
Using Power Laws to Estimate Conflict Size

Jeffrey A. Friedman¹

Journal of Conflict Resolution

1-26

© The Author(s) 2014

Reprints and permission:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0022002714530430

jcr.sagepub.com



Abstract

Casualty counts are often controversial, and thorough research can only go so far in resolving such debates—there will almost always be missing data, and thus, a need to draw inferences about how comprehensively violence has been recorded. This article addresses that challenge by developing an estimation strategy based on the observation that violent events are generally distributed according to power laws, a pattern that structures expectations about what event data on armed conflict would look like if those data were complete. This technique is applied to estimate the number of Native American and US casualties in the American Indian Wars between 1776 and 1890, demonstrating how scholars can use power laws to estimate conflict size, even (and perhaps especially) where previous literature has been unable to do so.

Keywords

Richardson's law, civilian casualties, civil wars, conflict, war

Estimating the severity of armed conflict is generally difficult. During recent violence in Iraq, Afghanistan, and the Balkans, for instance, casualty estimates ranged widely, despite efforts to record this information objectively. These figures can be the subject of high-profile controversies. More broadly, scholars' inability to generate rigorous casualty estimates prevents them from describing and analyzing the defining feature of armed conflict.¹

¹Dickey Center for International Understanding, Dartmouth College, Hanover, NH, USA

Corresponding Author:

Jeffrey A. Friedman, Dickey Center for International Understanding, Dartmouth College, 6048 Haldeman Center, Hanover, NH 03755, USA.

Email: jeffrey.a.friedman@dartmouth.edu

One of the main difficulties of estimating conflict size is that this is not just a matter of carefully gathering information. No matter how much effort analysts put into recording violence, the available record will almost always be incomplete, and thus, there will almost always be a need to estimate how comprehensively these data have been recorded. This article explains how power laws can be used to perform this kind of inference. A power law is a special probability distribution characterizing the relationship between the frequency and severity of many phenomena, including violent events. The notion that violent events often approximate power laws is a well-documented empirical regularity. This article instrumentalizes that fact to draw inferences about what the distribution of violent events might look like if the data were complete. In other words, the technique advanced here leverages the *distribution* of available data to evaluate the *comprehensiveness* of those data.

To demonstrate that technique, this article uses power laws to estimate Native American and US casualties during the American Indian Wars from 1776 to 1890.² There are two main reasons to focus on this experience. First, despite the central importance of these conflicts to North American history, there are no reliable estimates of how much fighting actually took place. Millions of Native Americans died as Europeans colonized North America and the tribes were subjected to coercive expansion by settlers and their governments. At the same time, most population loss resulted from disease, and some historians have recently argued that the frontier was far less violent than commonly believed.³ The historical significance of this violence is out of proportion with scholars' understanding of what it entailed.

Second, the American Indian Wars provide a context where estimating conflict size relies heavily on extrapolation. The US Army kept records of fighting with the tribes, and other sources help to flesh out this information. But it is highly unlikely that scholars can reliably estimate the severity of the American Indian Wars simply by accumulating data. Many violent events—especially small-scale engagements with little historical salience per se—have presumably gone unrecorded, becoming effectively invisible to event-count methodologies. Basing casualty estimates on the available record thus requires drawing inferences about how incomplete that record is.

The approach developed here is designed to do this. It indicates that roughly 50,000 Native Americans and roughly 12,000 US forces were killed, captured, or mortally wounded during the American Indian Wars. Combined, this exceeds the recorded figures for 90 percent of intrastate wars in the Correlates of War (COW) data.⁴ This historical contribution reinforces the article's methodological purpose, which is to demonstrate how power laws can be used to estimate conflict size, even (and perhaps especially) where previous literature has been unable to do so.

The article proceeds in five sections. The first section describes the American Indian Wars and introduces original data on 2,537 military engagements between Native American and US forces from 1776 to 1890. These are the most comprehensive event-level data on the American Indian Wars, but they doubtlessly remain incomplete. The second section explains how power laws can be used to estimate the comprehensiveness of these data. The third section implements this technique,

the fourth section discusses uncertainty surrounding the resulting estimates, and the fifth section the author's conclusions.

Data on the American Indian Wars

When the United States declared independence, most of its population lived near the Atlantic seaboard and the country had limited military resources. Settlers on the early frontier were vulnerable to attacks from nearby tribes, who in turn suffered frequent encroachments. These tensions regularly produced low-grade violence and sometimes spiraled into larger conflicts. In some cases, pan-tribal military alliances opposed the United States, leading to major engagements such as the 1794 Battle of Fallen Timbers and the 1811 Battle of Tippecanoe.

During the first half of the nineteenth century, the United States coercively relocated dozens of tribes to lands west of the Mississippi River. In some cases (as with the Choctaws), removal entailed relatively little violence, but in other cases (as with the Seminoles), tribes fought protracted conflicts to stay on their land. As US settlement expanded by mid-century, conflict ensued with prominent tribes such as the Cheyennes and Comanches, as well as with numerous, lesser-known tribes such as the Chetcos and Kalispels. Native Americans were increasingly confined to reservations during this period. Several tribes (such as Navajos and Nez Percés) forcibly resisted the reservation policy, and some (such as Apaches and Sioux) engaged in armed conflict when factions left reservations to live, hunt, or raid elsewhere. Some battles took place within the reservations themselves. In December 1890, the US Army fought Sioux conducting religious ceremonies near Wounded Knee Creek on a reservation in South Dakota. This is typically accepted by historians as marking the end of the American Indian Wars.⁵

The American Indian Wars thus span a wide range of time, geography, and participants. Clodfelter (2008) divides this experience into forty separate conflicts; Axelrod (1993) chronicles thirty-six "Indian Wars" after 1776. But as these authors acknowledge, dividing the period into discrete episodes is misleading, since much of the armed conflict between the United States and Native Americans involved protracted, low-level violence throughout the broader course of US expansion that took more than a century to complete.⁶ That movement constituted a major geopolitical shift: the United States emerged as a continental power, while previously hegemonic tribes lost vast amounts of people and land. Yet if violence played a leading role in precipitating this shift, there are no reliable estimates of how much violence actually transpired.

