Journal of Conflict Resolution

http://jcr.sagepub.com/

First Things First: Assessing Data Quality before Model Quality

Anita Gohdes and Megan Price

Journal of Conflict Resolution 2013 57: 1090 originally published online 15 November 2012
DOI: 10.1177/0022002712459708

Journal of Conflict Resolution
57(6) 1090-1108

DOI: 10.1177/0022002712459708
jcr.sagepub.com

First Things First: Assessing Data Quality before Model Quality

Anita Gohdes[1,2,3] and Megan Price[2]

Abstract

We address weaknesses in the Peace Research Insitute Oslo (PRIO) Battle Deaths Dataset, and as a result draw contradicting conclusions to those presented by Lacina and Gleditsch. Our analysis focuses on the availability of data on battle deaths within specific conflict-years and problems encountered when data from multiple types of sources are combined. We repeat Lacina, Gleditsch, and Russett's analysis of battle deaths over time, with an attempt to provide a more robust model and incorporate an estimate of the uncertainty present in the PRIO Battle Deaths Dataset. This reanalysis reveals that the data used to establish the PRIO Battle Deaths Dataset does not offer a clear answer as to whether battle deaths have decreased or increased since the end of the Second World War. We contend that while the PRIO Battle Deaths Dataset offers the most comprehensive assembly of battle deaths data available to date, it is not suitable for analysis across countries or over time.

Keywords

conflict data, battle deaths, selection bias.

[1]University of Mannheim, Mannheim, Germany
[2]Benetech Human Rights Data Analysis Group, Palo Alto, CA, USA
[3]Centre for the Study of Civil War, Peace Research Insitute Oslo, Oslo, Norway

Corresponding Author:
Megan Price, Benetech Human Rights Data Analysis Group, 480 S. California Ave., Suite 201, Palo Alto, CA 94306, USA.
Email: meganp@benetech.org

Introduction

In 2005, Lacina and Gleditsch introduced a new data set on global numbers of battle deaths, designed to address the "inappropriate use of incommensurate conflict statistics" in previous projects that led to "misleading impressions about patterns in global warfare" (Lacina and Gleditsch 2005, 145).[1] In presenting this new data set, the authors criticize the application of inconsistent definitions of the measurement of warfare and propose a definition of conflict severity that only includes deaths resulting from combat situations. In line with this new definition, the Peace Research Insitute Oslo (PRIO) Battle Deaths Dataset combines a multitude of publicly available data types and sources, including convenience samples, probability-based estimations, historic narratives, and expert opinions. The yearly averages and aggregates of this combined data lead the authors to conclude that both the absolute number of deaths in combat (Lacina and Gleditsch 2005) and the relative risk of dying in battle have declined in the last fifty years (Lacina, Gleditsch, and Russett 2006). Numerous publications have followed the use of the PRIO Battle Deaths Dataset as a measure for conflict severity (see Lacina 2006; Melander, Oberg, and Hall 2009; Hoddie and Smith 2009; HSRP 2010). Two recent monographs by Goldstein (2011) and Pinker (2011) have followed the contention that the world is becoming a more peaceful place, going so far as to argue that "war is really going out of style (Goldstein and Pinker 2011)."

Lacina and Gleditsch's conceptual critique of previous definitions used for war fatalities is an example of the increased use of disaggregated measurements of war in conflict research (see also Eck and Hultman 2007; Brück et al. 2010; Raleigh et al. 2010; Bussmann and Schneider 2011). These recently developed individual- and event-level accounts for conflict reflect the significant move from macro-level to micro-level theories and explanations of violence in conflict situations (see Kalyvas 2006; Weinstein 2007; Verwimp, Justino, and Brück 2009; Cederman and Gleditsch 2009; Raleigh and Hegre 2009). We welcome this change of focus, as it has stimulated a much-needed discussion on the variety of ways to count and estimate war fatalities (see Obermeyer, Murray, and Gakidou 2008; Spagat et al. 2009; HSRP 2010; Checchi 2010).[2]

The efforts of the authors to collect the best available information from different sources for the time period since the Second World War are commendable. Creating a database that unifies the publicly available information on battle violence is, in itself, a valuable task. However, contrary to their conclusion, we believe this data set is inappropriate for analyzing global trends in conflict severity and intensity. As noted earlier, numerous publications have relied on this data set to conduct just such analyses. As such we believe it is critical that the data set is examined carefully and that any limitations of the data are discussed fully in the literature.

In this article, we present conclusions that contradict those made by the authors of the PRIO Battle Deaths Dataset and outline why we believe the numbers gathered offer no clear answer to the question of whether battle deaths have in fact decreased, increased, or stayed the same. We identify two different issues that make the data set unsuitable for statistical analysis. These are addressed as problems of data "availability" and problems of data "quality." Regarding the first problem, for many conflicts, especially those that took place before the 1980s, very little data concerning deaths that can be attributed to combat even exist. In some cases, death tolls are only available for entire conflict periods. We find that 53.7 percent of battle death counts included in the entire data set, and 60.2 percent of observations before 1975 are not year specific. To obtain counts for individual years for these cases, conflict totals are simply divided by the number of years in the conflict period.[3] These yearly averages are combined with conflict numbers that document weekly, monthly, or yearly death tolls, where individual identification of the dead might even be available. This mixing of different levels of aggregated and averaged data leaves us with a collection of death tolls for the past half century that are not comparable over time or across different conflicts. We argue that battle deaths data that are not year specific are insufficiently disaggregated to draw conclusions about patterns over time.

The second problem is a paucity of high-quality data sources. We define quality as the extent to which the data included in the PRIO Battle Deaths Dataset can be used to obtain representative information on patterns of battle violence for the conflict situation covered. Using this definition, we distinguish between two different classes of sources included in the PRIO Battle Deaths Dataset: those that offer a statistically reproducible measure of uncertainty and those that are obtained through convenience sampling or based on expert opinions. We argue that the combination of these two classes, that is, probability-based estimations, on one hand, and convenience-based numbers and narratives, on the other hand, prevents an authoritative statement or prediction on global trends.

