

The Waning of War is Real: A Response to Gohdes and Price

Bethany Lacina¹ and
Nils Petter Gleditsch²

Journal of Conflict Resolution
57(6) 1109-1127
© The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0022002712459709
jcr.sagepub.com


Abstract

A number of recent studies argue that there is decline in armed conflict within and between nations. Gohdes and Price run against the grain in arguing that there is no evidence for a decrease in battle deaths in armed conflicts after World War II and that the trend reported in our earlier articles is spurious. However, they do not plausibly justify this thesis. We reexamine the argument for a decline, exploring nonlinearities in the data and potential biases due to measurement error. We find that very strong assumptions must hold in order for measurement errors to explain the trend in battle deaths.

Keywords

War death, battle death, conflict trend, conflict data

Several high-profile books have recently argued that both civil and interstate conflicts are in decline (Goldstein 2011; HSRP 2011; Pinker 2011).¹ Such trends in conflict are of much wider interest than an esoteric academic dispute about data. Thus, it is important, for both theory and policy, that flawed analysis not be allowed

¹University of Rochester, Centre for the Study of Civil War, PRIO, Rochester, NY, USA

²PRIO, Oslo, Norway

Corresponding Author:

Bethany Lacina, University of Rochester, Centre for the Study of Civil War, PRIO, University of Rochester Political Science, PO Box 270146, Rochester, NY 14627, USA.

Email: blacina@ur.rochester.edu

to obscure the very strong evidence that deadly violence in the world has been declining.

Gohdes and Price argue that the PRIO Battle Deaths Dataset does not provide convincing evidence of a decline in per capita battle deaths since 1946. However, there are a number of missteps in their analysis. In this article, we improve on their critique by laying out the strongest possible case for a spurious negative trend in battle deaths. We conclude that it is unlikely that model specification or measurement error explains the decline.

Our reply has four sections. The first deals briefly with some of the issues of data quality raised by Gohdes and Price (2012). The second addresses their discussion of the nonlinearity of the trend in battle deaths since 1945. Nonparametric tests show that the negative trend is not dependent on the assumption of linearity. Finally, we improve on Gohdes and Price's investigation of the uncertainty in the data. There are more sound and straightforward means of investigating measurement error than the simulation Gohdes and Price ran. We complete such an analysis and find that, in order for the negative trend to be an artifact of measurement error, multiple extreme assumptions must hold.

Issues of Data Quality

Gohdes and Price offer a valuable opportunity to discuss the role of data quality in the study of conflict. Their article also gives us a chance to revisit the trend in battle deaths documented in Lacina and Gleditsch (2005) and Lacina, Gleditsch, and Russett (2006), focusing on the robustness of that trend to model specification and measurement error. Before turning to such an analysis, we briefly discuss the sections of Gohdes and Price's article that deal with data quality. They make a few points with which we fundamentally disagree and one mistake that we wish to rectify.

First, Gohdes and Price have misread the data set's codebook in their analysis of missing best estimates. The data set is organized by conflict-year (e.g., Korean War 1950; Korean War 1951; Malaysian Emergency 1950). They correctly note that there are 771 conflict-years (40 percent) for which no best estimate is available, only a low and high estimate of deaths. They wrongly equate the lack of a best estimate to a reliance on the range of battle deaths reported in Uppsala Conflict Data Program (UCDP)/PRIO's Armed Conflict Dataset (Gleditsch et al., 2002), which is the list of conflicts upon which the PRIO Battle Deaths Dataset is based. When UCDP/PRIO's range of deaths is the only available information, the Battle Deaths Dataset records the bottom and top of that range as low and high estimates. The best estimate is missing.² But there are just 286 conflict-years coded in this manner. In the remaining 485 cases of a missing best estimate there is more detailed, conflict-specific information available for creating low and high estimates. However, there is insufficient information available for adjudicating between these values in order to choose a best estimate. Gohdes and Price's misunderstanding of the absent best estimates implies that there is less information in the data set than there in fact is.

On a more general point, Gohdes and Price argue that the absence of year-specific figures for a number of conflicts makes the PRIO Battle Deaths Dataset unsuitable for analysis of trends.³ This is an extreme conclusion. Even when the only available information is the total number of deaths in a conflict, the conflict itself is assigned to a specific interval of years. These intervals tend to be short. There are 57 multiyear conflict periods in the PRIO data for which only a total estimate of battle deaths is available. The median number of years in such a period is four. Conflicts for which no annual information is available also tend to be low intensity wars, such as the Cypriot war of independence or the Tripuri insurgency in India. The conflict periods of five or more years that lack trend information have a median high estimate of 28,000 total battle deaths, about 1.4 percent of the average five-year global toll in battle deaths. Such relatively low-intensity wars are unlikely to have much influence on global trend estimates regardless of which years in these conflicts were the most deadly.

Third, and most importantly, Gohdes and Price offer a confused discussion of the strengths of different sources of fatality estimates. They distinguish between methods that “offer statistically reproducible measures of uncertainty” (000) and data that “are not statistical estimates” (000). They correctly point out that one of the virtues of statistical sampling is that sampling-related uncertainty can be quantified. They contrast this advantage to data that are not based on random sampling, for which “no quantification of (possible) errors exists” (000). Such data are therefore “unrepresentative in an unknown way” (000).⁴

Uncertainty and unrepresentativeness (bias) are not the same, however. Sampling-based surveys, multiple-system estimation (MSE), and convenience data all have in common the problem that their biases are hard to quantify or even sign. A MSE fatality estimate, for example, comes with confidence intervals that capture uncertainty due to sampling. Yet, the bias in the estimate may be such that the true figure is much higher or lower than those confidence intervals imply. For example, bias may be due to a problem Gohdes and Price point to in their critique of convenience data: not all fatalities are equally likely to be detected by all compilers of information (Jewell, Spagat, and Jewell 2011).

Giving pride of place to statistical methods also leads Gohdes and Price to omit an entire category of high-quality fatality data, that collected through vital registration or enumeration. Militaries in particular tend to be enthusiastic enumerators. Thus, the US Department of Defense’s tally of soldiers killed in action does not come with estimates of uncertainty nor is it obtained through randomization. It nonetheless provides very credible data. To be sure, the availability of reliable count data is the exception rather than the rule in war. However, when such information is available it is often preferable to the data that could be obtained by statistical sampling. For instance, after the fall of the USSR, declassification provided researchers with the Soviet military’s estimates of rebels and soldiers killed during nationalist insurgencies in the Baltics between 1945 and the early 1950s (Anuškauskas, 2000). A survey probably could not improve on these figures. A survey’s recall period would be at least 62 years, there would be strong social-desirability bias in favor of reporting

involvement on the rebel side, and the local pro-Soviet population has presumably had high rates of out-migration. Any survey-obtained estimate of deaths would probably be of lower-quality than the archival figures.