This article thus attempts to estimate the number of casualties—defined as the number of people killed, captured, or mortally wounded in battle—that occurred during the American Indian Wars, both on the side of the Native Americans and on the part of the United States. The analysis includes military engagements between 1776 and 1890 that occurred within the continental United States or involved pursuits into neighboring territory (such as entering Mexico to capture Geronimo). The

analysis includes armed engagements of any size, involving regular or militia forces. Data include noncombatant casualties, but not losses from displacement or disease. As in many conflicts, it is sometimes ambiguous whether a particular event should be seen as belonging to “the American Indian Wars” as opposed to interpersonal violence that was essentially nonpolitical. This study approaches the issue inductively by including information from a broad range of sources, and thus letting the sources “say” which engagements belong in the data. To the extent that there is disagreement, the present study thus errs on the side of inclusion, while following the lead of the literature on which it aims to build.

Data collection comprised several anthologies recording armed conflict at the level of individual military engagements. The most comprehensive is Webb’s (1939) *Chronological List of Engagements between the Regular Army of the United States and Various Tribes of Hostile Indians*, which lists 1,177 engagements between 1790 and 1890. Webb’s book is itself a complication of official records produced by the US Army Adjutant General and the US Army War College Historical Section.⁷

Five additional sources flesh out the data. Michno’s (2003) *Encyclopedia of Indian Wars* describes 787 engagements occurring after 1850; a follow-on work, *Forgotten Fights* (Michno and Michno 2008) adds another 334 engagements dating to 1823. Axelrod’s (1993) *Chronicle of the Indian Wars*, Ratjar’s (1999) *Indian War Sites*, and Nunnally’s (2007) *American Indian Wars* all survey violence dating from 1776, covering 123, 559, and 940 engagements, respectively. These anthologies include transparent sourcing; consistent information on the date, location, casualties, and tribes involved in each engagement; and broad temporal and geographic coverage. In all, these data incorporate 3,920 event reports covering 2,537 separate engagements, recording 25,643 casualties sustained by eighty-six different Native American tribes⁸ and 10,476 casualties sustained by US forces.⁹ Yet there is surely a substantial amount of missing data here, and this impedes drawing inferences about the overall severity of the American Indian Wars.

How can scholars draw this kind of inference? We could examine the data and see if the results “look right,” but this requires making assumptions about the quantity we are trying to estimate. Polling subject matter experts poses a similar difficulty—since the magnitude of the American Indian Wars is the subject of scholarly debate, expert opinions are bound to differ, and the very purpose of this project is to help resolve that disagreement by producing an independent, objective estimate. For assessing recent conflicts, scholars can sometimes use sampling methods to cross-check event counts, but this approach is not available for conflicts, like the American Indian Wars, that concluded long ago.¹⁰

In summary, it is unlikely that data collection alone can ever fully (or even remotely) tell us how many casualties occurred during the American Indian Wars. Analysts must ultimately make some assessment of how comprehensive their data are, and thus how the sample of available information compares to the overall population of interest. The next section outlines an approach to dealing with this

challenge, based on the well-documented finding that the severity of violence often follows a special kind of probability distribution called a power law.

Estimation Strategy

Power laws characterize the relationship between the frequency and the severity of certain events. If discrete data follow a power law, then the probability of seeing an observation with magnitude x is given by the distribution function $p(x) = Cx^{-\alpha}$, where C is a constant ensuring the distribution sums to 1, and α is the *scaling parameter*. A distinctive feature of power laws is that when they are represented on a “log–log” plot (with the logarithm of the event’s severity on the x -axis and the logarithm of the probability of an event being at least that severe on the y -axis), the data will form a straight line.¹¹

Power laws characterize many phenomena. The sizes of cities, earthquakes, moon craters, and annual incomes have all been represented using power laws (Newman 2005). Richardson (1948) originally observed that violent events, ranging from homicides to world wars, also seem to follow this pattern. Power laws characterize the size of interstate wars (Cederman 2003) and terrorist attacks (Clauset, Young, and Gleditsch 2007). Bohorquez et al. (2009) plotted data on the severity of insurgent attacks within nine separate conflicts; in each case, the data resembled power laws. These findings have important implications. As Clauset, Young, and Gleditsch (2007, 59) explain with respect to their findings on terrorism: “The frequency-severity statistics of terrorist events [demonstrate that] there is no fundamental difference between small and large events; both are consistent with a single underlying distribution. This fact indicates that there is no reason to expect that major or more severe terrorist attacks should require qualitatively different explanations than less salient forms of terrorism.”

In general, the notion that power laws characterize violent events is a well-known empirical regularity, and this has received substantial attention in literature intended for broader audiences.¹² Figure 1 demonstrates that event-level data on the American Indian Wars also resemble a power law. Panels A and B plot the (logged) severity of armed engagements on the x -axis along with the (logged) probability that a randomly chosen engagement is at least that severe on the y -axis; Panel A represents casualties sustained by Native American forces, Panel B represents casualties sustained by US forces; both plots are approximately linear.

Using Power Laws to Extrapolate Missing Data

Yet even if these plots resemble linear relationships, they are not linear exactly. In particular, the left tail of the data in Panel A, representing casualties sustained by Native Americans, “sags” below what we would expect to see if the distribution followed a power law. The data contain noticeably fewer small-scale events than