We recognize the importance of setting clear definitions and categories of different types of deaths and killings and commend the authors' emphasis on categorizing these different types of violations that occur during and in the aftermath of a conflict (see also Hoddie and Smith 2009, 182). Critically examining and working to solve problems such as miscodings, misclassifications, issues of noncomparability, and the inclusion or exclusion of cases is an important step within the research process and vital for the analysis of conflict patterns. It ensures a common vocabulary and understanding of what is being measured and consequently sets out the scope and depth of the conclusions that can be drawn from such research. The authors of the PRIO Battle Deaths Dataset have helped to advance the field of conflict measurement by addressing these problems. Unfortunately, this work does not consider what we believe to be the larger problems of data availability and quality. The same critical eye that is used to examine the definition of conflict severity should be turned on the data to which these solutions are applied. Categories of war deaths are only useful if the information used to fill them is representative, consistent, and comparable across time and conflicts. A distinction between battle deaths and other forms of death that occur in conflicts is only helpful if the reports used as the basis for the measurements offer enough disaggregated information to distinguish between these

Table 1. Summary Statistics for the Battle Deaths Data.

	Best	High	Low
N	1,186	1,957	1,957
Mean	7,175	10,074	2,290
SD	30,190	4,9087	13,165
Min	13	25	0
Max	497,500	763,002	322,348
Missing values	771	0	0

categories. If this is not the case, and the data available are neither sufficient nor representative, conclusions about patterns of violence are likely to be misleading.

The next section of this article gives an in-depth review of the PRIO Battle Deaths Dataset, paying particular attention to the authors' own rating of the data availability as well as the differences between the best, lowest, and highest estimates for each conflict-year observation, over time. The section on Different Conflicts, Different Sources, Different Stories lays out the different challenges the PRIO Battle Deaths Dataset faces with respect to the data sources. Since the types of information sources that underlie this data set are diverse, a selection of the most common types of bias is discussed. In the section Is the Risk of Death in Battle Really Declining? we reanalyze the PRIO Battle Deaths Dataset by paying special attention to identifying trends in the data as well as applying a nonparametric model of uncertainty. The final section concludes with recommendations on how the study of conflict severity could be improved and what statistical techniques are available to reduce biased results of research in this field.

Different Times, Different Data Availability

The PRIO Battle Deaths Dataset contains information on battle deaths in conflicts that occurred between 1946 and 2008. The unit of observation is conflict-year, which means that each conflict and year is listed as an individual observation (e.g., El Salvador 1981). If multiple conflicts occurred in one country, then these are listed separately. Table 1 presents the summary statistics for the newest version of the PRIO Battle Deaths Dataset. For each observation, the PRIO Battle Deaths Dataset lists a best estimate for the number of battle deaths, as well a low and a high estimate (see Figure 1).

The most noteworthy information that the summary statistics in Table 1 provides is of the 1,957 observations, only 1,186 have a best estimate. This means that 39.4 percent of all conflict-year units lack a best value. The codebook offers an explanation for this: for all cases where no specific numbers are available, the UCDP (Uppsala Conflict Data Program)/PRIO code rules are applied and the upper and lower limits of these rules included as high and low estimates.[4] The coding rules

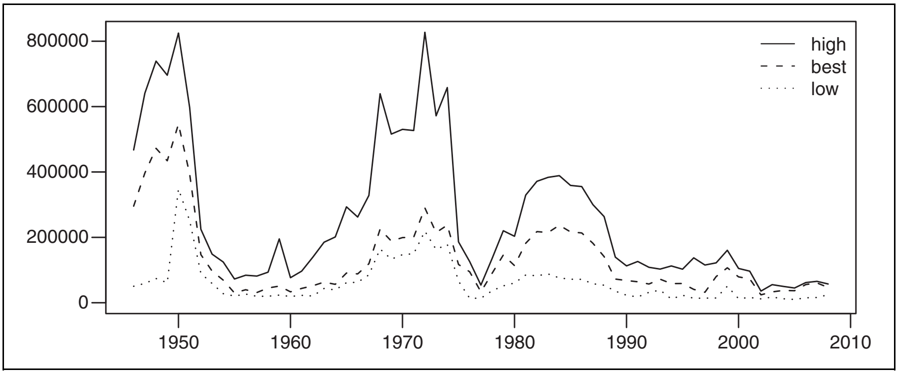

Figure 1. Battle deaths data that include a best estimate.

of the UCDP/PRIO divide all conflict-year observations into two categories: years where the number of battle deaths are estimated to be between 25 and 999, and years where the number is estimated to be between 1,000 and 9,999. For all the conflict-years lacking further information, there exists no best estimate and the low estimate is set either at 25 or at 1,000, and the high estimate is set either at 999 or at 9,999. The summary statistics reveal the differences between the three estimates: the best estimate reports a mean number of deaths of 7,175, the high estimate is 10,074, and the low estimate is 2,290.[5]

The best, high, and low estimates provided by the data set are numbers of battle dead that are found in the literature and that have been judged to be reasonable. They should not be mistaken for actual statistical estimates. Statistical estimates are based on probability theory, which enables the calculation of a confidence interval or other quantification of the uncertainty of the presented number. Since the numbers combined in the data set originate from a multitude of sources, we argue that the high and low estimates presented here are equally probable. We could, for example, take a minimalistic approach toward the assessment of all conflicts and assume that only the low estimates, indicated by the dotted line in Figure 1, are a reflection of the progression of global battle deaths over time. Following this assumption, we might infer that the number of deaths due to battle situations has continually decreased since the 1980s, with the exception of a slight surge in the late 1990s. Alternatively, if we decide to select the low estimates for the time prior to 1980 and select the high estimates for the most recent conflict-years recorded, we could reach the conclusion that the number of battle deaths have increased over time. The conclusions to be drawn on the risk of battle deaths throughout the last half century depend largely on which numbers are judged to be most "reliable" for each conflict-year.