We have touched on the most serious problems in Gohdes and Price's reflections on data quality. Ultimately, much of that discussion is irrelevant to whether the Battle Deaths Dataset can be used to study trends. Gohdes and Price need to establish the possibility of a temporally varying bias in the data set. We propose and investigate such an argument in later sections of this article. First, however, we discuss the nonlinearity in the downward trend in battle deaths.

The Uneven Decline in Battle Deaths

Gohdes and Price's statistical analysis begins by highlighting the nonmonotonicity of the trend in battle deaths. By adding a dummy variable for the period 1946–1950, fitting a cubic model to the data, and by excluding the earliest years of the time series from their simulation, they demonstrate the presence of nonlinearity and highlight the role of outliers.

Indeed, when plotted over time, battle deaths are a series of spikes and valleys. Figure 1 plots the low and high estimate of battle deaths per 100,000 people from 1946 to 2008.⁵ Lacina and Gleditsch (2005) point out that the data are dominated by several peaks, each somewhat lower than the last:

although there have been multiple major international security crises their military scale has progressively diminished. . . . The very large conflicts that these peaks represent almost overwhelm the rest of the curve (p. 155).

Lacina, Gleditsch, and Russett (2006, 674) pointed out how fitting a line to battle deaths data for 1900 to 1997 completely obscured any trends because the World Wars were such outliers:

[the] "flat-line" finding [reported by Sarkees, Wayman, and Singer (2003)] is driven primarily by the massive spikes in the middle of [the] timeline representing the two World Wars. These wars were several orders of magnitude more deadly than any conflicts before or since, and their presence in the regression line obscures other trends.

More recently, Pinker (2011) has argued that global deaths in war have fallen over in the very long run but that the decline has been far from a monotonic. Goldstein (2011) also highlights the multiple surges and lulls in conflict deaths since 1945.

One implication of the spikiness of battle deaths over time is that a polynomial will be able to fit the data more exactly than a line. It should also be obvious just by looking at Figure 1 that a flatter ordinary least squares (OLS) line will be produced by truncating the series at 1950 or allowing the observations for

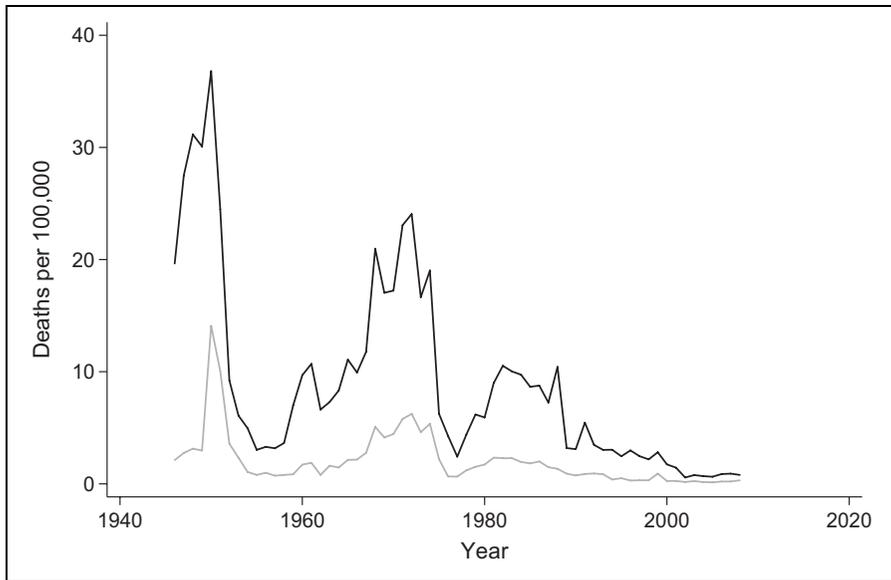


Figure 1. Low and high estimates of deaths in battle per 100,000 people. Calculated from the PRIO Battle Deaths Dataset, v 3.0.

1946–1950 to have a unique intercept. In either case, the fitted line will start and end in “valley” years and be rather flat.

However, there are more systematic means of dealing with nonlinearities in data than fitting an arbitrary polynomial or dropping outliers. For example, we can apply a smoothing function to the time series of battle deaths. By considering a moving window of data, a smoothing function also helps to address Gohdes and Price’s concern that the data for some conflicts cannot be trended by year. Figure 2 plots locally weighted linear regressions of the low and high estimates of worldwide battle deaths normalized by population.⁶ The resulting functions are not monotonically decreasing. However, they certainly preserve the appearance of an overall downward trend.

The smoothing functions do not produce test statistics. For hypothesis evaluation, we need a nonparametric test for a trend. We have chosen Kendall’s τ_a , a well-known rank test (Kendall 1938; Mann 1945).⁷ τ_a is computed by examining every pair of observations in the data and counting the number of pairs in which the higher number of battle deaths is recorded in the later of the two year; these cases are considered to be concordant pairs. Discordant pairs are those in which the larger number of battle deaths occurred in the earlier year. The test statistic is

$$\tau_a = \frac{\text{Number of concordant pairs} - \text{Number of discordant pairs}}{\text{Total number of pairs}}. \tag{1}$$