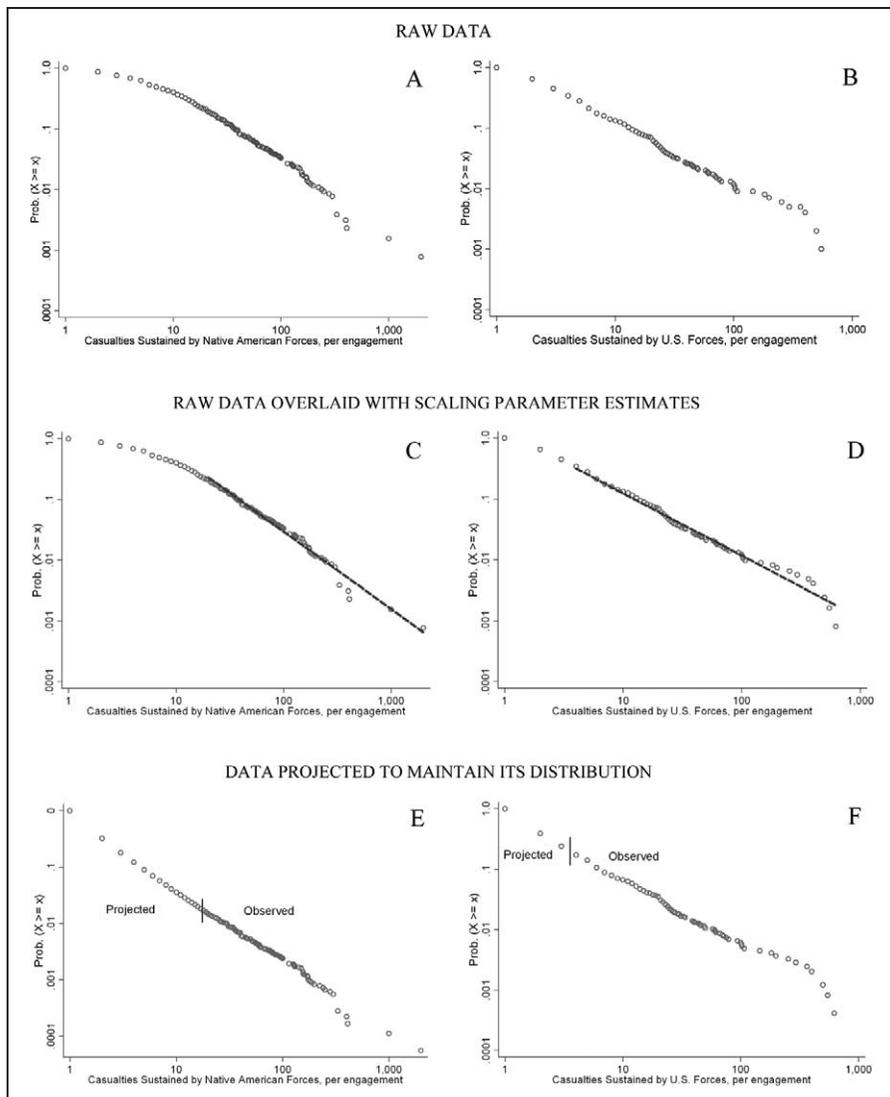


Figure I. Power laws in data on the American Indian Wars.

what a power law would predict, which is consistent with the expectation that these engagements are least likely to have been comprehensively recorded.

Panel B, representing data on US casualties, is not as noticeably concave on the left tail. This again makes sense, given that US casualties are more likely to have been documented at the time and preserved by available sources (especially since those sources include official US Army records). Together, Panels A and B

reinforce the expectation that when data on violent events are better measured, they will more closely approximate power laws.

This suggests a strategy for estimating the comprehensiveness of conflict data. Given a wide range of previous scholarship (and the way that the data themselves shape up), we can assume that the distribution of violent events during the American Indian Wars can be approximated with a power law. In addition, we expect that the data are relatively comprehensive when it comes to capturing high-intensity violent events, since those are most likely to have been recorded and preserved. If these assumptions are reasonable, then we can use the data on the right side of these distributions in order to develop an estimate of what we should see on the left if the data were complete.¹³

Figure 1 demonstrates this estimation strategy. Panels A and B provide scatterplots of the raw data. These data resemble power law distributions, and Panels C and D show how we can estimate their scaling parameters using a method that the next section describes in more detail. Panels E and F then show what the plots would look like if we added observations, so that the distributions maintained the same slopes throughout their left tails. Panels E and F are thus stylized projections of what these data might look like if they were complete, and we can use these projections to estimate how much information is missing from the original data.

Evaluating Key Assumptions

How plausible are the key assumptions driving this estimation strategy? The first of these assumptions, that data on the most violent events during the American Indian Wars are likely to be well-measured, seems fairly safe. Events like Andrew Jackson's Battle of Horseshoe Bend against the Creeks (1814), or George Custer's last stand against the Sioux (1876) were high profile, widely known and discussed at the time. For most of the period studied here, the US Army was small and soldiers did not expect to participate in much violence.¹⁴ Large battles were salient events, and since the data used in this study draw on Army records, it is doubtful that the right tail of this distribution is missing many observations. It is hard to say just how large an event would need to be before it would reliably appear in these data, but the technique described in the next section does not require making assumptions about where observed data begin deviating from a power law—this is something we can estimate directly.

The second key assumption driving the estimation strategy, that the severity of small-scale violent events should follow the power law which characterizes the rest of the distribution, is more contentious.¹⁵ Of course, most statistical estimation strategies rely on assumptions about probability distributions and functional forms; the operative question is not whether an empirical model is perfect, but whether its assumptions are reasonable for drawing inferences. The fact that so many different kinds of violent events resemble power laws throughout most of their distributions suggests that this is a reasonable baseline to use. And the fact