As mentioned in the Introduction, the availability of specific information on death tolls for individual observations is a determining factor for the PRIO Battle Deaths Dataset's statistical usefulness. Therefore, the question arises as to what extent there exists quantifiable evidence of battle violence for each individual conflict. Since the

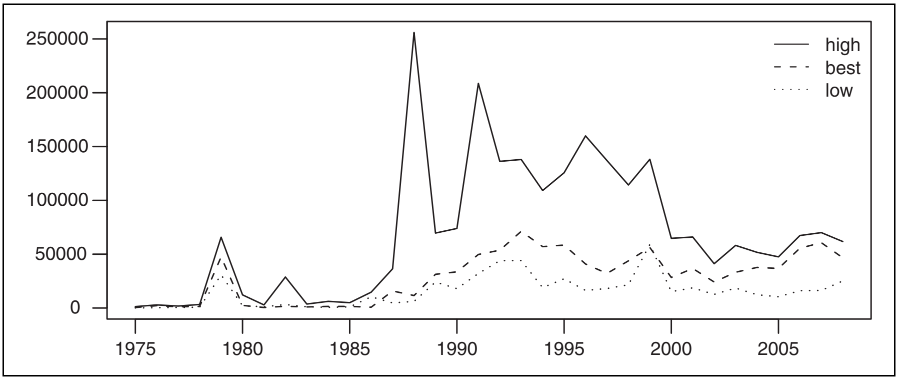

Figure 2. Battle deaths data that include yearly numbers and a best estimate, 1975–2008.

main interest of the PRIO Battle Deaths Dataset is stated to be the monitoring of trends (Lacina and Gleditsch 2005, 153), it seems appropriate to expect the information reflected in it to at least be disaggregated to yearly numbers. Ideally, each observation in the data set should thus offer a low, best, and high estimate of battle deaths that is specific to a single conflict in a single year. For example, for the Persian Gulf War in Iraq in 1991, the PRIO Battle Deaths Dataset gives a disaggregated best estimate of 28,245, a low estimate of 1,545, and a high estimate of 43,245 killed in combat situations. In contrast, for the Cambodian Civil War, the numbers of battle dead for the period from 1980 to 1986 are not year-specific. Instead the best estimate for this entire seven-year period is 48,800 and the high estimate is 70,000. The low estimate is 6,300, which is "trended per UCDP/PRIO coding rules."[6]

In addition to listing the best, low, and high numbers for a conflict-year, the data are also classified according to three different levels of "availability."[7] The first category of data offers an individually obtained number/estimate for each conflict-year unit. The example of the Gulf War in Iraq given earlier would fall into this category. The observations that fall into the first category represent the most detailed and informative data. Of all cases captured in the data set, 46.3 percent have estimates that rely on conflict-year specific information. Figure 2 shows the time trend of the data that falls into this first category. Since most of the observations for the earlier years recorded in the data set do not fulfill this criteria, the graph looks at the time trend from the middle of the 1970s onward.

We term the second category "trend," as it refers to data that were obtained by looking at the overall numbers of battle death for a conflict and adjusting the yearly mean number with the help of limited information on the conflict's dynamics. For example, for the civil war in El Salvador, the data available for the period between 1979 and 1999 are limited to a total number of all conflict deaths. However, year-specific numbers are available for the number of government losses only. Since the yearly number of all conflict deaths is assumed to be related to the losses on the

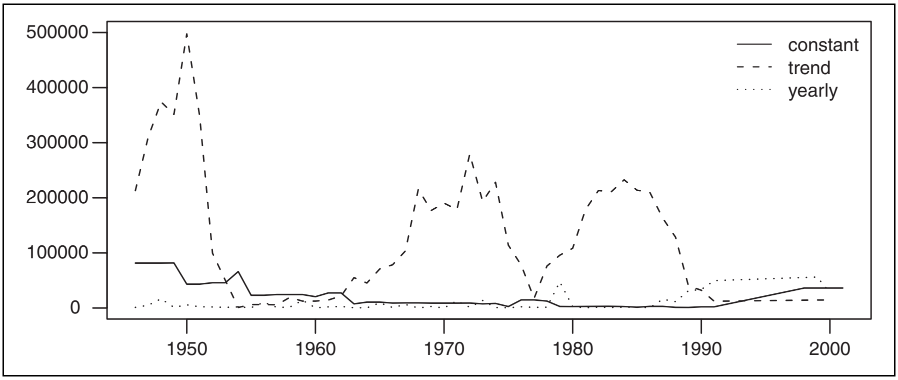

Figure 3. Best estimates distinguished by data availability.

government side, this information is used to adjust the simple average of overall battle death numbers available. For 18.3 percent of all conflict-year units in the data set, such "trend" numbers are available.[8]

The last category of data holds even less information than the previous one. The 35.4 percent of conflict-year units that fall in this category merely include a constant number that is derived by averaging across all battle deaths of the period of the respective conflict. For example, for the seven-year period between 1980 and 1986 of the Cambodian Civil War there only exists an aggregated estimate. The yearly observations for these conflict-year units therefore report this aggregated number—divided by seven.

Figure 3 gives an overview of the data when distinguishing between different types of availability. Turning to the actual number of battle deaths (as opposed to the number of conflict-year units), we see that the majority of battle deaths per year can be attributed to conflict cases that fall into the second category (*trend*), where no yearly numbers are available. Furthermore, we see that up until the late 1980s, more battle deaths are counted in the third category (*constant*), where merely an aggregated count of battle deaths is available for an entire conflict.

In summary, we find that 53.7 percent of the conflict-year units are not year specific, and that 60.2 percent of the units before 1975 do not even include information on conflict-internal trends. This means that although the definition of what is being measured remains constant across all of the data, less than half of the information used in the data set actually contains unit-specific numbers that could do the definition any justice. We conclude, therefore, that less than half of the data set is in fact suitable for the analysis of trends or the monitoring of battle deaths over time. However, the existence of yearly information for 46.3 percent of all observations does not imply that these numbers adequately represent the true patterns of battle violence as they occurred across the last half century. The numbers available for one conflict often differ profoundly between the best, low, and high estimate. Depending on the way in which the information was collected or recorded by the underlying sources, the probability that all three of the recorded numbers are incorrect is high. In the next section, we explore the problems concerning data quality and the challenge of quantifying uncertainty in counts of battle deaths.