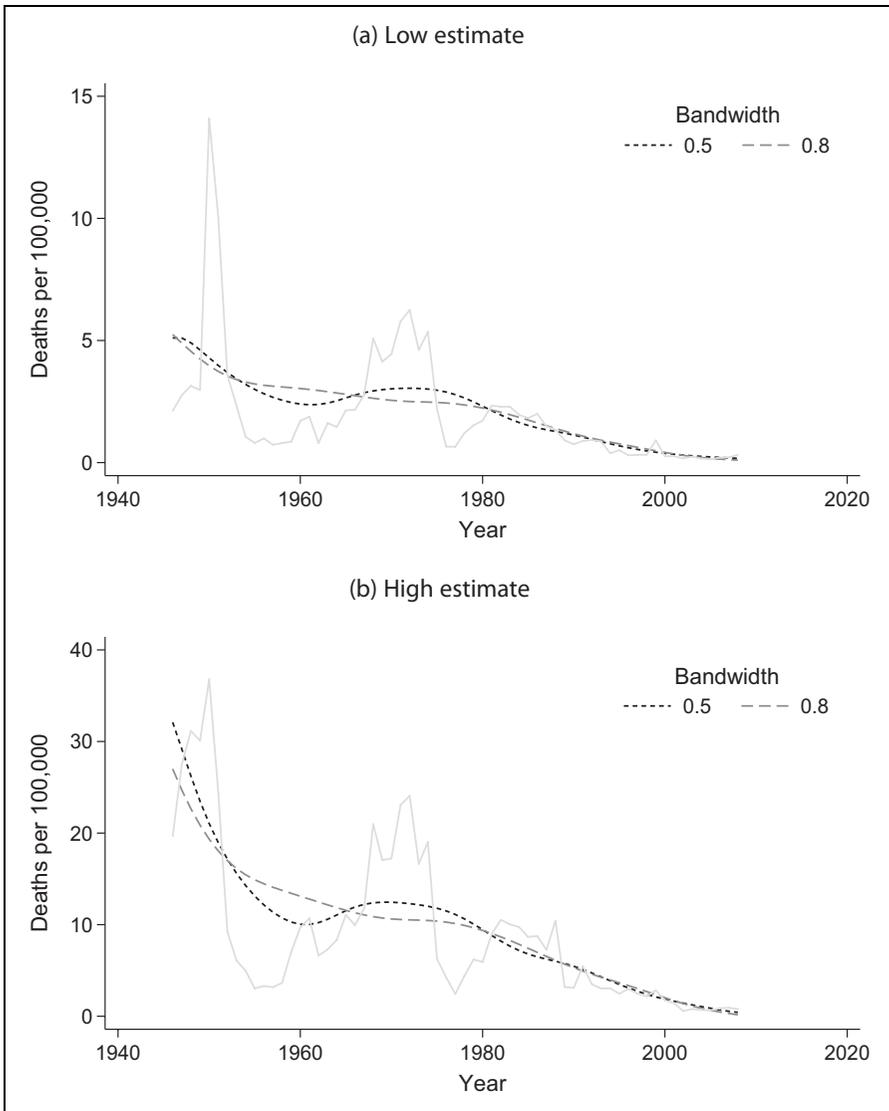


Figure 2. Locally weighted regressions of time versus low and high estimates of global battle deaths per 100,000 people.

Under the null hypothesis of no trend in the data, the expected value of τ_a is zero.⁸ The sign of τ_a indicates the direction of any estimated trend.⁹

For both the low and high series of estimates, we find a negative and statistically significant τ_a when looking at the periods 1946–2008, 1946–1997, and 1951–1997

(Table 1). A univariate OLS regression of battle deaths on time also estimates a statistically significant negative trend for any of these periods. Thus, moving from linear to nonparametric tests does not change our conclusion of a negative trend in battle deaths.

Estimating Uncertainty

The second task Gohdes and Price set to themselves is to model the uncertainty in the trend in battle deaths due to measurement error. To do so, they construct 1,000 time series by randomly choosing one of the available estimates (low, high, or best) for each conflict-year. They sum these conflict-year figures into annual global totals and normalize by population. Then, they calculate OLS regressions for each of the resulting time series. They report the average $\hat{\beta}_{Time}$ from these 1,000 regressions.

The expected value of battle deaths under this sampling procedure is the mean of the available estimates for each conflict-year. The simulation therefore recovers the same $\hat{\beta}_{Time}$ that would result from taking the mean of those estimates directly and then running a regression (Table 1, model 7). Gohdes and Price are more interested, however, in how many of these 1,000 regressions have statistically significant coefficients on the index for time. They can reject the null hypothesis of no trend at the 95 percent confidence level in 98.5 percent of their series when the data from 1946 to 1997 are used. If the pre-1951 observations are dropped, just 21 percent of these series return statistically significant coefficients on time. They argue that the fact that many estimated coefficients are statistically insignificant when they use data for only 1951 to 1997 should limit our confidence in the trend in the data.

Gohdes and Price do not provide a full-fledged justification for truncating the time series or for randomizing the selection of estimates instead of, for example, estimating a regression using all the high estimates or all the low estimates. Their procedure should not be confused with bootstrapping, which draws random samples of a data set and repeatedly estimates a statistic. Bootstrapping the estimates in Table 1 would involve drawing random samples of *years*.

A potential rationale for Gohdes and Price's procedure is to recover the range of $\hat{\beta}_{Time}$ that could be obtained from all the various permutations of low, high, and best estimates across conflict-years. The bounds on $\hat{\beta}_{Time}$ are arguably interesting. (Although, as we will discuss below, these bounds can be calculated directly.) However, Gohdes and Price report averages instead of bounds. For that average to be meaningful, it must be the case that each possible sample—that is, each possible combination of low, high, and best estimates across conflict-years—is equally likely to be correct. Such an assumption is entirely untenable. It implies, obviously, that there is no information whatsoever in the designation of one of the estimates as a best estimate; it is equally likely as the low and the high. More subtly, the equal probability assumption means that for a multiyear conflict the likelihood of different estimates being correct is uncorrelated across the years of the conflict. Given that the data sources available for the years of a conflict are often identical, the odds that all

Table 1. The OLS Regressions and Kendall's Rank Tests for a Trend in Battle Deaths Divided by Global Population, 1946–2008.

	Low estimates			High estimates			Averaged estimates
	1946–2008 (1)	1946–1997 (2)	1951–1997 (3)	1946–2008 (4)	1946–1997 (5)	1951–1997 (6)	1951–1997 (7)
OLS Results							
β_{Time}	-6.4e-07	-6.2e-07	-4.7e-07	-3.0e-06	-2.9e-06	-1.3e-06	-7.6e-07
(Standard error)	(1.8e-07)	(2.5e-07)	(2.3e-07)	(5.3e-07)	(7.3e-07)	(5.9e-07)	(3.9e-07)
p Value	.001	.017	.041	.000	.000	.037	.055
Kendall's rank test							
τ_a	-.49	-.30	-.24	-.53	-.34	-.22	-.21
p Value	.000	.002	.018	.000	.000	.033	.037
Observations	63	52	47	63	52	47	47

Note: p values are from a two-sided test of $H_0 : \beta_{Time} = 0$ or $H_0 : \tau_a = 0$. Heteroscedasticity robust standard errors are reported for OLS results.

the high estimates for one conflict are correct or all the low estimates are correct are much better than the odds that the correct values are a random sequence of low and high estimates. Finally, if every draw of the data is equally probable, there is an implied zero correlation in the likelihood of estimates from conflict-years in the same region or involving the same actor.