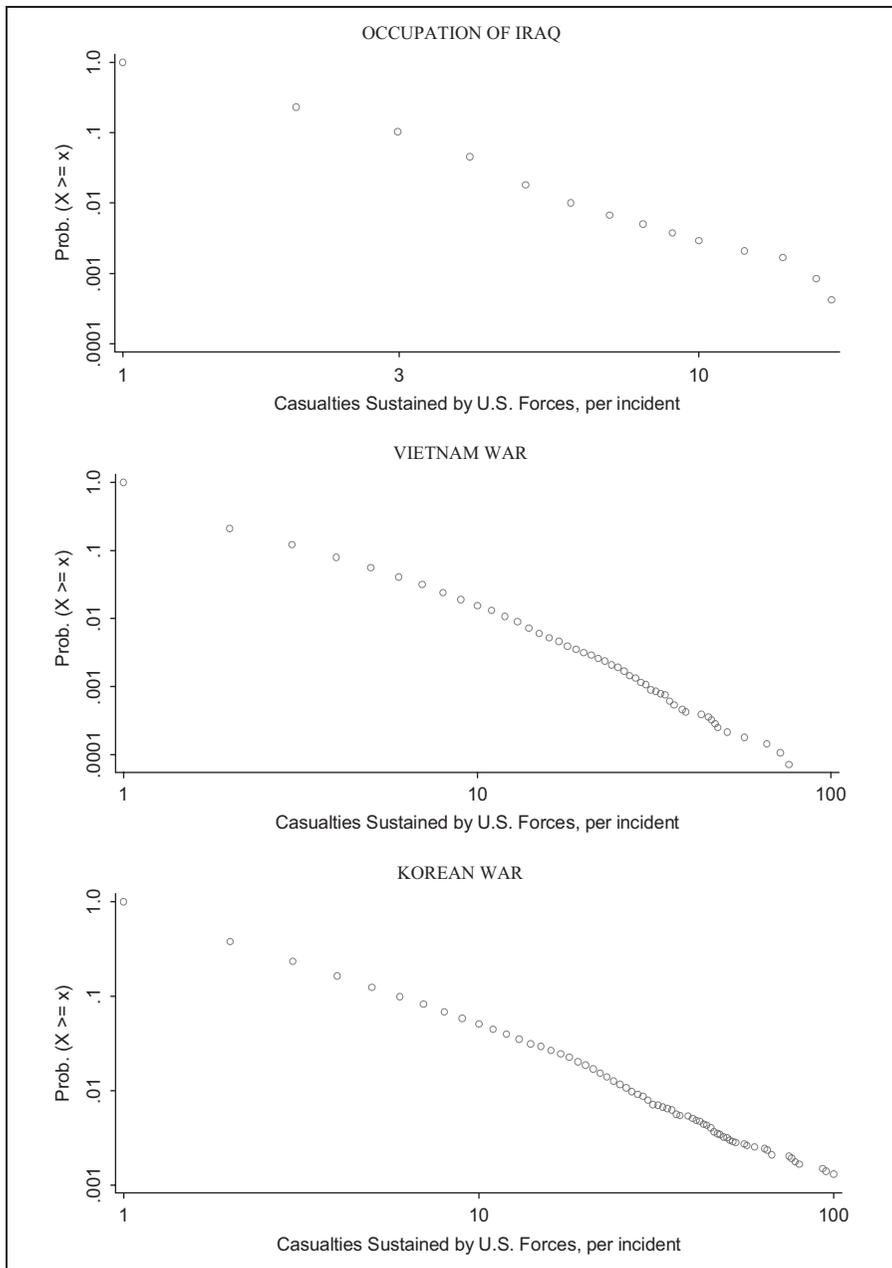


Figure 2. Distribution of US casualties by incident in Iraq, Vietnam, and Korea.

that data on US casualties in the American Indian Wars conform to a power law down to events of much lower magnitude than the data on Native American casualties reinforces the idea that better measured data should fit a power law more closely. Ideally, however, we would be able to go beyond suggestive evidence and directly examine this assumption.

To do so, we need data on violent events that are known to be complete (or come as close as possible to this standard), and for which casualties can be grouped by discrete incidents. Figure 2 plots three data sets that meet these conditions, representing US fatalities in Iraq, Vietnam, and Korea. For each of these conflicts, the US government (and for Iraq, media and nonprofit organizations) documented combat-related deaths in a manner that we can expect to be essentially comprehensive. Each data set provides not just the name of each soldier who died, but also their military unit, the type of incident leading to their death, and the location and date where this incident occurred. This is not the same as event-level data, but these individual-level factors can be used to determine which soldiers were killed as a result of incidents that occurred at the same place and time.¹⁶

All three data sets follow power laws throughout the extent of their left tails, as using the method described in the next section, we will estimate that the proper place to begin estimating the power law is for events causing one casualty or more. For that reason, the estimation technique advanced in this article will project that there are zero casualties missing from the data, and the estimated numbers of US fatalities in Iraq, Vietnam, and Korea exactly match the observed body counts of these wars, which we expect to be nearly comprehensive.

The distributional assumption driving this article's estimation strategy thus holds up well to a direct test, but of course this does not establish that the estimation strategy will work in all cases; perhaps there are some contexts where we would expect data on small-scale events to deviate substantially from the rest of the distribution, and no empirical phenomenon will conform precisely to a stylized model. Yet at the very least, we can say that the key assumptions invoked in this article are supported by clear empirical evidence. There is certainly more reason to think that we can approximate violence data with power laws than to think that error terms resemble normal distributions or that marginal effects are linearly additive in empirical studies of armed conflict. The next section describes in more detail how this assumption can be leveraged to estimate conflict size.

Implementation and Results

The implementation here is described specifically for casualties sustained by Native Americans. The same procedures were repeated for estimating the number of casualties sustained by US forces. The implementation employs statistical techniques developed by Clauset, Shalizi, and Newman (2009, henceforth CSN).¹⁷ This section leverages those techniques to provide a new approach for estimating conflict size.

Table 1. Extrapolating Data on Native American and US Casualties.

Panel A: Observed versus projected data, Native American casualties

Estimated x_{\min} : 20Estimated α : 2.21

Estimated total casualties sustained by Native American forces: 53,361

Casualties	Observed N	Projected N	Estimated missing observations	Estimated missing casualties
1	166	12,111	11,945	11,945
2	138	2,618	2,479	4,959
3	107	1,068	961	2,882
10	48	75	26	257
15	34	30	-3.5	-53
19	6	18	12	230
<i>Total</i>	1,297	17,976	16,679	27,718

Panel B: Observed versus projected data, US casualties

Estimated x_{\min} : 4Estimated α : 2.00

Estimated total casualties sustained by U.S. forces: 11,889

Casualties	Observed N	Projected N	Estimated missing observations	Estimated missing casualties
1	430	1,490	1,060	1,060
2	247	373	126	251
3	132	166	34	101
<i>Total</i>	1,233	2,075	1,220	1,413

See the fourth section for a discussion of statistical uncertainty surrounding these estimates.

We begin by estimating the scaling parameter. In order to do this, it is important to specify the range over which this parameter should be estimated. CSN present a maximum likelihood approach to determining the cutoff point, x_{\min} , below which the data do not conform to a power law, along with α , the scaling parameter characterizing all $x \geq x_{\min}$.¹⁸ Using this method, the fitted x_{\min} for data on Native American casualties is twenty and the scaling parameter characterizing the power law distribution above and including this point is $\alpha = 2.21$.

Carefully defining x_{\min} serves an important function beyond maximizing model fit. As described previously, the central thrust of the estimation technique developed here is to use observed data on the right side of a distribution (which generally conform to power laws) in order to characterize unobserved events on the left side of a distribution (where the data generally do not conform to power laws, but

are presumably undermeasured). CSN's method for defining x_{\min} objectively indicates where we should divide the data for this purpose. Once we have done so, and estimated the scaling parameter characterizing points $x \geq x_{\min}$, we can estimate how many data are missing for events of size $x < x_{\min}$.¹⁹

Table 1 (Panels A and B) presents the results. There are presumably a large number of missing observations for events that led to few Native American casualties—in fact, more than 90 percent of the events projected to be missing from the observed data correspond to engagements causing three Native American casualties or fewer. This is consistent with the expectation that these events would not only have been the most common, but also the most poorly recorded and preserved. Together, these projections require adding 19,786 casualties to the observed data.

The projected volume of missing data tapers quickly, however. When it comes to events causing ten Native American casualties, for example, Table 1 (Panel A) shows that we will only estimate about 26 missing engagements (implying that there are roughly 260 missing casualties associated with events of this size). Events causing 15 Native American casualties are actually overrepresented in the observed data (perhaps because this is a number to which observers or historians would have rounded uncertain estimates), and so we have to remove about 50 observed casualties here to make the data consistent with projected values.

