarse-grained trends. The exception are the dynamics of the number of events per month for thresholds of 1, 2 or 3 casualties per event (top panels of Figure 4(a)). Here the Engle-Granger test can not reject the null of no-cointegration (with *p*-values of Dickey-Fuller test equal to 0.653, 0.650 and 0.503 respectively), which suggests that even the long-term trends in the complete IBC and SIGACT datasets are statistically significantly different.

Overall, the differences in the monthly reporting of IBC and SIGACT are consistent with those observed in the aggregate statistics (Section 3.1). We also find the same casualty size dependent relative bias between the two datasets at the level of months. In particular, we again find significantly more small events in SIGACT compared to IBC in line with (I). However, this is only true during the 2006-2007 escalation of violence. In fact, before 2006 IBC even reports more small events and 2008 and onward the two datasets largely



agree. This is consistent with our assertion that reporting differs more noticeably the more intense the conflict (II) and also suggests that - apart from the escalation in 2006-2007 - IBC and SIGACT reporting of small events is, in fact, quite consistent. Note, however, that we also clearly see an overall tendency of IBC to report more events with many casualties almost all throughout the conflict (III). This attests to differences in reporting also in the less intensive phases of the conflict prior to 2006 and after 2007.

Figure 4(a) and (b) also suggest that there is not one threshold value for which IBC and SIGACT reporting agrees both in terms of number of events and casualties per month. While they show the best visual agreement with respect to casualty counts for a threshold of 2 (Figure 4(b), upper right panel), the corresponding events per month statistics differ markedly (Figure 4(a), upper right panel). Recall, however, that we argued before that coverage in IBC should be much more limited for small events than in SIGACT. This implies that we should actually not expect an agreement in the number of events per months for thresholds of 1 and 2. In fact, the number of events per month are most consistent for thresholds between 5 and 10 where media-based coverage should be more complete. Since the casualty counts in IBC are significantly larger for these thresholds, this appears to suggest that overall IBC systematically reports more casualties than SIGACT.

It is important to keep in mind, however, that we previously also identified a second possible source of bias that may lead to a similar effect: the reporting of one composite episode as several incidents with less fatalities in SIGACT. In fact, for large events in the SIGACT dataset one can typically find a counterpart in the IBC dataset within the same day or two. In contrast, quite a number of events reported by IBC do not have an equally sized counterpart in the SIGACT dataset (see also Section 3.3). Since there are typically many events within a short time window one can, unfortunately, typically not convincingly establish if there are a number of smaller incidents reported in SIGACT that taken together match or approximate the total casualty count of an episode in IBC. This makes it impossible to estimate the extent to which possible mis-reporting of episodes as separate incidents may affect the reporting in SIGACT. Overall, we can therefore only say with certainty that the differences in casualty reporting observed at a monthly level are consistent with IBC systematically reporting more casualties than SIGACT, mis-reporting of episodes as separate incidents in SIGACT, and/or a combination of both.

### 3.3 Daily time series comparison

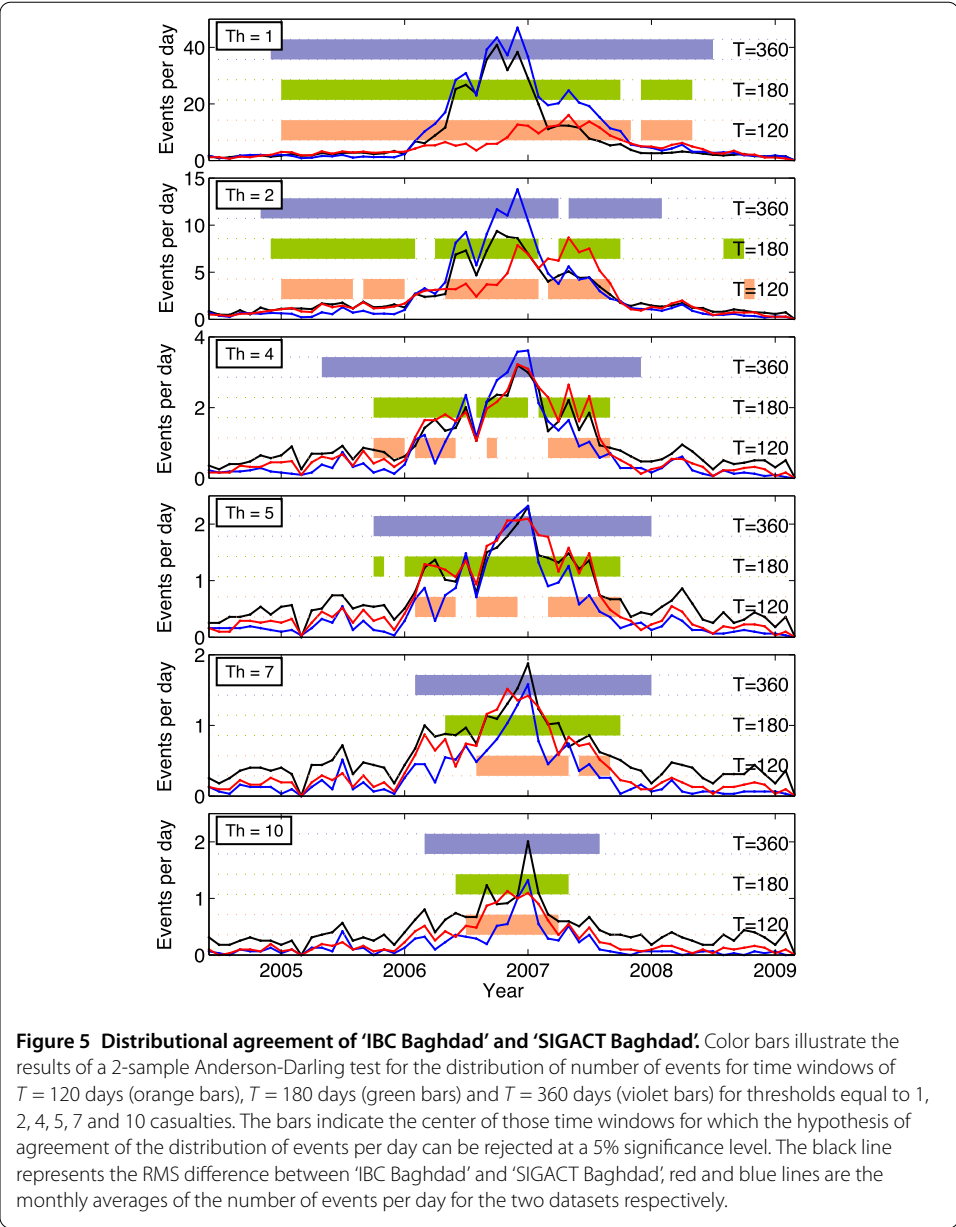
Many of the recent quantitative studies of the conflict in Iraq rely on detailed daily time series. We therefore now turn to a statistical analysis of deviations in the day-by-day microstructure of reporting between IBC and SIGACT. Note that in the period 2004-2009 both datasets exhibit a high degree of non-stationarity (see Figure 4(a)). In fact, the number of events in the second half of 2006 and first half of 2007 is up to 10 times larger than in 2005 or 2009. Any statistical analysis of these data thus requires us to explicitly model this non-stationarity, for instance using parametric methods. Alternatively, we can restrict our analyses to sufficiently small time windows, in which the dynamics can be assumed to be (approximately) stationary. In line with previous works (see for example [2]), we here pursue the latter approach and employ standard non-parametric tests to moving time windows. The choice of appropriate window size is subjected to trade-offs: it should be as small as possible to guarantee a stationary regime but also sufficiently large to contain sufficiently many events for robust statistical tests. We found that time windows ranging

from 4 months to half a year ( $T = 120$  days to  $T = 180$  days) fulfill both of these conditions.<sup>1</sup> However, we also performed our tests for a window size of 1 year ( $T = 360$ ) as a robustness check.