Different Conflicts, Different Sources, Different Stories

In this section, we examine the different types of sources that are used to compile the PRIO Battle Deaths Dataset and present specific issues that arise when gathering and combining data on conflict deaths. The authors draw from a large variety of information and have put every effort into combining all possible types of data that were available to them:

> [The] dataset draws on leading compendia of casualty statistics [. . .], on conflict monitoring projects [. . .], on the annual tables of major armed conflicts in the SIPRI Yearbook[. . .], as well as consultations with regional experts [. . . and] were augmented with studies of individual cases [. . .] archival materials from government sources [. . .] media sources and published studies based on compiled media data [. . .]; and original demographic and epidemiological work where it was available. (Lacina and Gleditsch 2005, 153)

Corresponding to the way we classified the different levels of availability of data, we classify the sources with respect to their quality. We define the quality of data by the extent to which it represents or fails to represent actual levels of violence within individual conflicts. Unfortunately, we are not able to do this in a comprehensive way as we did in the previous section. Since the sources used are so diverse and even vary within conflicts and individual yearly observations, our discussion of the data sources and their potential inaccuracies is kept at a more general level, with case examples serving as demonstrations.

We distinguish between two types of data "classes." The first class of data constitutes numbers that offer statistically reproducible measures of uncertainty, such as capture–recapture estimation techniques[9] or probability-based surveys. The second class refers to numbers that are not statistical estimates and are instead obtained from convenience samples and numbers quoting expert opinions.

Capture–recapture techniques and probability-based surveys describe statistical methods that entail scientifically reproducible measures of uncertainty. For example, death tolls of conflicts in Bosnia (Zwierzchowski and Tabeau 2010), Colombia (Lum et al. 2010), and Darfur (Degomme and Guha-Sapir 2010) have been estimated using such procedures. These numbers are usually produced using a transparent method and are therefore suitable for cross-checking and reproduction by fellow researchers.

The mere use of statistically reproducible estimation techniques does not guarantee accurate measurements of conflict severity. As Spagat and coauthors have shown (see Spagat et al. 2009; HSRP 2010), survey sampling methods, such as those used in Obermeyer, Murray, and Gakidou (2008) and by the International Rescue Committee (IRC) in the Democratic Republic of the Congo,[10] can seriously mismeasure death tolls (Burnham et al. 2006; Coghlan et al. 2006). It is, however, only because of the reproducibility of these estimates that they are even up for debate. As we will see, it is virtually impossible to evaluate conflict statistics that are based on convenience samples or expert opinions, as no quantification of (possible) errors exists. In the remainder of this section, we thus turn our attention to these nonprobability based data collection efforts and present potential sources of bias.

Convenience samples refer to data that are gathered without an underlying random selection process and instead record available, observable information. Unlike the name might imply, these samples are not necessarily the result of convenient data collection efforts. Equally, this term does not imply that the organization or institution leading the documentation process conducted their work in an unsystematic way. Even projects that set out to systematically collect information, no matter how well planned and organized, still often end up with data that does not represent the actual pattern of violence as it occurred during the conflict of interest. If the witnesses to be interviewed, the primary or secondary sources (such as press releases) to be reworked, or the graveyards to be analyzed are not determined with the help of a random process, the chances of obtaining unrepresentative results are almost unavoidable.

Any single data set that was established in a nonrandom way is thus unrepresentative. To make matters worse, it is unrepresentative in an unknown way and may include different proportions of the universe of conflict-related deaths for different time periods or geographic areas (Gohdes 2010; Krüger et al. forthcoming). First, the institution collecting the information can be the source of bias. Sometimes, the information gathered by an organization, such as a media outlet, might never have been intended to be used for statistical analysis or for drawing conclusions about trends in global combat (Davenport 2010; Earl et al. 2004; Smith et al. 2001). If multiple sources exist for the same conflict, it is likely that each institution will have its own definition of cases to be included in its respective collection efforts. This can pose a serious problem when synthesizing this information, and even a clear definition is of little use when the cases recorded cannot be classified accordingly. Furthermore, not all battle dead have the same probability of being counted. People killed in battle cannot report their own loss of life, and data collection projects will necessarily have to rely on other individuals or institutions to report them. The question of visibility is therefore crucial: the higher the visibility of the act of violence, the more likely it will be reported.

The intensity of the conflict is frequently a cause for inaccurate battle death numbers, since documenting violence can be a dangerous task and an increase in violence can inversely lead to a decrease in reported violence (Davenport and Ball 2002). Furthermore, the severity of a conflict usually varies by region (Buhaug and Gleditsch 2008), which means that regionally concentrated information on battle deaths is not representative of an entire country. If the only documentation available on battle deaths focuses on violent incidences that occurred in the capital city, extrapolations for other parts of the conflict zone are ultimately going to be biased (Kalyvas 2004).

The list of potential causes for bias in convenience samples is much longer than the selection presented here. The inaccuracies vary both between and within conflicts, and the variety of previously described conditions can lead to both undercounting and overcounting. Additionally there is no way of knowing just how inaccurate a single convenience sample is—multiple sources and inferential methods are needed to produce measures of uncertainty.

Finally, for many conflicts, the only available data are informed statements made by experts. Optimistically, these experts have detailed and extensive knowledge about conflict developments and dynamics. Such information provided by experts is extremely valuable for qualitative analyses and in-depth historical narratives that seek to understand causes and consequences of individual wars and conflict situations. Furthermore, expert opinions can provide an important contribution to quantitative models when used in the form of priors for Bayesian models. Here, region- or conflict-specific knowledge can be a useful tool for building more realistic and efficient models (see Prelec 2004; Gill and Walker 2005). Including expert statements in databases that are intended to serve the purpose of quantitative analysis is, however, the equivalent of a statistical "guess" and therefore not compatible with the idea of scientific reproducibility and quantifiable uncertainty.

Is the Risk of Death in Battle Really Declining?