In total, the method described here estimates that the data set is missing 16,678 engagements, which would have led to 27,718 Native American casualties. Adding these to observed figures, we can estimate that 53,361 Native Americans were killed, captured, or mortally wounded during the American Indian Wars between 1776 and 1890.²⁰

Table 1 (Panel B) also shows results of repeating these procedures to estimate casualties sustained by US forces. These data are still far from being complete—overall, they may be missing as many engagements as they record. Nevertheless, given that most missing engagements are presumably small in size, we can project that there are fewer than 1,500 US casualties missing from the data and that the US Army sustained 11,889 total casualties during the American Indian Wars.

Uncertainty and Sensitivity Analysis

One advantage of the estimation technique described here is that it is possible to characterize explicitly the uncertainty associated with estimated parameters along with how this affects resulting projections. This is generally not possible for casualty estimates that rely on expert opinion or secondary literature.²¹

Nonparametric bootstrapping is the logical approach. There are 1,297 engagements in the data set that resulted in recorded casualties sustained by Native American forces and 1,233 engagements in the data set that resulted in recorded casualties sustained by US forces. We can generate “bootstrapped” data sets of the same size by randomly sampling with replacement from the

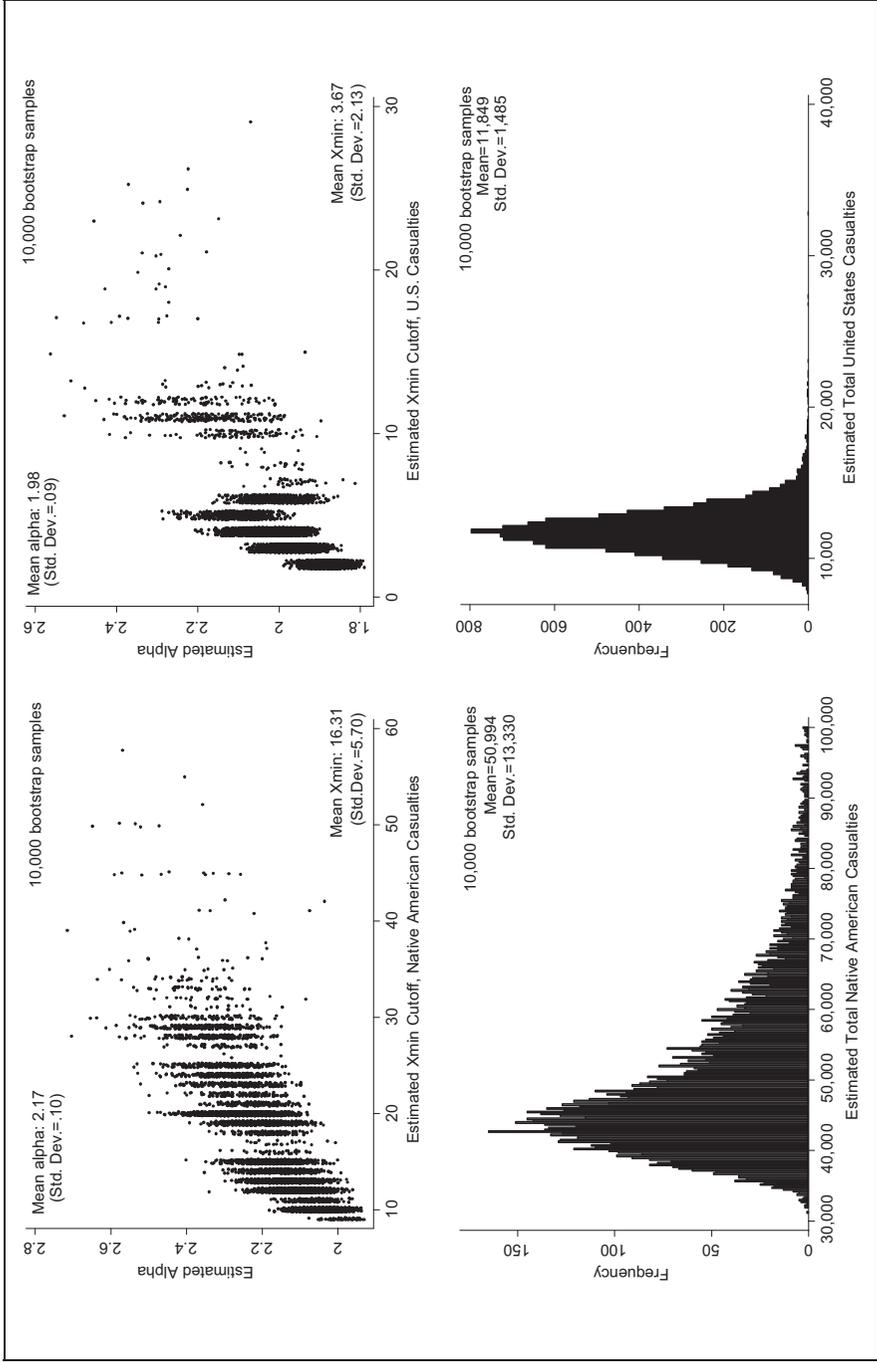


Figure 3. Bootstrapped estimates of X_{min} , α , and total casualties.

observed data. For each bootstrap sample, we can estimate maximum likelihood values for α and x_{\min} , resulting in the joint distributions shown in the top panels of Figure 3. Then for each pair of parameters, we can project the total number of casualties sustained by Native American and US forces, resulting in the histograms at the bottom of Figure 3.²² As expected, Figure 3 shows that there is substantially more uncertainty associated with estimating Native American casualties.

As shown in the bottom of Figure 3, bootstrap samples return mean casualty estimates that are close to the projections provided earlier: the mean estimate of 50,994 Native American casualties is 4 percent less than the projection given in Table 1 (Panel A), and the mean estimate of 11,849 US casualties is within 1 percent of the baseline model. Across 10,000 bootstrap samples, the standard error for estimates of total Native American casualties is 13,330, and the standard error for estimates of total US casualties is 1,485.

Analyzing the Data by Subset

Given that the American Indian Wars spanned a wide range of time, space, and actors, it makes sense to examine whether results would be meaningfully different if the estimation strategy were applied to subsets of the data rather than to all of them at once. In Figure 4, for example, the data are plotted for each of three periods—1776 to 1814, 1815 to 1864, and 1865 to 1890. These time periods correspond to logical historical breakpoints. Britain played a prominent role in backing tribes who fought the United States between the War of Independence and the War of 1812. After the Treaty of Paris, Britain curtailed this military support, and the continent's interior was opened to rapid US settlement (along with escalating demands for resettling tribes). Once the Civil War ended, the United States had a greatly expanded military infrastructure that it used to confine tribes to reservations. These periods thus had different political and military dynamics. Yet when the estimation strategy developed in this article is applied to each subset individually and the results are aggregated together, the projection of 55,689 total Native American casualties is just 4 percent larger than the output of the full-sample analysis. The projection of 13,669 US casualties is 15 percent larger than the original estimate.