For every window size  $T$  we slide the moving window across the whole range of data in steps of one month and extract the subset of events in both IBC and SIGACT within each time window. For each of the (approximately) stationary periods we can then compare the distribution of events per day as a measure of the day-by-day microstructure of the data using a two-sample Anderson-Darling test. The Anderson-Darling test rejects the hypothesis of both time-series being sampled from the same distribution if the statistic  $A^2$  is smaller than the critical level  $A_{0.05}^2$  for a significance level of 0.05. Since the number of samples (window size  $T$ ) is sufficiently large we use the large sample approximation for the critical level  $A_{0.05}^2 = 2.492$  [45]. Note that in contrast to the distribution of casualties per event (Figure 3), the distributions of events per day do not have fat-tails and typically decay almost exponentially (Figure S7 in Additional file 1). A Kolmogorov-Smirnov test would thus also in principle be applicable here [47]. However, in order to be consistent throughout our analysis and to account for the slower-than-exponential tails in case of small thresholds of 1 and 2 casualties per event, we here also rely on the more rigorous Anderson-Darling test.

Figure 5 graphically illustrates the results of the Anderson-Darling test for different thresholds and different window sizes. Color bars indicate the center of all windows of size  $T$  for which the null hypothesis of the number of events per day in both datasets being sampled from the same distribution can be rejected at a 5% significance level. The figure clearly illustrates that the two datasets significantly differ with respect to the distribution of events per day: the distributions in the two full datasets (threshold equal to 1, top panel) are statistically distinguishable from 2005 through 2007; only in the initial phase of the conflict and in the calmer phase after the U.S. military troop 'surge' in 2007 we can not detect significant differences. The higher the threshold, i.e., the more small events we exclude, the better the distributional agreement. It is important to note that in case of large differences in the numbers of events per day, the Anderson-Darling test will indicate significant deviations of one sample from another irrespective of the temporal characteristics. This certainly contributes to the strong disagreement for thresholds of 1 and 2 casualties in 2006-2007 but should not affect the results elsewhere where the numbers of events are much more similar. In general, the results for different window sizes are quite consistent and we can be confident that the exact choice of time window does not systematically drive our results.

The analysis in Figure 5 highlights that even though the average number of small events (thresholds 1 and 2) are relatively similar in IBC and SIGACT prior to 2006 and after 2007 the detailed daily reporting may still significantly differ, for example, in 2005 or in early 2008 (top panel). In the period 2006-2007 the daily structure of small events reported in the two datasets is almost everywhere significantly different except for a short episode in early 2007. For larger events (threshold 4 and larger) the average number of events per day is much more consistent throughout, but in the most intense phase of the conflict 2006-2007 the distributions of events per day remain statistically distinguishable. For events with 10 casualties and more the difference is only significant mid-2006 through early 2007 at the height of the escalation. The fact that the microstructures of the datasets become statistically indistinguishable does of course not imply that they necessarily correspond



to the same day-by-day occurrence of events. The test simply determines whether or not the overall distributions of events per day in a given (comparably large) time window are distinguishable or not. Consider, for instance, the very simple example of two time series with alternating 1 and 3 events on two subsequent days, but where the occurrence of events in the second series is shifted by one day. These time series have the same average number of events per day and are statistically absolutely not distinguishable even though each day their number of events differs by two, their average number of events per day.

In order to better quantify the actual day-by-day correspondence between IBC and SIGACT we therefore additionally consider the root mean square (RMS) difference of the number of events in IBC ( $n_{IBC}(t)$ ) and SIGACT ( $n_{SIGACT}(t)$ ) for a sliding window of size  $T_2 - T_1 = 1$  as a simple quantitative metric of (average) daily agreement (black line in

Figure 5):

$$\text{RMS} = \sqrt{\frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} (n_{\text{IBC}}(t) - n_{\text{SIGACT}}(t))^2}. \quad (2)$$

This difference can be directly compared to the average numbers of events per day in both IBC and SIGACT for the same moving time window (red and blue line in Figure 5 respectively):

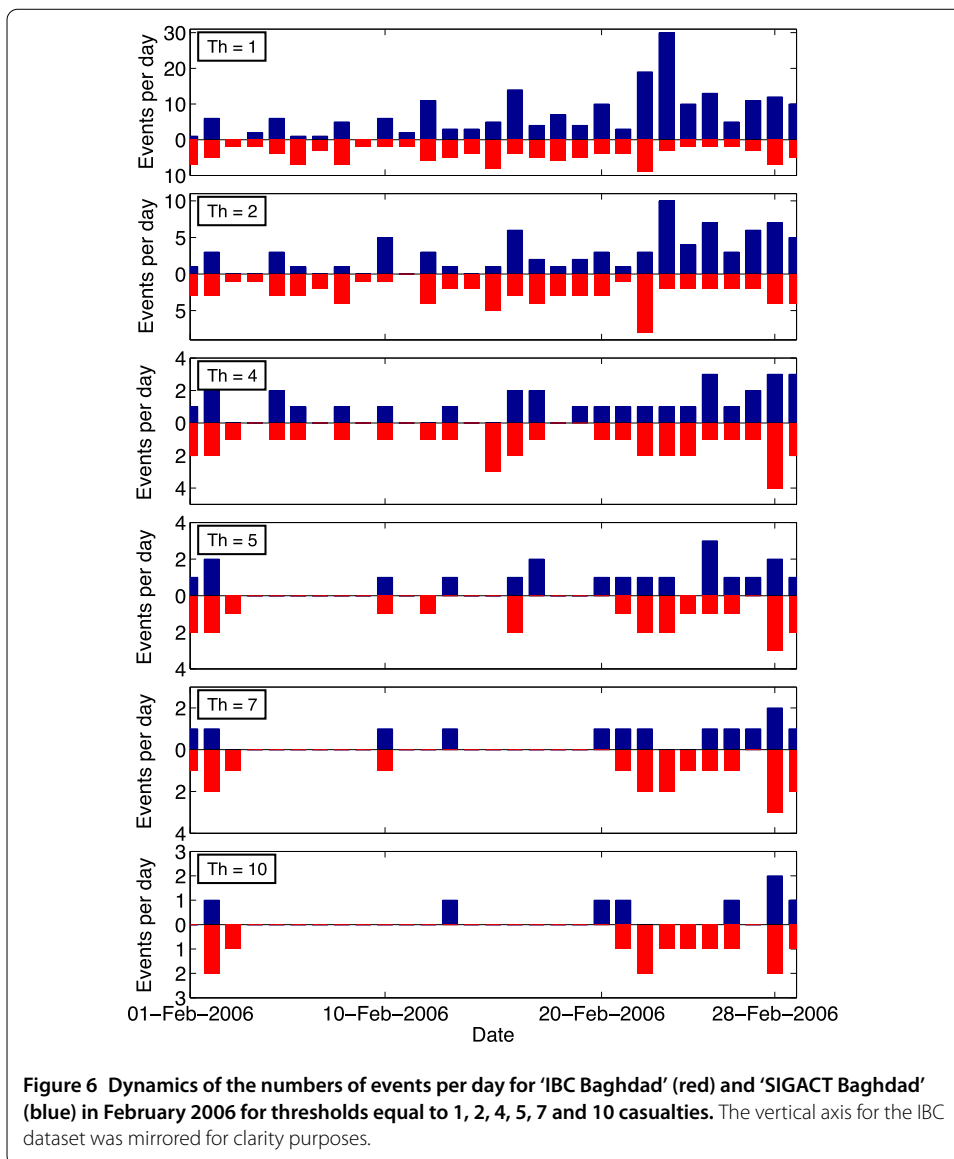
$$\overline{n_{\text{IBC}}} = \frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} n_{\text{IBC}}(t), \quad \overline{n_{\text{SIGACT}}} = \frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} n_{\text{SIGACT}}(t). \quad (3)$$

We find that the RMS difference is always of the order of magnitude of the average numbers of events per day for all thresholds we consider. In other words, the typical difference between two datasets is equal to the typical number of events per day. This is true even for intermediate thresholds of 5-10 casualties per event where the cumulative monthly number of events reported in IBC and SIGACT agree quite well. Note further that the RMS differences 2008 and onward is not significantly smaller than prior to 2006 contrary to our theoretical expectation that difference in reporting should be smallest in the later, less violent phases of the conflict (IV).