In the previous two sections, we argued that the data combined in the PRIO Battle Deaths Dataset does not adhere to the standards of quality and availability that would be needed to offer any conclusive information on trends of deaths in combat over time. Before we reanalyze the data with a nonparametric measure of uncertainty, we believe it is helpful to take a look at what the data can actually tell us about the changing risk of dying in battle.

Model I in Table 2 replicates the model proposed by Lacina, Gleditsch, and Russett (2006), but includes heteroskedasticity robust standard errors. As an indicator for the risk of dying in combat, we divide the yearly aggregates by the yearly lagged global population numbers taken from Gleditsch (2002), as Lacina, Gleditsch, and Russett (2006) do. In examining the predicted values of this model in Figure 4, it seems as if a nonlinear model might yield more explanatory power for this data. In order to stay within a comparable linear framework, we conduct a second- and third-degree polynomial trend analysis (models II and III in Table 2) and find that the third-degree polynomial produces the best fit for the data.[11] Interestingly, Model II including the squared time variable predicts an inverted U-shaped relationship, which would indicate that the risk of dying in battle reached its lowest point during the 1980s and is now increasing again.

Table 2. Reanalysis of Lacina, Gleditsch, and Russett (2006).

	I	II	III	IV
time	−2.02e−06	−2.29e−04	−9.07e−02	−8.82e−07
time²	(5.29e−07)***	(1.27e−04)	(2.55e−02)***	(4.59e−07)
		5.75e−08	4.60e−05	
time³		(3.22e−08)	(1.29e−05)***	
			−7.76e−09	
			(2.18e−09)***	
dummy				1.131e−04
				(3.548e−05)**
constant	4.03e−03	2.28e−01	5.97e+01	1.78e−03
	(1.05e−03)	(1.26e−01)	(1.68e+01)	(9.1e−04)

Note. Indicator for period between 1946 and 1950. Robust standard errors in parentheses.
***Sig: at 0.1%. **Sig: at 1%. *Sig: at 5%.

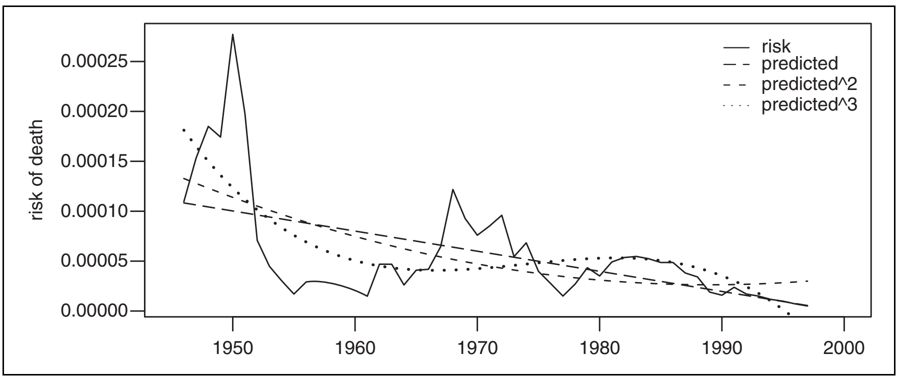

Figure 4. Predicted trends of polynomial analyses.

The improved fit achieved in the polynomial models demonstrates that a steady decline in the risk of dying in battle is not necessarily the obvious conclusion of analyzing this data. Adding to this, the linear framework suffers from a sensitivity toward outliers. The predicted values in Figure 4 give reason to assume that the direction of this relationship is principally being driven by battle deaths that occurred within the first five years following World War II. A closer look at the data reveals that the highest number of battle deaths occur in 1950, where over 69,000 combat deaths are recorded in the data set. In order to test this assumption, Model IV includes an indicator variable for this first five years of the period to be analyzed. As reported in Table 2, the indicator is significant, and the time variable is rendered insignificant. Controlling for the first five years of the period under analysis, there is no indication that the risk of dying in battle has actually decreased.

Table 3. Mean Results Using Permuted Data.

Mean Coef.	*SD*	*p* >.05	*p* < .05
−7.6e−07	1.9e−07	79.1%	20.9%

Note. Reanalysis using the model of Lacina, Gleditsch, and Russett, 2006 (1951–97). One thousand permutations used. Computed using robust standard errors.

We contend that this result might be driven by the data, as neither Lacina, Gleditsch, and Russett (2006) nor our analysis uses the lowest and highest estimate to check the robustness of the estimates. In a final step, we attempt to include a measure of uncertainty in our analysis. As argued earlier, we cannot quantify the likelihood that the low, best, or high numbers presented in the PRIO Battle Deaths Dataset are closer to the true number of battle deaths in a given conflict-year unit. For this reason, we choose a nonparametric model to test whether the uncertainty of estimates between the three values we have for each conflict-year has an impact on the conclusions we wish to draw from the data. We simulate 1,000 permuted samples of the original PRIO Battle Deaths Dataset. The samples have the same number and configuration of conflict-year units, with each observation either taking the best, low, or high estimate of the original data set.

We use each sample to test the hypothesis presented by Lacina, Gleditsch, and Russett (2006) of whether the risk of dying in battle has in fact decreased or not. We fit a model to each of the 1,000 permuted data sets with yearly aggregates divided by yearly lagged global population numbers as the outcome, and time as the regressor. Since we want to check whether the assertion that the model is being driven by the first five years under analysis is robust, we rerun the model for the period from 1950 to 1997. Table 3 reports the mean and standard deviation of the 1,000 coefficients. Further, we report the percentage of regressions where the coefficient is significant at the 5 percent level ($p < .05$) and the percentage where the regression model is not significant ($p > .05$). The mean coefficient of the 1,000 models with permuted data is negative, which would in fact indicate, that the number of battle deaths per capita has, even if only minimally, decreased. However, of the 1,000 regressions, 79.1 percent of all coefficients are not significant at the 5 percent level. The results of the original model are principally driven by the first five years under analysis.