Similar procedures were used to analyze data divided into Eastern, Plains, and Western regional subsets.²³ These regions varied by topography in ways that influenced military behavior: in the east, for instance, tribes often fought on foot using woodlands for cover and concealment, while tribes on the open Plains were more likely to fight on horseback, and armed conflict in the West typically took place on more rugged terrain. Tribes living in different parts of the continent also had systematic social and cultural differences; for example, tribes living on the Pacific Coast tended to have relatively decentralized social structures compared to those in the East. Yet when these regional subsets are analyzed

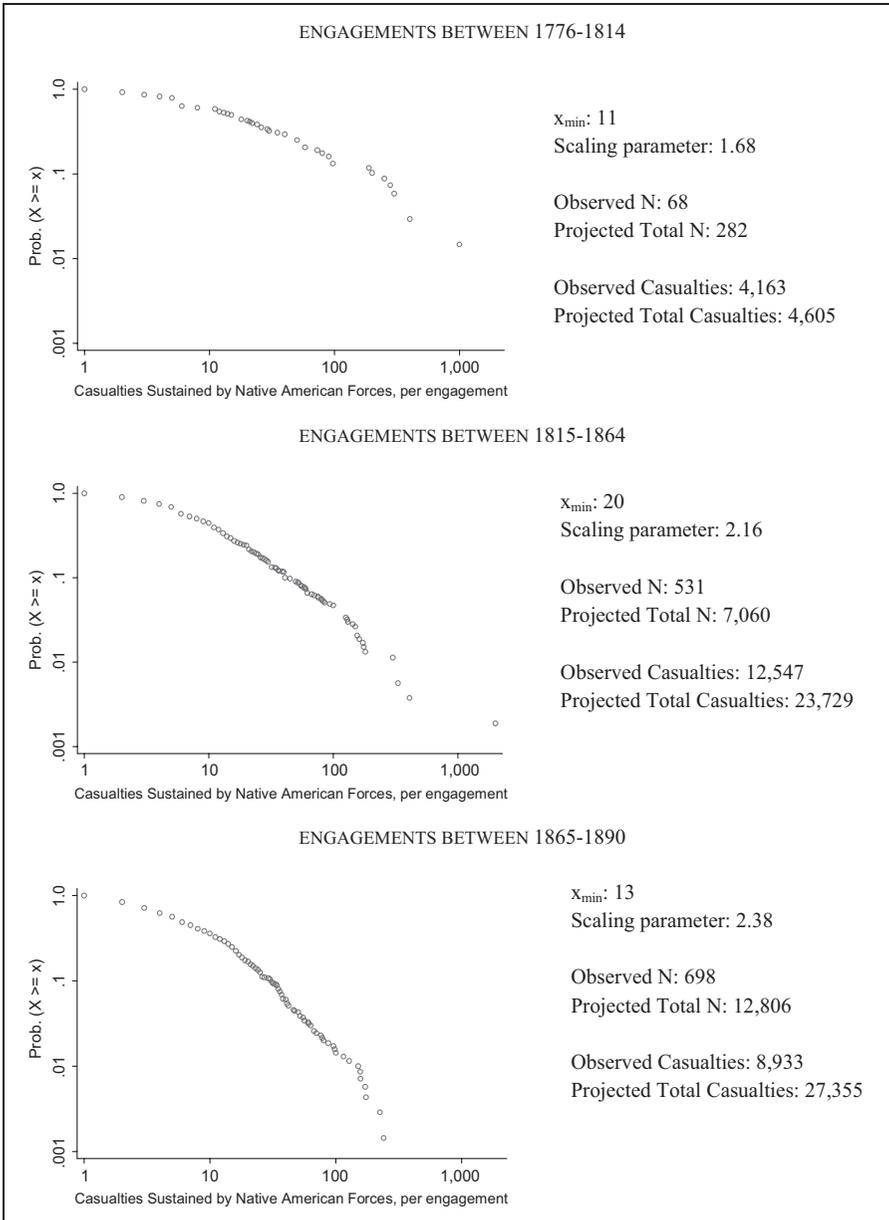


Figure 4. Analysis of period subsets.

Table 2. Summary of Sensitivity Analyses.

	Estimated Native American casualties	Estimated US casualties
	(with % deviation from baseline)	
Baseline model (parameters of best fit)	53,361	11,889
Mean of bootstrapped samples	50,994	11,849
... minus one standard deviation	-4%	-0.30%
	37,664	10,364
... plus one standard deviation	-30%	-13%
	64,324	13,334
	+20%	+12%
Data analyzed by ... temporal subsets	55,689	13,669
	+4%	+15%
... regional subsets	55,507	13,323
	+4%	+12%
... population subsets	62,122	11,902
	+16%	+0.1%

separately and the results are aggregated together, they again return estimates near to the original figures. The overall projections of 55,507 Native American casualties and 13,323 US casualties are 4 percent and 12 percent higher than the baseline estimates, respectively.

A third way to divide the data is based on the population of the tribes involved in each military engagement. The 86 tribes in the data set varied widely on this score, from tribes with fewer than 500 members (such as the Modocs and Wallawallas) to those with more than 10,000 (such as the Creeks and Cherokees). The mean tribe in these data had a population of about 5,000 (the standard deviation is roughly 4,000).²⁴ There are several reasons to think that tribes of different sizes might have varied in their military behavior, with the most obvious being that larger tribes would have been more able, all else being equal, to sustain and inflict larger numbers of casualties. If we divide the data into quartiles based on the populations of the tribes involved in each engagement, analyze each of these subsets individually and then aggregate the estimates together, then this projects a total of 62,122 Native American casualties during the American Indian Wars, 16 percent higher than the original estimate. The projection for US casualties based on these subsets of the data is 11,902, almost identical to the baseline projection.