To test our intuition for how day-by-day differences relate to distributional agreement, we analyze the daily agreement in IBC and SIGACT in February 2006. We chose this period specifically such that the two datasets are statistically distinguishable for small and indistinguishable for large thresholds (see Figure 5). Figure 6 graphically illustrates the direct comparison of the number of events reported in each dataset. It is visually apparent that the number of events per day with thresholds of 1 and 2 casualties (upper two panels) reported in SIGACT and IBC differ. Specifically, on some days SIGACT reports more events, on others IBC does, and there are also days when one of the datasets reports no event but the other one does. For larger events (up to 4 and 5 casualties, third and fourth panel) the numbers of events per day in both datasets are much more consistent but there are still significant differences. SIGACT, for example, at a threshold of 5 reports significantly more days with one event than IBC and less days with two events. For thresholds of 7 and larger (lower two panels) the distributions of events per day are statistically not distinguishable anymore. In the day-by-day comparison we see that each daily signature is dominated by days with no, one or two events and the occurrence of these days is overall quite similar. Note, however, that at the same time for well more than 50% of the days these counts do not coincide, which explains the day-by-day mismatch represented by the comparably large RMS differences (Figure 5).

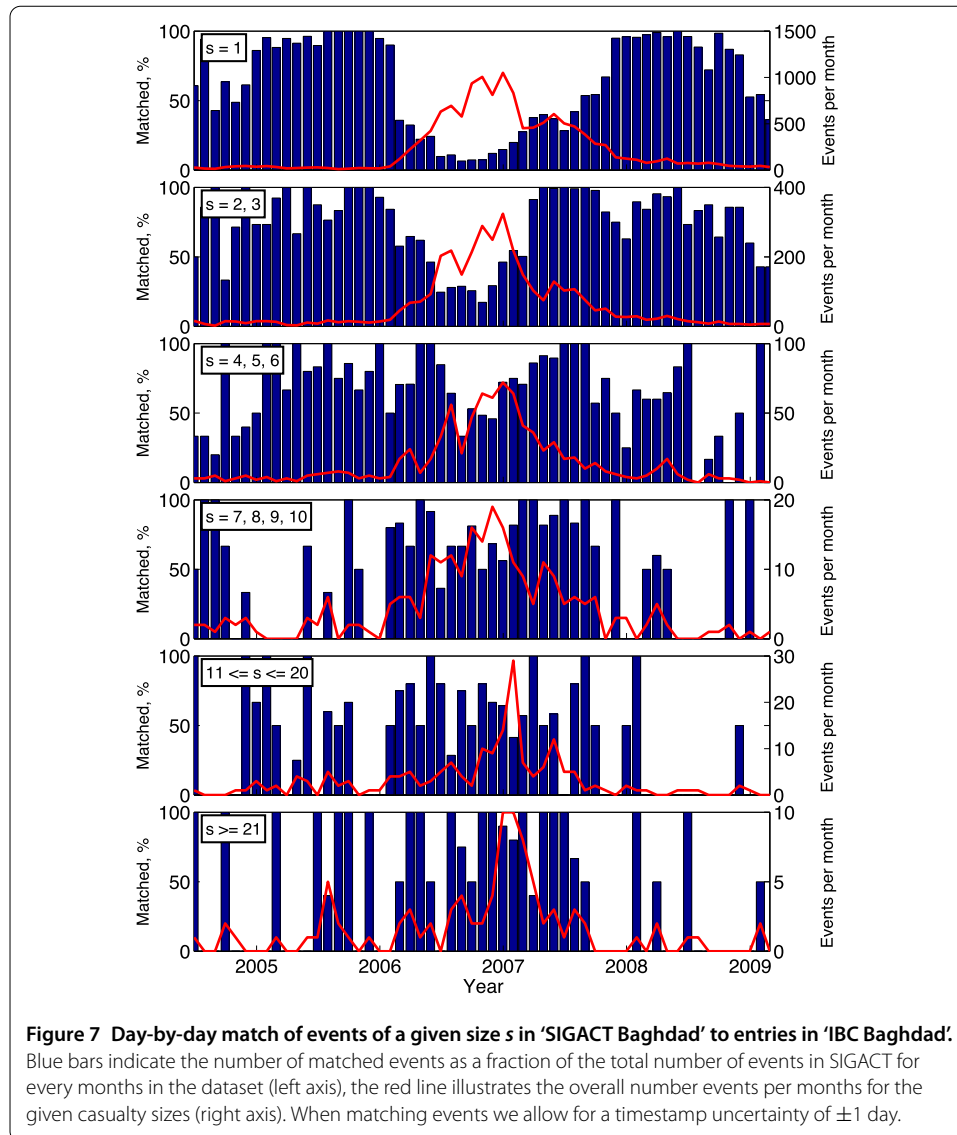
The large RMS difference we observe throughout the whole dataset should therefore be an indication that the day-by-day structure of event reporting in SIGACT and IBC does indeed significantly differ - despite the fact that they may be statistically indistinguishable at an aggregate or distributional level. In order to quantitatively estimate this daily mismatch, we compared how many events of a given size in SIGACT - the dataset with more events - can be matched to events in IBC. In matching events we allow for an uncertainty of  $\pm 1$  day. Please refer to Section 2 of Additional file 1 for the details of our automated





matching procedure. Figure 7 shows the number of matched events (blue bars) as a fraction of the total number of events in SIGACT (red line) for every month in the dataset. For simplicity we have grouped casualty sizes in categories. Note that for months with no events in a given casualty category, the fraction of matched events is set to 0 by default.

The figure suggests that daily SIGACT and IBC records are most consistent outside of the escalation of violence in 2006-2007 - this is particularly true for events with less casualties. Excluding the escalation phase 2006-2007 we find that on average 85.8% of the entries with 1 casualty and 82.3% of the entries with 2 or 3 casualties in SIGACT coincide with an entry with the same number of casualties within  $\pm 1$  day in IBC (Table 4). In contrast, during the period 2006-2007 only 24.6% of SIGACT reports with 1 casualty - by far the largest share of incidents - can be matched to IBC entries. In the same period, 50.9% of SIGACT records with 2 and 3 casualties have a corresponding entry in IBC within  $\pm 1$  day. For events with few casualties we can thus also confirm at a day-by-day resolution that differences in the reporting are generally larger the more intense the conflict (II). In con-



trast, the day-by-day agreement of events with 4 and more casualties is generally better in the 2006-2007 period (see Table 4 for details). Notice that especially the match of very large events (more than 20 casualties) is generally very good throughout (77.8% match). Finally, we do not find any systematic evidence that the detailed match of SIGACT and IBC has increased significantly after 2008, contrary to our theoretical expectation (IV).

It is important to emphasize here that we thus far only considered a one-sided comparison that matches SIGACT events to IBC. We previously observed that IBC reports more events with many casualties than SIGACT (Figure 4(a)), i.e., matching IBC to SIGACT events will yield a noticeably lower match. For example, the match of events with more than 20 casualties in this case is only 37.3% (please refer to Section 2 of Additional file 1 for the full comparison). The large RMS difference in Figure 5 reflects this mismatch. Note, too, that the RMS difference is a measure of daily agreement whereas we here allow for a timestamp uncertainty of  $\pm 1$  day - it is consequently a much more conservative estimate of the agreement of the two time series than the one tested here. As we would expect, us-

**Table 4** Number of SIGACT reports matched to IBC entries

Casualties	2004-05 & 2008-09			2006-07		
	matched	total	%	matched	total	%
s = 1	1,264	1,473	85.81	2,925	11,871	24.63
s = 2, 3	343	417	82.25	1,556	3,054	50.94
s = 4-6	86	133	64.66	480	693	69.26
s = 7-10	22	45	48.88	149	202	73.76
s = 11-20	18	36	50.00	83	143	58.04
s > 20	15	23	65.21	55	67	82.08

ing smaller tolerance ( $\pm 0$  days) to match events generally decreases agreement while using larger tolerance ( $\pm 3$  days) increases agreement of SIGACT events with IBC (see Section 2 of Additional file 1 for details). There is one notable exception though: very large events (with more than 20 casualties) are equally well matched for all tolerances suggesting that their reporting is clearly the most consistent.