From this reanalysis we conclude that there is no evidence for a significant decline in the risk of dying in battle. The polynomial analysis suggests that a strictly linear relationship provides a poor fit for the data. Controlling for the first postwar period, we find no significant relationship at all, even when including a measure of uncertainty into the analysis. Nevertheless, it is important that we can equally not conclude that there has been an increase in the severity of war. The only conclusion we draw is that the existing data on battle deaths cannot answer questions of trends over time. We contend that there also exists the possibility that the number of battle deaths is related to the frequency of wars through a power law distribution (see Richardson 1948). However, given the inaccuracy of the data and the fact that the detection of such relationships is fraught with difficulties, often leading to inaccurate parameter estimations (see Clauset et al. 2009, 1), we refrain from further discussing this possibility here.

Where Do We Go from Here?

In this article, we have attempted to demonstrate some of the problems that arise when documenting and analyzing violence. Collecting information on battle deaths data for individual conflicts is a challenging task with ample opportunities for misrepresentation of actual patterns of violence. When these individual numbers and estimates are combined from many different sources, it is virtually impossible to gain an accurate picture of battle deaths. Combining sources for half a century's worth of conflicts across the world, where more than half of the data does not even include a measure of time, is daunting and will most likely render the data unsuitable for quantitative analysis.

In examining the different types of sources that exist, we find that even in the best-case scenarios, where multiple independent organizations collected disaggregated and detailed information on battle violence for a complete conflict period, the potential biases and opportunities for misinterpretation are inherently diverse. Political, economic, social, ethnic, religious, geographical, periodical, and technological factors all influence the documentation and data collection efforts of organizations, such as nongovernmental organizations, truth commissions, or international institutions. These overcounts and undercounts vary both within and across conflicts, and it is impossible to account for them on an aggregated scale. Equally, expert statements drawn from qualitative observations of conflict dynamics cannot provide evidence-based facts that can be used for quantitative analysis and inference. The problems are thus case-specific and have to be assessed and addressed individually.

The need to address these issues on a case-by-case basis does not devalue the role of statistical analysis in the study of conflicts. On the contrary, examining conflict patterns in an evidence-based way is vital for our understanding of how violence is exercised (see, e.g., Blattman and Miguel 2010). However, we need to keep in mind that the quality of information that is available for analysis prescribes the methodological options we have and thus ultimately determines the scope of the conclusions that can be drawn from it. The final part of this article gives a brief overview of the way in which the data quality prescribes the method.

Ideally, we would like our data to include every single person who was killed in a combat situation for the conflict and time period we are interested in. The closest we can get to this is by having preconflict and postconflict census data, including information on migration flows that would allow us to exactly trace the individual fate of each person affected in the conflict region. It is clear that this kind of information is hardly ever available.[12] First, war zones and postconflict situations usually bring with them chaos and a lack of infrastructure, both of which seldom allow the conduct of censuses that attempt to systematically collect data on the entire population. Second, even if preconflict and postconflict census data are available, the time between the two might be so far apart that the information to be retrieved is only a rough estimate for larger periods of time.

In conflict or postconflict situations where little data are available but researchers have sufficient resources, the use of surveys to retrospectively assess the number of people killed in combat situations can provide unbiased and comprehensive information on the universe of incidences (for an overview, see Brück et al. 2010). When well prepared, executed, and evaluated, surveys generate high-quality data that is ideal for further quantitative analysis of causes and consequences of conflicts. The quality of the data retrieved is, however, dependent on the resources, available capacity, and the infrastructural conditions in the field that might or might not be conducive for the implementation of a survey. If the scope of the survey only includes parts of a conflict territory, then the data generated can only make partial statements about the conflict.

Sweeping generalizations only undermine the general credibility of the information that can actually be evaluated from the survey and discredit other, well-conducted survey projects.[13] In analyzing survey data, it is thus important for researchers to be modest in the formulations of their conclusions.

If more than one institution documents incidences of violence throughout a conflict, researchers might be able to use the different sources to estimate the total population of people killed in combat. If the information collected is detailed enough to match individuals between the different sources, the overlap structure between the sources can be used to extrapolate the missing number of battle deaths that were not recorded in any of the sources. Projects in Bosnia and Herzegovina, Guatemala, Kosovo, Peru, Colombia, and East Timor (Zwierzchowski and Tabeau 2010; Ball et al. 2002, 2003; Lum et al. 2010; Silva and Ball 2008) have used multiple recapture techniques to calculate such totals (Agresti 1994; Bishop, Fienberg, and Holland 1975). Depending on the availability of data, in terms of overlap between the sources, and the richness of the data, in terms of victim information, the results of this analysis will be more or less informative. For example, if a lot of data is available that includes information on the year a person was killed and whether the person was male or female, capture–recapture models can be used to estimate the unreported killings of both men and women for each year of the conflict. If no information on victim characteristics and time of death is available, then such multiple recapture models will only be able to estimate overall conflict numbers. This estimation method clearly demonstrates how the conclusions drawn are dependent on the quality of data that is used in the analysis.

Finally, for some conflicts it is impossible to obtain information that is suitable for quantitative analysis, be it because the conflict dates too far back or because neither the resources nor the infrastructure is in place to conduct fact-finding projects. In these cases, we need to clearly and modestly demonstrate that the lack of data does not support inferences about patterns of violence. Communicating that the boundaries of our research and conclusions are dependent on the data we use is vital if we want the analyses conducted in many conflict regions throughout the world to be rewarded with more credibility in the future. This credibility rests on the defensibility and legitimacy of our methods and conclusions. Limitations in data must be translated into limitations in conclusions.