Table 2 summarizes these estimates. All are within 30 percent of the baseline projection, and many are much closer. This stands in sharp contrast with extant debates about the sizes of individual armed conflicts, where estimates often range by an order of magnitude or more.²⁵

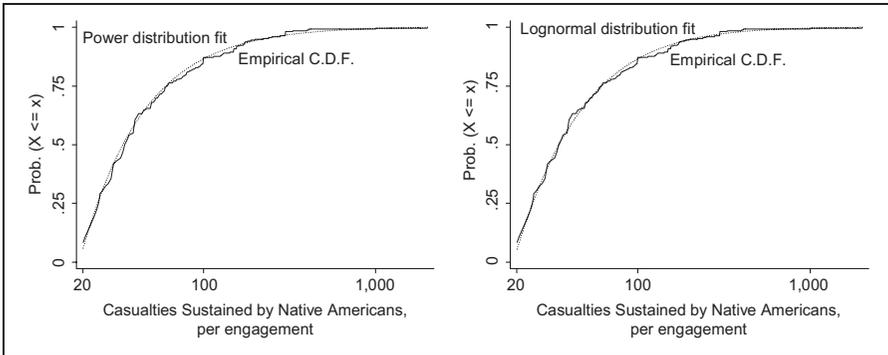


Figure 5. Comparing power law and lognormal models.

Considering Alternative Distributions

Another way to evaluate how power laws characterize the data is to examine this fit relative to other distributions. As CSN explain, several alternatives can easily be mistaken for power laws when they take on certain parameters. In examining twenty-four phenomena that have been claimed to follow power laws, CSN found that it is especially difficult to determine whether data are more representative of power law or lognormal distributions Clauset et al. (2009, 679-89).

Figure 5 draws a similar comparison for the Native American casualties data used in this article, showing how the power law and lognormal fits are hard to distinguish.²⁶ If we measure model fit by mean squared error (predicted vs. actual observations at each magnitude), the power law and the lognormal perform identically to two decimal places.²⁷ If we compare model fit with a likelihood ratio test, then the fits are similarly indistinguishable ($p = .93$).²⁸ And even though these distributions fit the data similarly well, they have noticeably different implications: since the lognormal distribution is thinner than a power law on its left tail, it predicts that there are fewer missing observations, leading to an estimate of 36,892 Native American casualties. This is 31 percent less than the 53,361 Native American casualties projected using power laws. Substituting a lognormal fit makes less of a difference for estimating US casualties: the resulting estimate of 11,502 US deaths is 3 percent less than the estimate based on a power law. This is what we would expect given that the x_{\min} cutoff for the US data is much lower; to the extent that the lognormal and power law distributions have differently shaped left tails, a lower cutoff gives these divergences less room to affect overall projections.

What should we make of these comparisons? It will almost always be possible to make plausible estimates based on alternative distributions; ultimately, there are an infinite number of distributions to choose from, and for any given empirical data set, some are bound to fit better than a power law. With that said, the power law is

appealing because it is simple (requiring analysts to estimate only one parameter), it is generalizable (as scholars have found that violence data appear to conform to power laws widely, both within and across conflicts), and we have seen that when data are better measured then they approximate power laws more closely. At the very least, this seems to be a reasonable starting point for the analysis. And even if this section has demonstrated that varying distributional assumptions can cause estimates of Native American casualties to change by 30 percent, this uncertainty still pales in comparison to the vastly different figures often given in scholarly and public debates about conflict size. The following section concludes by drawing connections to these broader issues.

Discussion

This article offers two main contributions: one is historical, speaking to knowledge about the American Indian Wars in particular, and the second is methodological, speaking to broader academic questions about estimating conflict size.

The principal historical relevance of this article is to provide the first systematic casualty estimates for the American Indian Wars. These conflicts are central to understanding how the United States became a continental power, yet historians do not know how violent this period actually was. This is largely because any historical analysis of the subject will inevitably confront missing data and necessitate drawing difficult inferences. This article offers a novel approach to drawing those inferences based on empirical regularities that characterize armed conflict generally. This technique suggests that roughly 50,000 Native Americans and roughly 12,000 US forces were killed, captured, or mortally wounded during the American Indian Wars.

To put these figures in perspective: if US forces sustained 12,000 casualties in the American Indian Wars, then this would be about as large as the totals from the War of Independence, the War of 1812, the Mexican–American War, and the Spanish–American War combined.²⁹ This is roughly a third of combat losses US forces sustained in Korea, and about twice the number of losses sustained during the occupations of Iraq and Afghanistan, as of this writing. And these comparisons are all in absolute terms. Per capita, the American Indian Wars are even more pronounced as one of the most violent periods in US history.

For Native Americans, these conflicts exacted a far greater toll. Fifty thousand combat casualties are roughly what US forces sustained in Vietnam, and again, the populations in question are far different. From 1776 to 1890, the population of Native Americans living in the continental United States averaged roughly 400,000 (Reddy 1993). Against this reference point, a figure of 50,000 casualties is massive, and it is unclear how many populations have sustained more battle deaths per capita. Obviously, it is difficult to draw direct comparisons in terms of aggregate counts because losses during the American Indian Wars were distributed across so much time and so many actors, and thus, it is also instructive to describe the intensity

of this violence. As mentioned earlier, the average tribe enumerated in the data set had roughly 5,000 members at the time of its first engagement with the United States. Fifty thousand Native American casualties, distributed across the 86 tribes and 115 years covered in this analysis, amounts to approximately one of every thousand Native Americans killed per annum—for the United States today, this would be equivalent to about 300,000 annual fatalities.

Of course, we do not need to invoke power laws to establish that the American Indian Wars were costly, especially for Native Americans—this is common knowledge. But from a historical perspective, it is useful to have an objective sense of relevant magnitudes. For instance, Pinker (2011, 195) cites a figure of 20 million Native American deaths during his discussion of the “annihilation of the American Indians” as being one of the most violent phenomena in world history, but the vast majority of this population loss was a result of disease and not armed conflict. It is important to understand the impact of these diseases, but they were largely independent of the fighting that took place during the American Indian Wars, and the resulting population loss was so huge that it can obscure the military aspects of the conflict, which were important and devastating in their own right. This article addresses the military costs of the American Indian Wars directly, and this is its main historical contribution.

More generally, this article has offered a new method for estimating conflict size, a subject that is often controversial within scholarship and public debates. During the occupation of Iraq, for instance, a study published in *The Lancet* (Burnham et al. 2006) used survey methods to estimate that roughly 600,000 civilians had died as a result of the war. This figure was an order of magnitude higher than contemporary estimates provided by the US government. The *Lancet* article received substantial attention, as well as rebuttals from both the US government and academics.³⁰

Among scholars, these kinds of disagreements often go well beyond assessing individual cases, as prominent data sets measuring conflict size are often themselves the subject of dispute. For example, two recent studies (Obermeyer, Murray, and Gakidou 2008; Gohdes and Price 2013) argued that the most widely used data set on battle deaths since World War II (Lacina and Gleditsch 2005) had systematically underestimated the costs of recent conflicts; furthermore, both studies claimed that that correcting these estimates eliminates the perception that violence has been declining over time.³¹ These debates are thus not just about the relative merits of different estimation techniques per se, but about the existence of an empirical phenomenon that is central to international relations scholarship. This only highlights the fact that while violence is the defining feature of armed conflict, measurement issues impede scholars from analyzing it systematically on even the broadest dimensions.