We validated our day-by-day comparison by comparing it to results of a study performed at Columbia University. In the study, a small random sample of SIGACT events with civilian casualties was compared to entries in the IBC database [51]. Specifically, students were tasked to manually match SIGACT entries to IBC events following a specific detailed protocol. The analysis revealed that only 23.8% of the events in their SIGACT sample had corresponding entries in IBC. The Columbia researchers noted though that most of the events in their sample had only very few casualties - a consequence of the fact that by randomly sampling events for their study they mainly selected incidents during the period 2006-2007 where by far the most SIGACT events were recorded. In fact, the large majority of records in this period reports only one casualty per event (see Table 4). In our analysis we find an agreement of 24.6% for these events in the 2006-2007 period, which is very consistent with the Columbia estimate. For events with more than 20 casualties 94.1% of the SIGACT entries could be matched to entries in IBC in the Columbia study. The estimate of 82.1% based on our automated comparison is similar but clearly more conservative. Note that the specification of timestamp uncertainty of  $\pm 1$  day used in our automated procedure is equivalent to the matching prescription used in the Columbia study (see Section 2 of Additional file 1 for details).

It is important to emphasize two key shortcomings of the manual, in-depth comparison performed in the Columbia study. Most importantly, the random selection of events across the whole dataset effectively limits their analysis to the period 2006-2007 - the period in which all of our previous analyses find the most significant disagreement between IBC and SIGACT. Their findings thus likely systematically underestimate the overall match of events. In fact, our analysis shows that for the full period of analysis 38.5% of all SIGACT records could be matched to IBC entries with the same number of casualties. This is significantly more than the 23.8% reported in the Columbia study. Furthermore, manual comparisons are only possible for small (random) subsets of event. Having verified that we obtain results consistent with an in-depth comparison by human coders, the clear advantage of an automated comparison is its coverage, i.e., it efficiently yields estimates of the correspondence of daily reports in IBC and SIGACT for the full period of analysis.

In summary, our results strongly suggest that at any level of analysis - aggregate statistics, monthly statistics, detailed distributional level and daily time series - IBC and SIGACT reporting differ significantly, most strongly for events with few casualties but also for

larger event sizes where aggregate event statistics are comparably more consistent. Consequently, we can expect that the choice of dataset would strongly affect any inference we draw from these data, simply because the conflict dynamics represented in each datasets at any level of analysis are indeed quite different.

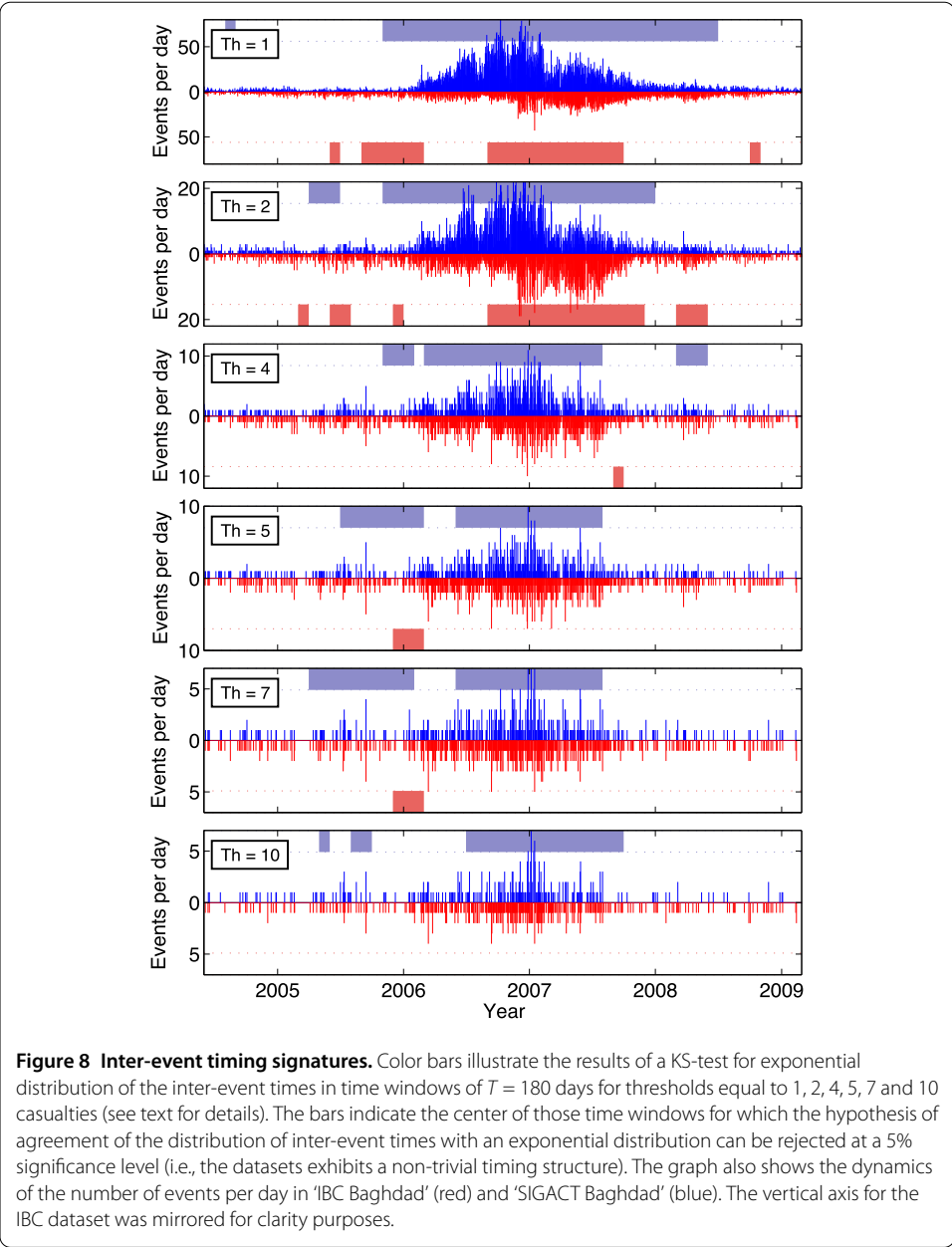
In the following sections we complement these comparative insights with an in-depth analysis of the reporting in each dataset. Specifically, we explore if and where the two datasets contain non-trivial timing information - i.e., information about the occurrence of subsequent events - and how robust these are to uncertainty in timestamps. This is, of course, a critical precondition for the use of the datasets for any kind of timing or causal analysis. It is complementary to our prior comparative analysis in the sense that both, either or neither of the datasets may actually be suitable to study event dynamics in Baghdad, regardless of the relative differences in reporting we have already identified.

### 3.4 Distributional signatures

In Section 3.3 we used the distribution of events per day to characterize day-by-day event dynamics. A second very common measure that captures the micro-structure of event data is the distribution of times between incidents, or inter-event times [3]. The latter is always favorable if the data resolution is more fine-grained than days. Inter-event timing distributions at a resolution of hours, for example, provide a much more detailed characterization of the dynamics of subsequent events. We here chose to rely on the distribution of inter-event times because it also tends to be more sensitive to differences in the distribution of sparse data for which it is generally more difficult to detect deviations from a trivial timing signature. As before, we consider the dynamics in a given time window of length  $T$  within which the conflict dynamics can be assumed to be (approximately) stationary. Notice that the results for the event per day statistics are substantively equivalent; please refer to Section 5 of Additional file 1 for details.

In a structureless datasets, i.e., in datasets where the timing of events is statistically independent, the distribution of events per day simply follows a Poisson, the corresponding distribution of inter-event times an exponential distribution. The deviation of timing signatures from a Poissonian or exponential is thus mainly indicative of the usefulness of the dataset because a featureless dataset is essentially useless for any kind of quantitative (causal) inference or timing analysis. We would, however, also like to note that empirically and theoretically it is not plausible that the timing of conflict events in Iraq is completely independent. In fact, most theories of political violence prominently feature mechanisms that emphasize reciprocity and reactive dynamics [8, 52], spatial spillover effects or diffusion of violence [5].