Acknowledgments

We are grateful to Patrick Ball, Jule Krüger, Gerald Schneider, Christian Davenport, Johannes Kleibl, and the anonymous JCR reviewers, whose comments helped to substantially improve this article. We would like to thank the authors of the PRIO Battle Deaths Dataset for their generous comments and help in improving this article. This article benefitted greatly from discussions with the team of the Benetech Human Rights Data Analysis Group as well as the workshops on improving the analysis of political violence at Yale and Notre Dame. All errors are, of course, our own.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Notes

1. All data and code necessary to reproduce the Results, tables, and figures presented here can be found in the replication materials available at http://www.sowi.uni-mannheim.de/lspol4/. The PRIO Battle Deaths Data set can be accessed via http://www.prio.no/CSCW/Datasets/Armed-Conflict/ Battle-Deaths/The-Battle-Deaths-Dataset-version-30/. For the descriptive analysis of the data, we used the newest version available at the time of submission. For the replication, we used the version that was used by Lacina, Gleditsch, and Russett (2006).
2. This analysis focuses solely on the Battle Deaths Data, used by Spagat et al. (2009), and thus does not explicitly deal with the controversy between Obermeyer, Murray, and Gakidou (2008) and Spagat et al. (2009). The benefits and pitfalls of survey sampling methods are briefly addressed in the section Where Do We Go from Here? but require a separate analysis.
3. For example, for the Cambodian Civil War, the number of battle dead for the period from 1980 to 1986 is assumed to lie between 6,300 and 70,000, with a "best estimate" of 48,800 (see PRIO Documentation of Coding Decisions, 102). To obtain a yearly low, high, and best "estimate," these figures are divided by the number of years they cover—in this case seven years.
4. See http://www.prio.no/sptrans/973829835/PRIObd3.0_codebook.pdf.

5. Since over one-third of the observations are missing a best value, it is not clear whether the best estimate is in fact generally closer to the higher value than to the lower value. Figure 1 gives a complete picture of the relationship between the three estimates over time.
6. See http://www.prio.no/sptrans/1555324504/PRIObd3.0_documentation.pdf, page 102.
7. The variable *annualdata*, which is included in the data set, offers this information.
8. The dashed line in Figure 3 gives the time trend for this subset. The aggregated data by year in this category includes peaks in 1950 and at the beginning of the 1970s. An increase can be seen in the 1980s. The high numbers in 1950 can be attributed to the Korean War and the numbers in the 1970s result mostly from conflicts in Vietnam and Cambodia. The increase in the 1980s is a result of a multitude of conflicts where only limited data are available on the conflicts' dynamics.
9. Also commonly referred to as multiple systems estimation, tag-recapture, and other names depending on the area of application.
10. See http://www.theirc.org/special-reports/congo-forgotten-crisis.
11. We thank an anonymous reviewer for this suggestion.
12. Zwierzchowski and Tabeau approach this ideal with their use of the 1991 Population Census to match lists of war-related deaths in Bosnia and Herzegovina (Zwierzchowski and Tabeau 2010).
13. The controversy surrounding the IRC-led surveys in the Democratic Republic of Congo is a good example of this. See http://www.theirc.org/special-reports/congo-forgotten-crisis, http://www.humansecurityreport.info/2009Report/2009Report_Complete.pdf, and http://www.hrdag.org/resources/blog_2010-01-21.shtml for an overview.

References

Agresti, Alan. 1994. "Simple Capture-Recapture Models Permitting Unequal Catchability and Variable Sampling Effort." *Biometrics* 50 (2): 494–500. http://www.jstor.org/stable/2533391.

Ball, Patrick, Jana Asher, David Sulmont, and Daniel Manrique. 2003. *How Many Peruvians Have Died? An Estimate of the Total Number of Victims Killed or Disappeared in the Armed Internal Conflict Between 1980 and 2000*. Washington, DC: Report to the Peruvian Commission for Truth and Justice (CVR).

Ball, Patrick, Wendy Betts, Fritz Scheuren, Jana Dudukovich, and Jana Asher. 2002. *Killings and Refugee Flow in Kosovo March - June 1999*. Washington, DC: A Report to the International Criminal Tribunal for the Former Yugoslavia.

Bishop, Yvonne M. M., Stephen Fienberg, and Paul H. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press.

Blattman, Christopher, and Edward Miguel. 2010. "Civil War." *Journal of Economic Literature* 48 (1): 3–57.

Brück, Tilman, Patricia Justino, Philip Verwimp, and Alexandra Avdeenko. 2010. "Identifying Conflict and Violence in Micro-Level Surveys." HiCN Working Paper 79 Households in Conflict Network (HiCN).

Buhaug, Halvard, and Kristian Skrede Gleditsch. 2008. "Contagion or Confusion? Why Conflicts Cluster in Space." *International Studies Quarterly* 52 (2): 215–33.

Burnham, Gilbert, Riyadh Lafta, Shannon Doocy, and Les Roberts. 2006. "Mortality after the 2003 Invasion of Iraq: A Cross-Sectional Cluster Sample Survey." *Lancet* 368 (9545): 1421–28.

Bussmann, Margit, and Gerald Schneider. 2011. "Accounting for the Dynamics of One- Sided Violence: Introducing KOSVED." Working Paper.

Cederman, Lars-Erik, and Kristian Skrede Gleditsch. 2009. "Introduction to Special Issue on 'Disaggregating Civil War'." *Journal of Conflict Resolution* 53 (4): 487–95.

Checchi, Francesco. 2010. "Estimating the Number of Civilian Deaths from Armed Conflicts." *Lancet* 375 (9711): 255–57. http://www.sciencedirect.com/science/article/pii/S0140673609621280.

Clauset, A., C. Rohilla Shalizi, and M. E. J. Newman. 2009. "Power-Law Distributions in Empirical Data." Accessed on November 1, 2011. http://arxiv.org/pdf/0706.1062v2.

Coghlan, Benjamin, Richard J. Brennan, Pascal Ngoy, David Dofara, Brad Otto, Mark Clements, and Tony Stewart. 2006. "Mortality in the Democratic Republic of Congo: A Nationwide Survey." *Lancet* 367 (9504): 44–51.

Davenport, Christian. 2010. *Media Bias, Perspective, and State Repression: The Black Panther Party.* Cambridge, UK: Cambridge University Press.

Davenport, Christian, and Patrick Ball. 2002. "Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977-1996." *Journal of Conflict Resolution* 46 (3): 427–50.

Degomme, Olivier, and Debarati Guha-Sapir. 2010. "Patterns of Mortality Rates in Darfur Conflict." *The Lancet* 375: 294-300. Accessed on November 10, 2011. http://linkinghub.elsevier.com/retrieve/pii/S014067360961967X.