Since the assumption of the equal probability of each possible draw of the low, high, and best conflict-year figures is not tenable, averaging the values of $\hat{\beta}_{Time}$ makes no sense. Nor is there any reason to tabulate the percentage of times that $\hat{\beta}_{Time}$ is statistically significant. Doing so is analogous to reasoning that because there are two outcomes to a lottery, winning and losing, the expected value of a lottery ticket is one-half of the jackpot.

Measurement Error and Reanalyzing the Trend in Battle Deaths

Gohdes and Price lack a coherent explanation of how uncertainty in the battle deaths data would translate into a spurious trend and their simulation does not properly investigate this possibility. In this section, we explore the most plausible reasons that measurement error might produce a negative trend in battle deaths.

The argument that Gohdes and Price need to make is that the uncertainty in the battle deaths data is neither random nor merely changing in variance over time. Rather, they should argue that the measurement errors have a directional bias that varies with time. The worst case scenario for our hypothesis would be if the early estimates are systematically inflated by measurement error and the later estimates systematically deflated by measurement error. More plausibly, measurement error may run in a single direction but be more severe in some periods than in others. For example, perhaps lower in quality data are systematically inflated. The downward trend might then be an artifact of increased accuracy. A spurious trend could also be generated by relatively accurate early data combined with later data that have systematic negative errors, although this scenario seems unlikely, given that indicators of data quality, such as presence of a best estimate and availability of trend information, increase over time.

Underestimates or Overestimates?

To make a well-supported argument about how measurement error might create a spurious trend in battle deaths, we need to consult the literature on probable biases in the various sources in the PRIO data set.

First, there is near consensus that compiling incidents from media sources leads to underestimates of deaths (Spagat et al. 2009, fn12). Davenport and Ball (2002) and Restrepo, Spagat, and Vargas (2006) have argued that media sources underreported deaths in conflicts in Guatemala and Colombia. Earl et al. (2004) find underreporting of protests even in the media-saturated United States. The downward bias of media-based data may influence the PRIO Battle Deaths Dataset, given that one

of its sources is UCDP's fatalities data set (UCDP 2011b) available for 1989 through 2011. The UCDP data are based primarily on media report coding and therefore are a potential source of underestimation in the PRIO data.¹⁰ However, we doubt that the use of UCDP's figures represents a major confound. In the post-1989 period, there are usually alternative estimates available to weigh against UCDP's figures. As a result, estimates in the PRIO data set are almost always higher than those in UCDP's data.

The PRIO data set rarely has relevant surveys or MSEs from which to draw information. Those that are available tend to be for more recent conflicts. Both under- and overestimation of mortality are possible when these methods go awry. However, as Gohdes and Price point out, the intense scrutiny that studies of this kind frequently receive helps to pinpoint errors in particular cases. Thus, we doubt that these sources create a systematic bias in the PRIO battle deaths data.

Official death statistics, generally compiled by enumeration, are another major category of battle deaths' data. Such information may be released immediately; revealed through declassification (e.g., Ryan, Finkelstein, and McDevitt, 2003); or collected in connection with legal proceedings (e.g., Tabeau and Bijak 2005). Some states do not have the capacity to collect this kind of information and available figures may be suspect because of the agenda of the government releasing the data. We do not know of research that has established whether governments, on average, have an incentive to exaggerate or minimize their losses. There may, however, be some systematic bias in government-released enumerations toward underestimation, because in any enumeration-based procedure some deaths can be missed entirely.

Finally, the PRIO Battle Deaths Dataset draws heavily on expert assessments, a potential source of overestimation. Expert estimates of deaths might be influenced by the ease with which people remember events over nonevents. This availability bias can lead to "disproportionate risk assessment . . . due to the exposure to negative outcomes even if the event is itself rare" (Kynn 2008, 244). Another possible source of upward biases in expert estimates is the blurring between expertise and advocacy. Experts may overestimate deaths because they seek to draw attention to ongoing conflicts or to underline the importance of the conflict on which they specialize (Cohen and Green 2012).

Given that the biases in PRIO's sources run in different directions, can we say anything about how measurement error has evolved over time? PRIO's estimates for more recent conflicts tend to draw on a larger number of sources than those for earlier conflicts. Incident-based analyses, MSE, and surveys are particularly rare in the early parts of the data; official figures and expert assessments are more important. We have suggested these two sources have countervailing biases, the former toward underestimation and the latter toward overestimation. The relative magnitude of these biases is not known. However, the most plausible case for measurement error creating a spurious trend would argue that errors in official sources are relatively unimportant, perhaps because these errors are random or small, while expert assessments create systematic overestimation. Higher quality recent data do not have these errors. The negative trend in the data is thus a product of increased accuracy. The

next section evaluates how likely it is that changes in measurement error over time account for the trend in battle deaths.

The Potential Effects of Measurement Error

As already noted, the worst-case scenario for our claim of a decline in battle deaths is the case in which the low estimates of battle deaths are correct in early years in the data set and the high estimates are correct in later years. In order to explore this scenario, we can regress battle deaths on time or perform a rank test using data series of battle deaths composed of only low estimates prior to a certain year and only high estimates thereafter. This procedure can also be used to find the upper bound on $\hat{\beta}_{Time}$, that is the greatest value that could be obtained from any combination of the low, high, and best estimates for all conflict-years. The maximum $\hat{\beta}_{Time}$ is produced by fitting a line to a data series composed of low estimates before the mean year in the series and high estimates for later years. (We refer here to the maximum $\hat{\beta}_{Time}$ in terms of the number line and not the maximum absolute value. A proof of this claim is included as Appendix A.)

To begin our analysis, we construct time series with low estimates before a cutoff year and high estimates after that year. We create 63 such series, using every year from 1946 to 2008 as the cutoff between low and high estimates. Then, we estimate an OLS regression and find Kendall's τ_a for each of these 63 series.¹¹

Figure 3 plots the values of $\hat{\beta}_{Time}$, with 95 percent confidence intervals, from these OLS regressions. The x -axis is the cutoff year between low and high estimates. The graph shows that there is no possible permutation of estimates that will yield a point estimate of a positive trend. $\hat{\beta}_{Time}$ is always negative. However, the confidence intervals for the regression include zero or a weakly positive slope if the cutoff year between the low and high estimates falls between 1961 and 1988.