Figure 8 shows the number of events per day for both datasets and graphically illustrates the results of a Kolmogorov-Smirnov test for a moving window of 180 days (results for larger window sizes are consistent and are discussed in Section 5 of Additional file 1). Specifically, bars indicate the center of time windows for which the Kolmogorov-Smirnov test rejects the hypothesis of agreement of the distribution of inter-event times with an exponential distribution at a 5% significance level. The analysis suggests that in the full SIGACT Baghdad dataset the timing of events deviates significantly from that of a Poisson process all throughout 2006 to mid-2008. In the much calmer periods prior to 2006 and after mid-2008 the timing signature, however, does not deviate significantly from that of a featureless process. For events larger than thresholds of 2, 4, 5, 7 and 10 casualties,



SIGACT still consistently features periods where the timing of events does not follow a featureless Poisson process, mainly in the most violent period mid-2006 to mid-2007.

In the full IBC dataset and for events with more than 2 casualties the timing of events also has a significant non-trivial timing structure that allows to reject the null hypothesis of Poisson dynamics for periods throughout late 2005 to 2007. This finding, however, is much less robust than for the SIGACT data. In fact, there is a half-year stretch in early 2006 for the full IBC dataset that features only a trivial timing signature. For a threshold of 2, the inter-event signature is also not distinguishable from a Poissonian in a period from late 2005 to late 2006; notice that in both periods the number of events per day is quite large. The differences between the signatures in IBC and SIGACT are most pronounced for subsets of events with minimally 4 or more casualties. Even though the overall num-

ber of events in SIGACT and IBC is comparable for those subsets, there is hardly any time window for which the timing signature in IBC significantly differs from that of a featureless process. This is especially obvious in the escalation phase mid-2006 to mid-2007 where the timing of events in IBC is statistically independent everywhere but deviates significantly from a featureless process in SIGACT.

As emphasized before, based on theories of political violence, we would expect that the timing of events should not be independent. The empirical narrative of the conflict in Iraq similarly suggests that events tend to be related. It is, however, in general not possible to decide whether or not the absence of non-trivial signatures in these periods is a consequence of incomplete reporting or evidence that the timing of events of a given size is indeed uncorrelated. The fact that both datasets feature time windows with trivial timing signatures thus simply suggests that it would be ill-advised to use the respective datasets in these periods to study (causal) relations between the timing of events. This is true for large parts of the IBC data - especially for larger thresholds - whereas SIGACT generally features more and longer time windows with non-trivial timing signatures (Figure 8). Notice though that in the low intensity conflict phases prior to 2006 and also after mid-2008 our statistical tests do not indicate any non-trivial timing signatures in SIGACT either.

Overall IBC appears to be much less suitable to study timing dynamics and thus to infer (causal) relationships between events. This is consistent with our observation in Section 2.2 that the reporting of timestamp in IBC may be more constrained through the use of approximate - or possibly misreported - timing of events provided in newspaper articles. It is important to keep in mind though that we only tested for non-trivial timing signatures in data drawn from the whole Baghdad area - significant correlations in the timing of events may, for example, simply be limited to smaller geographic scales.

### 3.5 Uncertainty of timestamps

We now turn to a systematic test of the effect of timestamp uncertainty on the distributional features analyzed in the previous section. In other words, we address the question of how robust the timing signatures we find are to uncertainties in the coding of timestamps. The robust coding of event timestamps is critically important for any quantitative technique where inferences hinge on the (causal) order of events. Examples of commonly used techniques using such time-ordered data include point process models, such as self-excited Hawkes processes [53, 54], Autoregressive Conditional Durations (ACD) [55, 56] or Autoregressive Conditional Intensity (ACI) [57]. Note that in both IBC and SIGACT the reporting of event timing may, in principle, be subject to systematic coding inaccuracies. The media sources IBC relies on may report events with a delay, provide only approximate timing information or may misreport the timing of an event altogether. SIGACT data is compiled from field reports, which may also systematically miscode the true timing of an event. Common problems include delayed reporting in situations of heavy engagement with enemy forces, reporting *post hoc* on incidents that a unit was not directly involved in and for which the timing is not precisely known, or summary reports filed at the end of a day (see also Section 2.2).

In order to statistically characterize the effect of timestamp inaccuracies on the day-by-day signatures of events, we again rely on the distribution of inter-event times  $\tau_i = t_i - t_{i-1}$ . We further assume that both IBC and SIGACT report events with timestamp uncertainties  $\Delta_{\text{IBC}}$  and  $\Delta_{\text{SIGACT}}$ . Note that the IBC dataset only codes timing of events

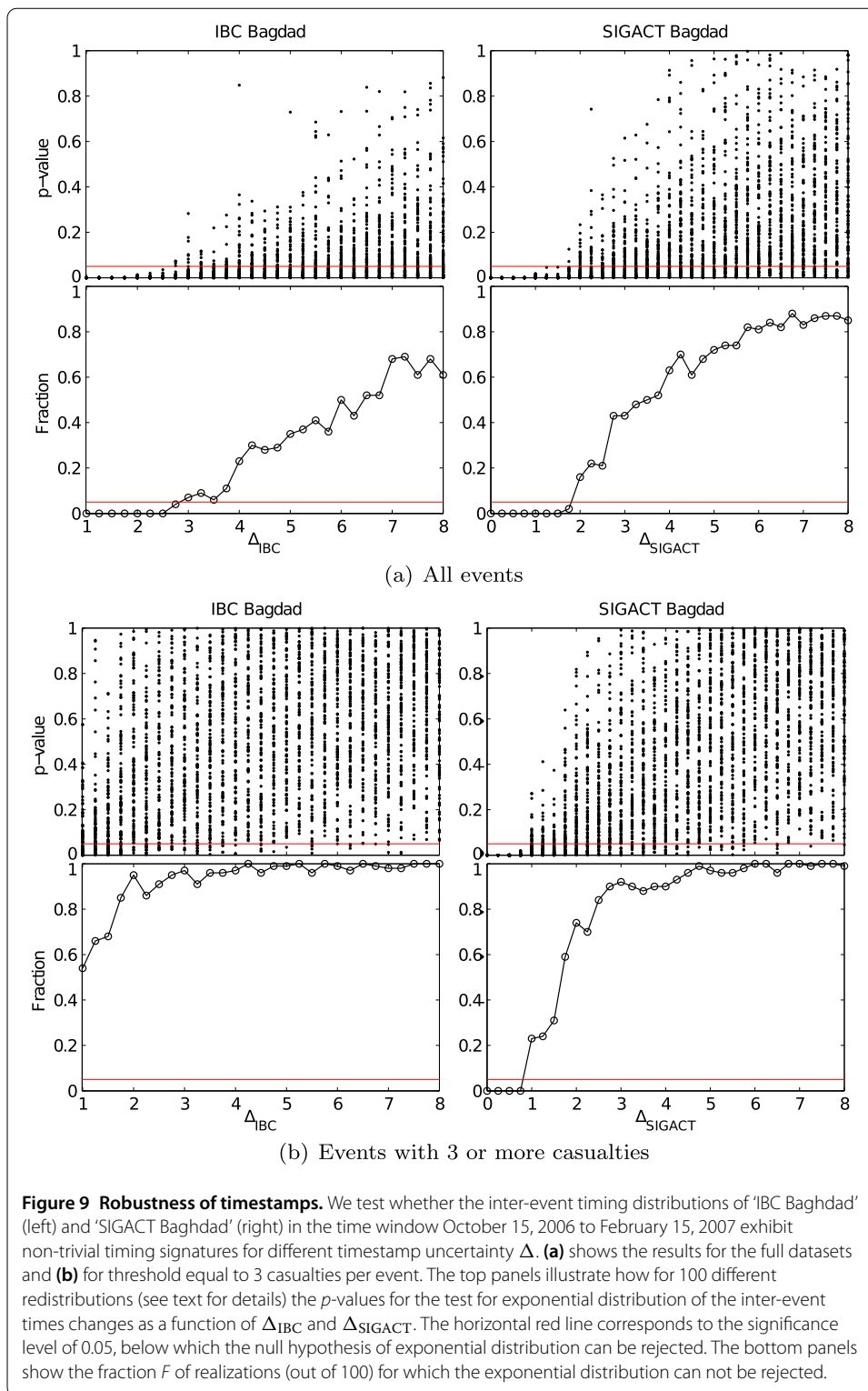
with a precision of days, i.e.,  $\Delta_{IBC} \geq 1$  day. SIGACT on the other hand carries much more precise timestamps with a resolution of minutes and thus does not have this constraint. In order to account for uncertainties  $\Delta$  in the timestamps we adopted the methodology proposed in [58] and assume that the difference between the real time of an event  $\tilde{t}_i$  (which is unknown) and the timestamp  $t_i \geq \tilde{t}_i$  is some effective ‘noise’  $\xi_i = t_i - \tilde{t}_i < \Delta$ .