Earl, Jennifer, Andrew Martin, John D. McCarthy, and Sarah A. Soule. 2004. "The Use of Newspaper Data in the Study of Collective Action." *Annual Review of Sociology* 30: 65–80. Accessed on November 10, 2011. http://www.jstor.org/stable/29737685.

Eck, Kristine, and Lisa Hultman. 2007. "One-Sided Violence Against Civilians in War." *Journal of Peace Research* 44 (2): 233–46.

Gill, Jeff, and Lee D. Walker. 2005. "Elicited Priors for Bayesian Model Specifications in Political Science Research." *Journal of Politics* 67 (3): 841–72. Accessed on November 1, 2011. http://dx.doi.org/10.1111/j.1468-2508.2005.00342.x.

Gleditsch, Kristian Skrede. 2002. "Expanded Trade and GDP Data." *Journal of Conflict Resolution* 46 (5): 712–24. Accessed on July 1, 2010. http://jcr.sagepub.com/act.

Gohdes, Anita. 2010. "Different Convenience Samples, Different Stories: The Case of Sierra Leone." *Benetech Human Rights Data Analysis Group.*

Goldstein, Joshua S. 2011. *Winning the War on War: The Decline of Armed Conflict Worldwide.* New York: Dutton.

Goldstein, Joshua S., and Steven Pinker. 2011. "War is Really Going Out of Style", *op ed New York Times.* Accessed on December 20, 2012. http://www.nytimes.com/2011/12/18/opinion/sunday/war-really-is-going-out-of-style.html.

Hoddie, Matthew, and Jason M. Smith. 2009. "Forms of Civil War Violence and Their Consequences for Future Public Health." *International Studies Quarterly* 53 (1): 175–202.

HSRP (The Human Security Report Project). 2010. "Human Security Report 2009: The 'Shrinking Costs of War.'" Oxford University Press, Oxford, UK.

Kalyvas, Stathis. 2004. "The Urban Bias in Research on Civil Wars." *Security Studies* 13 (3): 160–90.

Kalyvas, Stathis. 2006. *The Logic of Violence in Civil War.* Cambridge, UK: Cambridge University Press.

Krüger, Jule, Patrick Ball, Megan Price, and Amelia Hoover Green. in press. "It Doesn't Add Up: Methodological and Policy Implications of Conflicting Casualty Data." In *Civilian Casualties and Strategic Peacebuilding*, edited by Taylor Seybolt, Baruch Fischhoff and Jay Aronson. Oxford University Press.

Lacina, Bethany. 2006. "Explaining the Severity of Civil Wars." *Journal of Conflict Resolution* 50 (2): 276.

Lacina, Bethany, and Nils P. Gleditsch. 2005. "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths." *European Journal of Population/Revue Européenne de Démographie* 21 (2): 145–66.

Lacina, Bethany, Nils P. Gleditsch, and Bruce Russett. 2006. "The Declining Risk of Death in Battle." *International Studies Quarterly* 50 (3): 673–80.

Lum, K., M. Price, T. Guberek, and P. Ball. 2010. "Measuring Elusive Populations with Bayesian Model Averaging for Multiple Systems Estimation: A Case Study on Lethal Violations in Casanare, 1998-2007." *Statistics, Politics, and Policy* 1 (1): 2.

Melander, E., M. Oberg, and J. Hall. 2009. "Are 'New Wars' More Atrocious? Battle Severity, Civilians Killed and Forced Migration Before and After the End of the Cold War." *European Journal of International Relations* 15 (3): 505.

Obermeyer, Ziad, Christopher J. L. Murray, and Emmanuela Gakidou. 2008. "Fifty Years of Violent War Deaths from Vietnam to Bosnia: Analysis of Data from the World Health Survey Programme." *British Medicine Journal* 336 (7659): 1482–86.

Pinker, Steven. 2011. *The Better Angels of Our Nature.* New York, NY: Viking.

Prelec, Draen. 2004. "A Bayesian Truth Serum for Subjective Data." *Science* 306 (5695): 462–66. Accessed on November 1, 2011. http://www.sciencemag.org/content/306/5695/462.abstract.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset." *Journal of Peace Research* 47 (5): 651–60.

Raleigh, Clionadh, and Håvard Hegre. 2009. "Population Size, Concentration and Civil War: A Geographically Disaggregated Analysis." *Political Geography* 28 (4): 224–38.

Richardson, Lewis F. 1948. "Variation of the Frequency of Fatal Quarrels With Magnitude." *Journal of the American Statistical Association* 43 (244): 523–46. Accessed on November 2, 2011. http://www.jstor.org/stable/2280704.

Silva, Romesh, and Patrick Ball. 2008. "The Demography of Conflict-Related Mortality in Timor-Leste (1974–1999): Reflections on Empirical Quantitative Measurement of Civilian Killings, Disappearances, and Famine-Related Deaths." *Statistical Methods for Human Rights* 117–139. New York: Springer.

Smith, Jackie, John D. McCarthy, Clark McPhail, and Boguslaw Augustyn. 2001. "From Protest to Agenda Building: Description Bias in Media Coverage of Protest Events in Washington, D.C." *Social Forces* 79 (4): 1397–423. Accessed on November 10, 2012. http://www.jstor.org/stable/2675477.

Spagat, Michael, Andrew Mack, Tara Cooper, and Joakim Kreutz. 2009. "Estimating War Deaths: An Arena of Contestation." *Journal of Conflict Resolution* 53:934–50.

Verwimp, Philip, Patricia Justino, and Tilman Brück. 2009. "The Analysis of Conflict: A Micro-Level Perspective." *Journal of Peace Research* 46 (3): 307–14.

Weinstein, J. M. 2007. *Inside Rebellion: The Politics of Insurgent Violence.* Cambridge, UK: Cambridge University Press.

Zwierzchowski, Jan, and Ewa Tabeau. 2010. "The 1992-95 War in Bosnia and Herzegovina: Census-Based Multiple Systems Estimation of Casualties' Undercount." Conference Paper for the International Research Workshop on "The Global Costs of Conflict".