The black line in Figure 4A plots the p values from a two-tailed test of the null hypothesis that the OLS coefficient on time is zero. The black line in Figure 4B does the same for Kendall's rank test ($H_0 : \tau_a = 0$).¹² In each graph there is a range of years that, if they are taken as the dividing line between low and high estimates, produce statistically insignificant test statistics. Thus, it is not impossible that systematic positive errors in the early years of the data and negative errors in the later years are generating a spurious negative trend in battle deaths.

However, relaxing the extreme assumptions used to build these data series quickly produces more robust trend estimates. Consider what happens if we modify the time series just described by using the best estimates for conflict-years where they are available. If there is no best estimate, we record the low estimate of deaths in a conflict-year before a cutoff year and the high estimates after that cutoff year. We then sum to obtain the global annual battle deaths. This procedure admits the possibility of inflation in earlier estimates and deflation in later estimates for conflict-years where little information is available. However, where best estimates

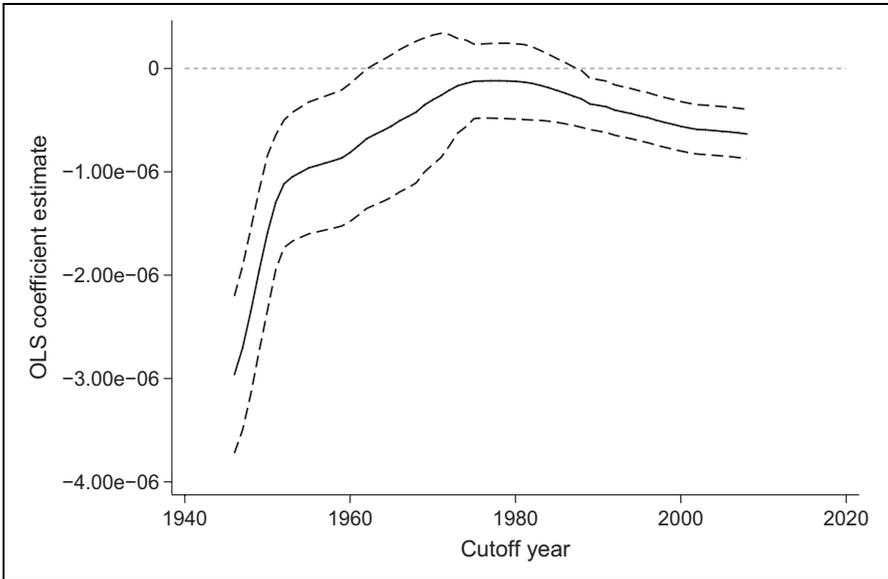


Figure 3. $\hat{\beta}_{Time}$, with 95 percent confidence intervals, from an OLS regression using a time series of low estimate of battle deaths before a cutoff year and high estimates thereafter.

are available, they are presumed to be the most likely values. The p values from statistical tests using these time series are plotted as gray lines in Figure 4A and B. Now we obtain a statistically significant trend estimate regardless of the year at which the low estimates give way to high estimates.

Thus far, we have entertained the possibility that the average measurement error in the data runs in opposite directions in different eras. However, we can think of no plausible reason why average measurement error would have flipped direction. Instead, we believe the most plausible argument for a spurious trend is that heavier reliance on expert assessments in earlier conflicts means that the data have an upward bias that diminishes over time. A negative trend is generated by increasing accuracy of the data. To explore this possibility, we construct another set of 63 series based on using low estimates before a cutoff year and best estimates thereafter.¹³

It is also possible that the early data are roughly correct but the later data are systematically biased downward. Such errors would also produce a spurious negative trend. Our fourth set of sixty-three data series uses best estimates before a cutoff year and high estimates thereafter.¹⁴

The p values from OLS regressions and rank tests using the two sets of data just described are plotted in Figure 5. The test statistics are all statistically significant at the 90 percent confidence level and generally at 95 percent, in both the OLS and rank test results. Thus, the trend appears robust to the possibility of unidirectional measurement errors that vary in severity over time.