To test the impact of a given uncertainty  $\Delta$  on the timing signature in each time series we then proceed as follows. For a given time window  $T$  we draw random variables  $\xi_{i,IBC}$  and  $\xi_{i,SIGACT}$  from the uniform distributions  $U([0, \Delta_{IBC}])$  and  $U([0, \Delta_{SIGACT}])$  respectively. We then construct time series  $\hat{t}_{i,IBC} = t_{i,IBC} - \xi_{i,IBC}$  and  $\hat{t}_{i,SIGACT} = t_{i,SIGACT} - \xi_{i,SIGACT}$ , and calculate the distribution of inter-event times  $\hat{\tau}_{i,IBC} = \hat{t}_{i,IBC} - \hat{t}_{i-1,IBC}$  and  $\hat{\tau}_{i,SIGACT} = \hat{t}_{i,SIGACT} - \hat{t}_{i-1,SIGACT}$  for each. Note that the values  $\hat{\tau}_i$  represent proxies for the unobserved real values of inter-event times  $\tilde{\tau}_i$ . We then apply a two sample Anderson-Darling test to the distributions of these inter-event times (for both IBC and SIGACT independently). We repeat this procedure  $M = 100$  times, generating a set of binary values  $\{h_{j,IBC}\}$  and  $\{h_{j,SIGACT}\}$ ,  $j = 1, \dots, M$ , where  $h_j = 0$  if we can reject the null hypothesis at a 5% significance level, and  $h_j = 1$  if the null hypothesis can not be rejected.

The effective measure for whether or not the timing distributions of the two time series with uncertainties are distinguishable is then simply the fraction of cases when the null hypothesis can not be rejected:  $F_{IBC} = \sum_{j=1}^M h_{j,IBC} / M$  and  $F_{SIGACT} = \sum_{j=1}^M h_{j,SIGACT} / M$ . If the value of  $F_{IBC}$  (or  $F_{SIGACT}$ ) is close to 0 we can be certain that the distributions of inter-event times  $\hat{\tau}_{i,IBC}$  (or  $\hat{\tau}_{i,SIGACT}$ ) are different from an exponential distribution - independently of particular values of the ‘noise’ terms  $\xi_{i,IBC}$  (or  $\xi_{i,SIGACT}$  respectively). This also implies that the real inter-event times  $\tilde{\tau}_{i,IBC}$  (or  $\tilde{\tau}_{i,SIGACT}$ ) exhibit non-trivial clustering. Similarly, a value of  $F$  close to 1 suggests that for most of the cases we can not reject the null hypothesis for the proxy values  $\hat{\tau}_i$ . This, in turn, implies that we will most likely not reject the null hypothesis at the same significance level for the real (unobserved) values  $\tilde{\tau}_i$ .<sup>j</sup> Effectively the fraction  $F$  may thus be referred to as the ‘likelihood’ of the time series to have been generated by a Poisson process.

From a conceptual point of views, the random time shifts  $\hat{t}_i = t_i - \xi_i$  simply introduce bias to the time-series: the larger  $\Delta$ , the larger the ‘randomness’ in our proxy time-series  $\hat{t}_i$ . Note that the more robust the timing signatures in the data, the larger the uncertainty  $\Delta$  at which  $\hat{\tau}_{i,IBC}$  and  $\hat{\tau}_{i,SIGACT}$  start to only represent *iid* random samples drawn from an exponential probability distribution. The functional dependence of  $F$  on  $\Delta$  is thus a quantitative measure for the robustness of the timing signatures. In particular, we will identify the critical value of  $\Delta_c$  for which we can be more than 95% certain, i.e.,  $F < 0.05$ , that uncertainties in timestamps do not destroy the non-trivial signature in  $\hat{\tau}_{i,IBC}$  and  $\hat{\tau}_{i,SIGACT}$ .

Figure 9 shows the  $p$ -values of the KS-test and the fraction  $F$  as a function of the value of  $\Delta$  for the time window October 15, 2006 to February 15, 2007 - a period specifically chosen to reflect a situation where both full datasets show non-trivial timing signatures, but where for larger thresholds this signature breaks down in IBC. For both IBC and SIGACT the figure clearly demonstrates that the non-trivial timing distributions in the full datasets are quite robust to uncertainties in timestamps with  $\Delta_{c,IBC} \simeq 3$  days and  $\Delta_{c,SIGACT} \simeq 2$  days respectively (Figure 9(a)). Notice, too, that the transition to Poissonian dynamics for increasing  $\Delta$  is continuous and relatively slow. At uncertainties of about 5 days (IBC) and 4 days (SIGACT) 50% of the reshuffled datasets are indistinguishable from featureless data. Note that we also analyzed events with 3 or more casualties (Figure 9(b)). Here IBC clearly



does not feature robust non-trivial timing signature since already at the minimal uncertainty of one day  $F$  is close to 1. For SIGACT we do observe a non-trivial signature and  $\Delta_{c,SIGACT} \simeq 2$  suggests that this signature is similarly robust as that observed for the full dataset.



Our analysis thus suggests that - where they exist - the non-trivial timing signatures for the full IBC and SIGACT data are indeed quite robust against uncertainty of timestamps. In fact, the signatures are robust enough that even if event timing may have been miscoded by up to 2 days, we could still expect to see non-trivial timing dynamics. Note that this does, of course, not imply that timestamp uncertainties of up to 2 days would not affect the inferences we draw from day-by-day and even distributional comparison - it only suggests that some timing information will be preserved.

#### 4 Discussion and conclusion

In this study we systematically identified a number of key quantitative differences between the event reporting in media-based IBC data and field report-based SIGACT military data. In fact, we find significant differences in reporting at all levels of analysis: aggregate, monthly, distributional and day-by-day comparisons. These relative biases are consistent with a number of structural differences of the reporting in IBC and SIGACT. We further showed that even for subsets of events where both datasets were found to be most consistent at an aggregate level, the daily time series of events were significantly different. Overall this suggests that at any level of analysis the specific choice of dataset may have a critical impact on the quantitative inferences we draw - at the extreme using IBC or SIGACT data might, in fact, lead to substantially different results.

In an individual analysis of each dataset we further showed that SIGACT and IBC differ markedly with regard to their usefulness for event timing analyses - a key application for both datasets. In fact, IBC was found to have only trivial timing signatures, i.e., signatures indistinguishable from an *iid* random process, for much of the time period analyzed. In comparison SIGACT codes much more non-trivial timing dynamics and is thus generally more suitable for the analysis of event timing. In the low intensity conflict phases prior to 2006 and after mid-2008, however, even SIGACT generally does not feature non-trivial timing dynamics. This strongly suggests that any analysis of event timing and causal relationships between events using SIGACT should best be restricted to the period 2006 to 2008. Our analysis, however, also confirmed that where non-trivial timing signatures for the full datasets exist these signatures are quite robust against uncertainties in timestamps of events.

In order not to be systematically affected by geographically biased coverage, our quantitative analysis focused exclusively on the case of Baghdad. We contend, however, that the relative as well as absolute differences in reporting of IBC and SIGACT extend beyond this 'best case' scenario to all of Iraq. In other words, for the full Iraq datasets reporting differences are at best what we found here, but they are likely even more pronounced due to fundamentally more limited event coverage outside of the greater Baghdad area.

Our findings have a number of concrete implications for recent studies analyzing the conflict in Iraq. First, we would like to re-emphasize that the substantial disagreement between the two datasets suggests that using one or the other will likely yield substantively different results. This applies to studies using IBC data at a distributional [2] or aggregate level [17], but most notably to studies using IBC [3, 18, 19] or SIGACT [8, 20] data at a daily resolution where the differences are most substantial. The lack of simultaneous agreement with regard to number of events and casualty counts per months implies in particular that time series analysis with models that describe both event occurrence and casualties - for instance, models of marked point processes [59] - may lead to substantially

different results depending on which dataset is used, even if focusing on subsets of events of certain minimal sizes.

Second, the absence of non-trivial timing signatures for significant parts of both datasets may pose a substantial problem if data is used for detailed timing (or causal) analysis. In fact, none of the above mentioned studies using either IBC or SIGACT data at a daily resolution confirmed whether they actually feature robust timing signatures. The analyses in [18, 19], for example, employ a Hawkes point process model [53, 54] to study event timing dynamics. However, our analysis suggests that the IBC data used is almost featureless at short time-scales, having only long-term non-stationary trends for long periods in 2005, 2006 and 2008. It is therefore clearly not suitable for this kind of analysis. Moreover, given the daily resolution of timestamps in IBC and the corresponding clustering of events on a given day, we strongly caution against the direct calibration of a Hawkes model even where robust timing signatures exist, simply because the resulting model fits will be (falsely) rejected by standard goodness-of-fit methods. Instead, it is better to rely on randomization techniques such as those proposed in [58] and used for the timestamp analysis in our study. Note also that the absence of non-trivial timing signatures in SIGACT prior to 2006 and after mid-2008 may affect the inferences regarding causal relationship between events in [8, 20] - this applies particularly for [20] which analyzes event dynamics exclusively in the first six months of 2005.

The growing number of recent contributions addressing issues of bias in conflict event data [13–16] points to an increased awareness for data related issues in conflict research. Our study contributes to this literature by systematically analyzing relative biases in conflict event data and relating them to structural differences in reporting. The sources of systematic bias discussed here are, however, clearly not restricted to conflict data. For researchers using data on other social processes that may be subject to similar biases our analysis suggests two important ‘lessons learned.’ First, the often very substantial differences between the two datasets analyzed here should raise awareness that data bias is not an afterthought but a critical issue worthy of our fullest attention. In particular, if analyses are meant to provide concrete policy advice we must be especially wary that substantive findings do not arise from biased inference. Second, we showed how structural differences in reporting directly translate into relative biases. This demonstrates, that a careful *a priori* understanding of the strength and limitations of a given dataset allows to anticipate possible biases in subsequent analyses - even if there is only one dataset that covers the case in question. If more than one comparable dataset exists one can either directly analyze their relative bias or, at least, perform the same analysis for all datasets to verify that the substantial conclusions drawn are robust and consistent. We also showed that statistical tests may help identify datasets that are more suitable than others for the analysis at hand.

To date most studies using these data unfortunately neither address potential biases nor systematically test the robustness of their findings. There is certainly not one comprehensive strategy to mitigate bias in empirical data but the present study suggests that researchers can at least actively address it. Especially with the growing availability of large and highly-resolved datasets it will be more important than ever that issues of data quality are taken seriously. As the case of the conflict in Iraq shows, if unaccounted for, we otherwise face the risk that the ‘views to a war’ will indeed be driving our substantial findings.

## Additional material

**Additional file 1: Supplementary Information.**  
**Additional file 2: IBC and SIGACT data for the greater Baghdad area.**

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

KD and VF conceived and designed the study, KD prepared the data, VF analyzed the data and KD and VF wrote and approved the final version of the article.

### Author details

<sup>1</sup>Department of Humanities, Social and Political Science, Chair of Sociology, Modeling and Simulation, ETH Zürich, Clausiusstrasse 50, Zürich, 8092, Switzerland. <sup>2</sup>Department of Management, Technology and Economics, Chair of Entrepreneurial Risks, ETH Zürich, Scheuchzerstrasse 7, Zürich, 8092, Switzerland.

### Authors' information

KD is a PhD student in the Department of Humanities, Social and Political Science at ETH Zürich (Switzerland), Chair of Sociology, Modeling and Simulation. In his research he uses detailed, disaggregated empirical violence data and a range of statistical and computational modeling techniques to study micro-level conflict processes. Focusing mainly on asymmetric intra-state conflict he has worked on the Israeli-Palestinian conflict, Jerusalem in particular, and on the conflict in Iraq.

VF is a senior researcher in the Department of Management, Technology and Economics at ETH Zürich (Switzerland), Chair of Entrepreneurial Risks. His research is mainly focused on self-excited point process models for the description of dynamics in complex systems, with a particular interest in financial applications such as modeling market microstructure effects.

### Acknowledgements

We are grateful to Ryohei Hisano, Spencer Wheatley, Didier Sornette, Michael Mäs, Thomas Chadeaux, Sebastian Schutte, Ryan Murphy and Dirk Helbing for fruitful discussions and comments on earlier versions of this article.

### Endnotes

- <sup>a</sup> The estimates of the total fatalities over the course of the Iraq war differ substantially. For a detailed discussion please refer to <http://www.iraqbodycount.org/analysis/beyond/exaggerated-orb/>.
- <sup>b</sup> For reactions by leading conflict researchers to the release of the data see [60], for more general statements regarding their relevance and impact see [61]. We contend that the data can be used in a responsible manner for academic research, given that the empirical analysis does not in any way and under any circumstances harm or endanger individuals, institutions, or any of the political actors involved. Note in particular that all data used here has been intentionally stripped of any detailed information on specific incidents beyond information on timing, severity and location of attacks.
- <sup>c</sup> Details on data format, preparation etc. are provided in Section 1 of Additional file 1. Data used in this study is provided as .csv files for download (see Additional file 2).
- <sup>d</sup> We include all SIGACT events independent of perpetrator identity consistent with the coverage of IBC.
- <sup>e</sup> Events in Baghdad make up about 35% of all events in IBC and 50% in SIGACT suggesting that there is indeed an element of relative geographic reporting bias.
- <sup>f</sup> In our analysis we always rely on the lower bound as it is the most conservative estimate; see Section 3 of Additional file 1 for details and sensitivity analyses.
- <sup>g</sup> In the U.S., for example, the geographic coverage of different providers varies significantly, independent of population density.
- <sup>h</sup> Note that some of the 'missing' small events in IBC might at least be partially accounted for in the aggregated monthly (morgue or hospital) reports that were excluded from our study.
- <sup>i</sup> A previous analysis of the number of events per day in Iraq also used a half year temporal window size [2].
- <sup>j</sup> As a consequence of the nature of the statistical test used here we reject the correct null hypothesis in 5% of the cases by chance and we thus effectively expect to obtain  $F_{\max} = 0.95$  even if the dataset is completely featureless.