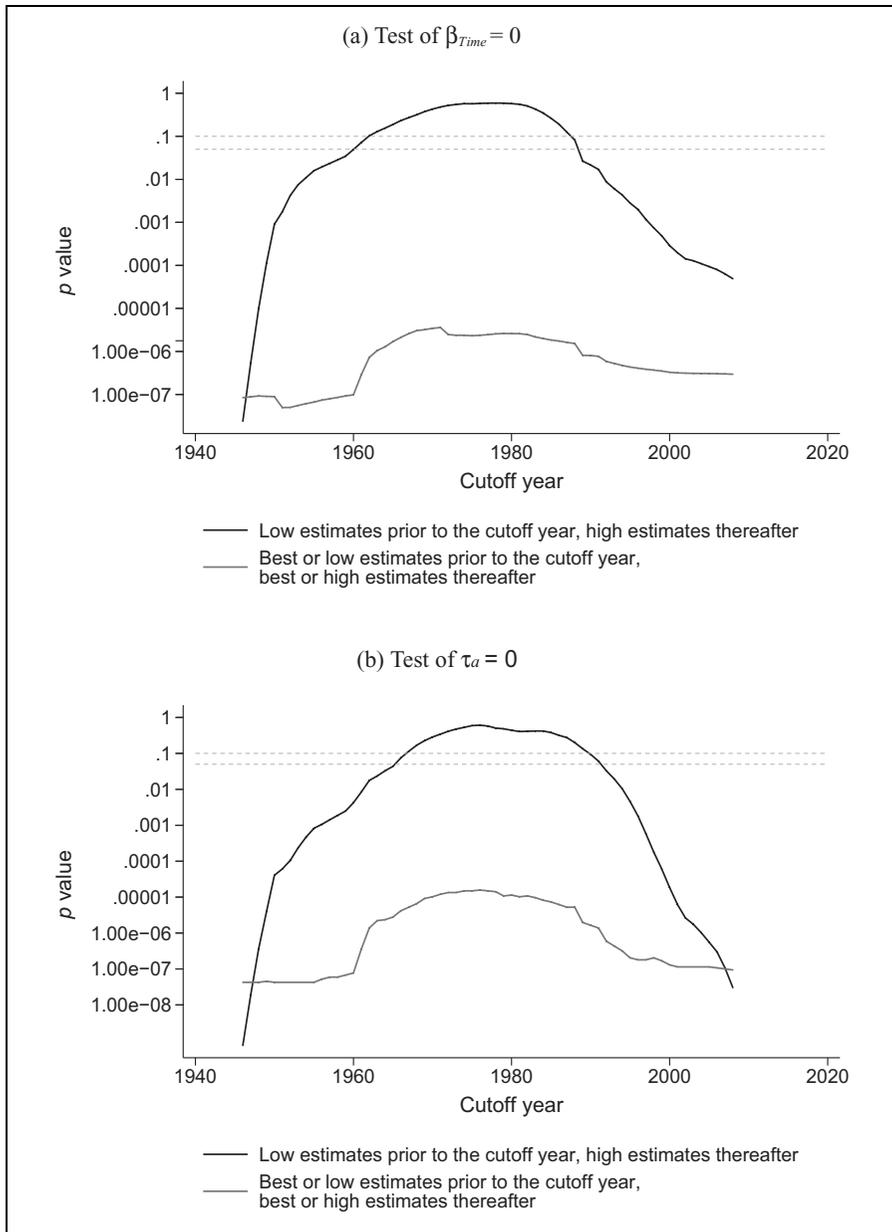


Figure 4. Statistical significance tests from OLS regressions of battle deaths on time and of Kendall's τ_a using data series incorporating the possibility of measurement errors that change direction over time.

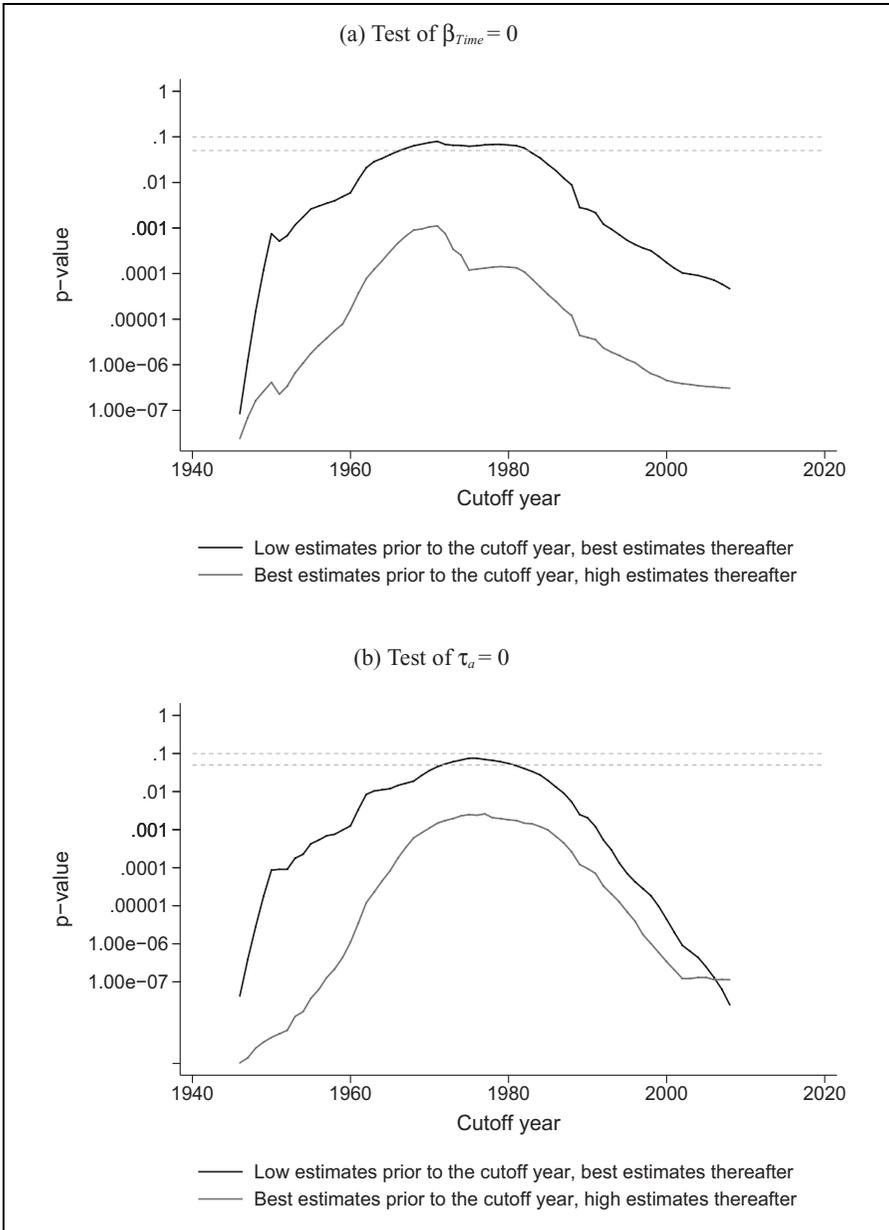


Figure 5. Statistical significance tests from OLS regressions of battle deaths on time and of Kendall's τ_a using data series incorporating the possibility of unidirectional measurement errors that change in magnitude over time.

The preceding analysis shows what needs to hold in order for the negative trend in battle deaths between 1945 and 2008 to be the result of measurement error. Such error must have produced systematically inflated estimates in early data and systematically deflated estimates in later data. If the bias in the data runs in only one direction but varies in magnitude over time, the estimate of a negative trend is robust (cf. Figure 4 and Figure 5). After reviewing the potential biases in our sources, we do not believe that it is plausible that measurement error flips in this way. Even if it does, when the best estimates of battle deaths are used where they are available, a statistically significant negative trend obtains even under the extreme assumption of errors that reverse direction over time (cf. estimates in Figure 4). Thus, in order for measurement error to explain the observed trend, we have to assume that the best estimates contain very little information about the most likely values in the data and that the direction of the bias in the data is changing over time.

Conclusion

Gohdes and Price recognize that the PRIO data are the most comprehensive information available on battle deaths after World War II. They are not optimistic that new investigations of long-past conflicts could substantially improve on the data set. Thus, researchers interested in trends in conflict must decide whether the data constitute a preponderance of evidence or whether there is too much uncertainty to warrant a conclusion. Our investigations suggest overwhelming evidence for a negative trend in battle deaths between 1946 and 2008. Nonparametric modeling shows that the trend is not an artifact of a linearity assumption. It is also very unlikely that the trend is the result of measurement error. In sum, Gohdes and Price's conclusion that the available evidence of a trend is insufficient requires an extraordinary tolerance for Type II error—the error of seeing nothing when we have an opportunity to learn about the world.

Appendix

Consider two parallel series of data for the years $i = \{1, 2, \dots, n\}$: a series of low estimates of deaths in each year denoted $l_i = \{l_1, l_2, \dots, l_n\}$ and a series of high estimates denoted $h_i = \{h_1, h_2, \dots, h_n\}$, where $l_i < h_i \forall i$.