Received: 20 December 2013 Accepted: 19 September 2014 Published online: 25 October 2014

### References

1. Clauset A, Young M, Gleditsch KS (2007) On the frequency of severe terrorist events. *J Confl Resolut* 51:58-87
2. Bohorquez JC, Gourley S, Dixon AR, Spagat M, Johnson NF (2009) Common ecology quantifies human insurgency. *Nature* 462(7275):911-914
3. Johnson N, Carran S, Botner J, Fontaine K, Laxague N, Nuetzel P, Turnley J, Tivnan B (2011) Pattern in escalations in insurgent and terrorist activity. *Science* 333(6038):81-84
4. Zammit-Mangion A, Dewar M, Kadiramanathan V, Sanguinetti G (2012) Point process modelling of the Afghan war diary. *Proc Natl Acad Sci USA* 109(31):2414-12419
5. Schutte S, Weidmann N (2011) Diffusion patterns of violence in civil wars. *Polit Geogr* 30(3):143-152

6. Weidmann N, Salehyan I (2013) Violence and ethnic segregation: a computational model applied to Baghdad. *Int Stud Q* 57:52-64
7. Bhavnani R, Miodownik D, Choi HJ (2011) Three two tango: territorial control and selective violence in Israel, the West Bank, and Gaza. *J Confl Resolut* 55:133-158
8. Linke AM, Witmer FD, O'Loughlin J (2012) Space-time granger analysis of the war in Iraq: a study of coalition and insurgent action-reaction. *Int Interact* 38(4):402-425
9. Bhavnani R, Donnay K, Miodownik D, Mor M, Helbing D (2014) Group segregation and urban violence. *Am J Polit Sci* 58:226-245
10. Raleigh C, Linke A, Hegre H (2010) Introducing ACLED: an armed conflict location and event dataset. *J Peace Res* 47(5):651-660
11. Sundberg R, Lindgren M, Padsokocimaite A (2010) UCDP GED codebook version 1.5-2011. Available online at <http://www.ucdp.uu.se/ged/>
12. Lyall J (2010) Are coethnics more effective counterinsurgents? Evidence from the second Chechen war. *Am Polit Sci Rev* 104:1-20
13. Eck K (2012) In data we trust? A comparison of UCDP GED and ACLED conflict event datasets. *Coop Confl* 47:124-141
14. Chojnacki S, Ickler C, Spies M, Wiesel J (2012) Event data on armed conflict and security: new perspectives, old challenges, and some solutions. *Int Interact* 38:382-401
15. Raleigh C (2012) Violence against civilians: a disaggregated analysis. *Int Interact* 38(4):462-481
16. Weidmann NB (2013) The higher the better? The limits of analytical resolution in conflict event datasets. *Coop Confl* 48(4):567-576
17. Condra LN, Shapiro JN (2012) Who takes the blame? The strategic effects of collateral damage. *Am J Polit Sci* 56:167-187
18. Lewis E, Mohler GO, Brantingham PJ, Bertozzi AL (2012) Self-exciting point process models of civilian deaths in Iraq. *Secur J* 25:244-264
19. Lewis E, Mohler GO (2011) A nonparametric EM algorithm for multiscale Hawkes processes. Preprint
20. Braithwaite A, Johnson SD (2012) Space-time modeling of insurgency and counterinsurgency in Iraq. *J Quant Criminol* 28:31-48
21. Chadeaux T (2014) Early warning signals for war in the news. *J Peace Res* 51:5-18
22. Golder SA, Macy MW (2011) Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* 333(6051):1878-1881
23. IBC (2014) Iraq body count. <http://www.iraqbodycount.org/>
24. Kagan K (2009) *The surge: a military history*. Encounter Books, New York
25. Petraeus D (2006) Learning counterinsurgency: observations from soldiering in Iraq. *Mil Rev* 86:2
26. Petraeus D (2010) Counterinsurgency concepts: what we learned in Iraq. *Global Policy* 1:116-117
27. Rogers S (2010) Wikileaks Iraq: data journalism maps every death. <http://www.theguardian.com/news/datablog/2010/oct/23/wikileaks-iraq-data-journalism> (accessed: 09/03/2013)
28. Berman E, Shapiro JN, Felter JH (2011) Can hearts and minds be bought? The economics of counterinsurgency in Iraq. *J Polit Econ* 119(4):766-819
29. Rogers S (2010) Wikileaks Iraq: what's wrong with the data? <http://www.theguardian.com/news/datablog/2010/oct/25/wikileaks-iraq-data/> (accessed: 08/07/2013)
30. Earl J, Martin A, McCarthy JD, Soule SA (2004) The use of newspaper data in the study of collective action. *Annu Rev Sociol* 30:65-80
31. Oliver PE, Maney GM (2000) Political processes and local newspaper coverage of protest events: from selection bias to triadic interactions. *Am J Sociol* 106(2):463-505
32. McCarthy JD, McPhail C, Smith J (1996) Images of protest: dimensions of selection bias in media coverage of Washington demonstrations, 1982 and 1991. *Am Sociol Rev* 61(3):478-499
33. Davenport C, Ball P (2002) Views to a kill: exploring the implications of source selection in the case of Guatemala 1977-1995. *J Confl Resolut* 64(2):427-450
34. DoS (2009) *U.S. government counterinsurgency guide*. U.S. Department of State
35. González MC, Hidalgo C, Barabási AL (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779-782
36. Leetaru KH, Wang S, Cao G, Padmanabhan A, Shook E (2013) Mapping the global Twitter heartbeat: the geography of Twitter. *First Monday* 2013:18(5-6). <http://firstmonday.org/article/view/4366/3654>
37. PewResearch (2013) 72% of online adults are social networking site users. <http://www.pewinternet.org/2013/08/05/72-of-online-adults-are-social-networking-site-users/> (accessed: 06/26/2014)
38. Samoilenko A, Yasseri T (2014) The distorted mirror of Wikipedia: a quantitative analysis of Wikipedia coverage of academics. *EPJ Data Sci* 3:1
39. Thelwall M, Haustein S, Larivière V, Sugimoto CR (2013) Do altmetrics work? Twitter and ten other social web services. *PLoS ONE* 8(5):64841
40. Abramowitz M, Stegun IA (eds) (1965) *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Dover Publications, New York
41. Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. *SIAM Rev* 51(4):661
42. Richardson LF (1948) Variation of the frequency of fatal quarrels with magnitude. *J Am Stat Assoc* 43:523-546
43. Lars-Erik Cederman CW, Sornette D (2011) Testing Clausewitz: nationalism, mass mobilization, and the severity of war. *Int Organ* 65(4):605-638
44. Maillart T, Sornette D (2010) Heavy-tailed distribution of cyber-risks. *Eur Phys J B* 75(3):357-364
45. Pettitt AN (1976) A two-sample Anderson-Darling rank statistic. *Biometrika* 63:161-168
46. Scholz FW, Stephens MA (1987) K-sample Anderson-Darling tests. *J Am Stat Assoc* 82(399):918-924
47. Frederick J (2006) *Statistical methods in experimental physics*, 2nd edn. World Scientific, Singapore
48. Engle RF, Granger CWJ (1987) Co-integration and error correction: representation, estimation, and testing. *Econometrica* 55(2):251-276

49. Dickey DA, Fuller WA (1979) Distribution of the estimators for autoregressive time series with a unit root. *J Am Stat Assoc* 74(366):427-431
50. Said SE, Dickey DA (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 71(3):599-607
51. Carpenter D, Fuller T, Roberts L (2013) WikiLeaks and Iraq body count: the sum of parts may not add up to the whole - a comparison of two tallies of Iraqi civilian deaths. *Prehosp Disaster Med* 28(3):223-229
52. Haushofer J, Biletzki A, Kanwisher N (2010) Both sides retaliate in the Israeli-Palestinian conflict. *Proc Natl Acad Sci USA* 107(42):17927-17932
53. Hawkes AG (1971) Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58:83-90
54. Hawkes AG (1971) Point spectra of some mutually exciting point processes. *J R Stat Soc, Ser B, Methodol* 33(3):438-443
55. Engle RF, Russell JR (1997) Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model. *J Empir Finance* 4(2-3):187-212
56. Engle RF, Russell JR (1998) Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66(5):1127-1162
57. Russell JR (1999) Econometric modeling of multivariate irregularly-spaced high-frequency data. University of Chicago. Working paper
58. Filimonov V, Sornette D (2012) Quantifying reflexivity in financial markets: toward a prediction of flash crashes. *Phys Rev E* 85(5):056108
59. Daley DJ, Vere-Jones D (2008) An introduction to the theory of point processes. volume II: general theory and structure. *Probability and its applications*, vol 2, 2nd edn. Springer, Berlin
60. Bohannon J (2010) Leaked documents provide bonanza for researchers. *Science* 330(6004):575
61. The Guardian (2010) Iraq war logs: experts' views.  
<http://www.theguardian.com/commentisfree/2010/oct/28/iraq-war-logs-experts-views/> (accessed: 08/07/2013)

doi:10.1140/epjds/s13688-014-0025-x

**Cite this article as:** Donnay and Filimonov: Views to a war: systematic differences in media and military reporting of the war in Iraq. *EPJ Data Science* 2014 **3**:25.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](http://springeropen.com)

---