The equation to be estimated by OLS is $y_i = \alpha + \beta i + \epsilon_i$, where each y_i is either the low or the high estimate available for that year—that is $y_i \in \{l_i, h_i\}$. Our claim is that, out of the 2^n possible series of low and high estimates, the one that will produce the maximum OLS estimate of β is the series composed of l_i for $i < \bar{i}$ and h_i for $i > \bar{i}$. Note that we are referring here to the maximum value of $\hat{\beta}$ in terms of the real number line not in terms of absolute value.

Consider an arbitrary sequence of y_i s where each $y_i \in \{l_i, h_i\}$. Will $\hat{\beta}$ be larger if we set any one observation (y_j) to the high estimate (h_j) or to the low estimate (l_j),

while leaving the rest of the time series unchanged? Call the estimate that results from using h_j in the time series $\hat{\beta}_{h_j}$ and the estimate that results from using l_j in the time series $\hat{\beta}_{l_j}$. We want to show that $\hat{\beta}_{h_j} > \hat{\beta}_{l_j}$ if $j > \bar{i}$ and that $\hat{\beta}_{h_j} < \hat{\beta}_{l_j}$ if $j < \bar{i}$.

The OLS estimator of β is

$$\hat{\beta} = \frac{\sum iy_i - \frac{1}{n} \sum i \sum y_i}{\sum i^2 - \frac{1}{n} (\sum i)^2}. \tag{A1}$$

By substituting \bar{i} for $\frac{1}{n} \sum i$ and rewriting the remaining summation signs in terms of j and $i \neq j$, equation (A1) is restated as

$$\hat{\beta} = \frac{\sum_{i \neq j} iy_i - \bar{i} \sum_{i \neq j} y_i}{\sum_{i=1}^n i^2 - \frac{1}{n} (\sum_{i=1}^n i)^2} + \frac{jy_j - \bar{i}y_j}{\sum_{i=1}^n i^2 - \frac{1}{n} (\sum_{i=1}^n i)^2}. \tag{A2}$$

By equation (A2), $\hat{\beta}_{h_j} > \hat{\beta}_{l_j}$ if

$$\begin{aligned} \frac{\sum_{i \neq j} iy_i - \bar{i} \sum_{i \neq j} y_i}{\sum_{i=1}^n i^2 - \frac{1}{n} (\sum_{i=1}^n i)^2} + \frac{jh_j - \bar{i}h_j}{\sum_{i=1}^n i^2 - \frac{1}{n} (\sum_{i=1}^n i)^2} &> \frac{\sum_{i \neq j} iy_i - \bar{i} \sum_{i \neq j} y_i}{\sum_{i=1}^n i^2 - \frac{1}{n} (\sum_{i=1}^n i)^2} + \frac{j l_j - \bar{i} l_j}{\sum_{i=1}^n i^2 - \frac{1}{n} (\sum_{i=1}^n i)^2} \\ \frac{jh_j - \bar{i}h_j}{\sum_{i=1}^n i^2 - \frac{1}{n} (\sum_{i=1}^n i)^2} &> \frac{j l_j - \bar{i} l_j}{\sum_{i=1}^n i^2 - \frac{1}{n} (\sum_{i=1}^n i)^2} \\ jh_j - \bar{i}h_j &> j l_j - \bar{i} l_j \\ h_j(j - \bar{i}) &> l_j(j - \bar{i}) \\ j - \bar{i} > 0 & \text{ Recall that } l_i < h_i \forall i \\ j &> \bar{i}. \end{aligned}$$

Similar reasoning can show that $\hat{\beta}_{l_j} > \hat{\beta}_{h_j}$ if $j < \bar{i}$ and that $\hat{\beta}_{l_j} = \hat{\beta}_{h_j}$ if $j = \bar{i}$. Note that determining whether $\hat{\beta}_{l_j} > \hat{\beta}_{h_j}$ or $\hat{\beta}_{l_j} < \hat{\beta}_{h_j}$ does not depend on knowing the values of y_i for $i \neq j$.

Applying the analysis above to the entire time series, $\hat{\beta}$ is maximized when $y_i = l_i \forall i < \bar{i}$ and $y_i = h_i \forall i > \bar{i}$. $\hat{\beta}$ is minimized when $y_i = h_i \forall i < \bar{i}$ and $y_i = l_i \forall i > \bar{i}$.

Acknowledgments

We would like to thank Bruce Russett for advice and input throughout the writing of this paper. Thanks are also due to Kevin Clarke, Kristian Skrede Gleditsch, Joshua Goldstein, Håvard Hegre, Gary King, Joakim Kreutz, Andy Mack, Håvard Nygård, Steven Pinker, Mike Spagat, and Gerdis Wischnath. Gleditsch acknowledges funding from the Research Council of Norway. The PRIO Battle Deaths Dataset can be downloaded from www.prio.no/cscw/cross/battledeaths. Replication data for this article are available at jcr.sagepub.com and www.prio.no/cscw/datasets.

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 disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Nils Petter Gleditsch would like to acknowledge the financial support of the Research Council of Norway.

Notes

1. See also skeptical analyses of that trend in Sarkees, Wayman, and Singer (2003) and Obermeyer, Murray, and Gakidou (2008) and the relevant rejoinders in Lacina, Gleditsch, and Russett (2006) and Spagat et al. (2009), respectively.
2. Use of the Armed Conflict Dataset's estimated range of deaths is indicated with a dummy variable named *Source*. See the codebook for details (www.prio.no/cscw/cross/battledeaths).
3. For each conflict-year, an indicator variable notes whether the battle death figures are based on sources that estimate fatalities for that year; based on aggregated figures divided between conflict-years using limited information about trends or based on an aggregate figure averaged across conflict-years. Eighteen percent of conflict-years are coded based on limited information about trends, for example annual data on losses for one side of a conflict. Thirty-five percent of conflict-years are based on aggregate estimates averaged over a span of time.
4. Gohdes and Price also imply that surveys conducted through random sampling and multiple-system estimation (MSE) have the unique virtues of transparency and replicability. Transparency is obviously not the particular providence of either technique nor even of quantitative methods. As for replicability, Gohdes and Price presumably do not mean to insist on this criterion in a strong sense, given that conflict data involve a specific population in a specific time frame, rather than a system that can be replicated many times, as in a laboratory. Among the methods used to estimate conflict mortality, only surveys can take advantage of resampling. Yet even resurveying is a very imperfect means of replication, as time elapsed since earlier studies will have altered the target population through a variety of poorly observed factors, such as migration rates.
5. The population data are those used in Lacina, Gleditsch, and Russett (2006), supplemented with information from the Penn World Table (Heston, Summers, and Aten, 2011).
6. At each point in the time series, a line is fit based on a window of data around that point. A fitted value is recorded based on the estimated line. Then the process is repeated with the next observation in the series. The width of the window of data used at each point is called the bandwidth. A bandwidth of 0.5, for example, means that 50 percent of the total number of observations are included in the window. A tri-cube weighting function is also applied to the regressors so that proximate observations are weighted more heavily (Cleveland, 1979).

7. Kendall's τ_a and τ_b are identical if there are no tied values in the data, which there are not in our case.
8. Hypothesis testing is based on an approximated normal distribution, $\tau_a \sim N\left(0, \frac{2(2n+5)}{9n(n-1)}\right)$. Kendall's rank test requires only the assumption that we have independent observations.
9. A value of -1 would be obtained if the number of battle deaths were strictly decreasing over time. A value of 1 would be obtained if the number of battle deaths were strictly increasing over time.
10. The UCDP data are also conservative by design (see UCDP 2011a).
11. All the results in this section are similar if data from 1946 to 1997 are used.
12. τ_a is negative for all the series constructed.
13. If the best estimate is missing for an observation that falls after the cutoff year, we use the high estimate. Using low estimates in place of missing best estimates would be more consistent with the idea of unidirectional measurement errors that diminish in significance over time. Such a procedure produces even stronger evidence in favor of a negative trend in battle deaths.
14. If the best estimate is missing for an observation before the cutoff year, the low estimate is used. Substituting high estimates for missing best estimates strengthens the evidence in favor of a negative trend in battle deaths.

References

- Anuškauskas, A. 2000. "A Comparison of the Armed Struggles for Independence in the Baltic States and Western Ukraine." In *The Anti-Soviet Resistance in the Baltic States*, edited by A. Anuškauskas, 63–70. Vilnius, Lithuania: Du Ka.
- Cleveland, W. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74 (368): 829–36.
- Cohen, D., and A. H. Green. 2012. "Dueling Incentives: Sexual Violence in Liberia and the Politics of Human Rights Advocacy." *Journal of Peace Research* 49 (3): 445–58.
- Davenport, C., and P. Ball. 2002. "Views to a Kill: Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977–1995." *Journal of Conflict Resolution* 46 (3): 427–50.
- Earl, J., A. Martin, J. D. McCarthy, and S. A. Soule. 2004. "The Use of Newspaper Data in the Study of Collective Action." *Annual Review of Sociology* 30: 65–80.
- Gleditsch, N., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand. 2002. "Armed Conflict 1946–2001: A New Dataset." *Journal of Peace Research* 39 (5): 615–37.
- Gohdes, A., and M. Price. 2012. "First Things First: Assessing Data Quality before Model Quality." *Journal of Conflict Resolution* 56.
- Goldstein, J. S. 2011. *Winning the War on War: The Decline of Armed Conflict Worldwide*. New York, NY: Dutton.
- Heston, A., R. Summers, and B. Aten. 2011. *Penn World Table, Version 7.0*. Philadelphia: Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania. Accessed March 19, 2012. www.pwt.econ.upenn.edu.

- HSRP (The Human Security Report Project). 2011. *Human Security Report 2009/2010: The Causes of Peace and the Shrinking Costs of War*. New York: Human Security Report Project and Oxford University Press.
- Jewell, N. P., M. Spagat, and B. Jewell. 2011. "Capture-Recapture Estimation of War Deaths: Foundations, Assumptions, and Challenges." Paper presented at the Joint Statistical Meetings of the American Statistical Association on August 3. Miami Beach, FL.
- Kendall, M. 1938. "A New Measure of Rank Correlation." *Biometrika* 30 (1-2): 81-89.
- Kynn, M. 2008. "The 'Heuristics and Biases' Bias in Expert Elicitation." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 171 (1): 239-64.
- Lacina, B., and N. P. Gleditsch. 2005. "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths." *European Journal of Population* 21 (2-3): 145-66.
- Lacina, B., N. P. Gleditsch, and B. Russett. 2006. "The Declining Risk of Death in Battle." *International Studies Quarterly* 50 (3): 673-80.
- Mann, H. 1945. "Nonparametric Tests against Trend." *Econometrica* 13 (3): 245-59.
- Obermeyer, Z., C. J. L. Murray, and E. Gakidou. 2008. "Fifty Years of Violent War Deaths from Vietnam to Bosnia: Analysis of Data from the World Health Survey Programme." *British Medical Journal* 336 (7659): 1482-86.
- Pinker, S. 2011. *The Better Angels of Our Nature: Why Violence Has Declined*. New York: Viking.
- Restrepo, J., M. Spagat, and J. Vargas. 2006. "The Severity of the Colombian Conflict: Cross-country Datasets Versus New Micro-Data." *Journal of Peace Research* 43 (1): 99-115.
- Ryan, M. A., D. M. Finkelstein, and M. A. McDevitt, eds. 2003. *Chinese Warfighting: The PLA Experience Since 1949*. London: M.E. Sharpe.
- Sarkees, M., F. Wayman, and J. Singer. 2003. "Inter-State, Intra-State, and Extra-State Wars: A Comprehensive Look at their Distribution Over Time, 1816-1997." *International Studies Quarterly* 47 (1): 49-70.
- Spagat, M., A. Mack, T. Cooper, and J. Kreutz. 2009. "Estimating War Deaths: An Arena of Contestation." *Journal of Conflict Resolution* 53 (6): 934-50.
- Tabeau, E., and J. Bijak. 2005. "War-Related Deaths in the 1992-1995 Armed Conflicts in Bosnia and Herzegovina: A Critique of Previous Estimates and Recent Results." *European Journal of Population* 21 (2-3): 187-215.
- UCDP (Uppsala Conflict Data Program). 2011a. *UCDP Battle-Related Deaths Dataset Codebook: Definitions, Sources and Methods for the UCDP Battle-Related Death Estimates*. Uppsala, Sweden: Uppsala Conflict Data Program. Accessed March 29, 2012. <http://www.ucdp.uu.se>.
- UCDP (Uppsala Conflict Data Program). 2011b. *UCDP Battle-Related Deaths Dataset v.5-2011*. Uppsala, Sweden: Uppsala Conflict Data Program. Accessed March 29, 2012. <http://www.ucdp.uu.se>.