This article offers a new way to approach such debates. The method described here is based on event counts, but it explicitly seeks to go beyond them, using the data’s distribution in order to draw inferences about their comprehensiveness. In order to explain the technique and demonstrate its utility, this article examined how

power laws can be used to estimate the severity of the American Indian Wars, but scholars can use the same methodology in order to evaluate the size of many other conflicts as well. In order to employ this technique, observed data must be relatively comprehensive when it comes to capturing large-scale events, but one of the appealing aspects of the estimation strategy is that it is specifically designed to deal with missing data for small-scale events. Political scientists can play an important role on this score by developing objective ways of going beyond historical material or government records, which are usually incomplete. It will almost always be necessary to supplement data that are available with inferences about information that has not been preserved. This article has shown how power laws can be used as a tool for that purpose.

Author's note

A previous version of this article was presented at the 2013 Annual Meeting of the American Political Science Association. Replication materials are available from the author and at the *Journal of Conflict Resolution* website, <http://jcr.sagepub.com/>.

Acknowledgment

Special thanks to Bear Braumoeller, Aaron Clauset, Allan Dafoe, Scott Gates, Colin Gillespie, Neil Johnson, Joshua Kertzer, Richard Nielsen, Arthur Spirling, Brandon Stewart, Benjamin Valentino, and two anonymous reviewers for providing helpful comments that greatly improved the manuscript. All errors are the author's.

Declaration of Conflicting Interests

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

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. On debates about casualty estimation in general, see Andreas and Greenhill (2010) and Seybolt, Aronson, and Fischhoff (2013).
2. Consistent with academic convention, this article will refer to the historical period as the "American Indian Wars" and otherwise use the terms "American Indian" and "Native American" interchangeably.
3. See Dykstra (2003) on this view and Michno (2003) for disagreement.
4. Though only one of those tribes—the Sioux—actually appears in the COW given the data set's threshold of violence required for inclusion. COW data span conflicts between 1815 and 2007. 174 of the 334 entries in this data set contain casualty estimates for more than one side. with missing or incomplete casualty figures are not included in the reference group, but no actor in these remaining conflicts is recorded as having sustained more

battle deaths than the Native Americans. Among the 431 combatants for which the COW records estimated battle deaths, only 17 sustained more than the Native Americans if the estimate advanced in this article is accurate.

5. Prominent histories of the American Indian Wars adopting a largely US-centric viewpoint include Prucha (1969), Utley (1967), and Utley (1973). Prominent sources taking tribes to be the principal units of analysis include Richter (2001), Calloway (2003), Blackhawk (2006), and Hämäläinen (2008).
6. As Axelrod (1993, 257) writes, “Wars between Indians and whites were rarely officially declared or officially concluded, and even when they ostensibly were, violence often preceded the official declaration and persisted after the official cessation of hostilities.”
7. These records certainly suffer omissions and inaccuracies, but they are generally well regarded. In a related discussion, Michno (2003, 353) writes, “The numbers of casualties used in this study are inherently biased toward army estimations, since it was the army who kept the records. [But . . .] there is reason to be confident that the army estimates were reasonably accurate . . . [T]he army lived by a strict code of honor, particularly in its official reports, and even if the unit leader tried to inflate numbers, he would face correction and perhaps ostracism by other soldier-witnesses.” Delay (2008) similarly discusses the potential accuracy of his event-level data covering engagements between Native Americans and Mexicans.
8. These tribes are Achumawi, Arapaho, Arikara, Bannock, Blackfeet, Caddo, Cahto, Cahuilla, Cayuga, Cherokee, Chetco, Cheyenne, Chickasaw, Chinook, Chippewa, Chiricahua Apache, Comanche, Coquille, Coeur d’Alene, Creek, Crow, Delaware, Fox, Goshute, Gros Ventre, Haida, Havasupai, Hupa, Jicarilla Apache, Kalispel, Kansa, Karankawa, Kickapoo, Kiowa, Kitanemuk, Kitsai, Klamath, Koasati, Lipan Apache, Luiseno, Mescalero Apache, Miami, Mingo, Miwok, Modoc, Mohave, Mohawk, Navajo, Nez Perce, Nomlaki, Osage, Ottawa, Paiute, Palouse, Pawnee, Pomo, Ponca, Potawatomi, Quechan, Santee Sioux, Sauk, Seminole, Seneca, Shasta, Shawnee, Southern Coast Salish, Spokane, Squaxon, Tahltan, Takelma, Taos, Tenino, Teton Sioux, Tolowa, Tonkawa, Tututni, Umatilla, Umpqua, Ute, Walapai, Wallawalla, Western Apache, Wichita, Wintu, Yakama, and Yokut. Tribal designations were based on the Smithsonian Institution’s *Handbook of North American Indians* (Sturtevant 1978-), which pays extensive attention to enumerating tribal subgroups and synonymies.
9. In almost all cases, casualty counts are reproduced directly from the event reports. In cases where sources offered different estimates for the same engagement, those estimates were averaged together. There were thirty-one cases in which US casualties were listed as being “several,” or “a few,” or “a number”—in each instance, this number was approximated with a coding of 5, comprising less than 2 percent of total recorded US casualties. Seventy event reports were similarly interpolated for Native American casualties, again comprising less than 2 percent of total recorded casualties.
10. On sampling methods for estimating conflict size (including mortality surveys and multiple systems estimation), see Lum et al. (2010); Asher (2013); Klingner and Silva (2013); Manrique-Vallier, Price, and Gohdes (2013); and Jewell, Spagat, and Jewell (2013).

11. The logarithm of a power law's distribution function is $\log p(x) = \log C - \alpha \times \log(x)$, and thus any unit change in $\log(x)$ will reduce $\log p(x)$ by the value of the scaling parameter. Power law data are "scale invariant" in the sense that scaling the argument of the function by a constant always results in a proportional scaling of the value of the function. This is because $p(\beta x) = C(\beta x)^{-\alpha} = C\beta^{-\alpha}x^{-\alpha}$, where $C\beta^{-\alpha}$ is a constant.
12. For example, Hayes (2002), Pinker (2011), and Silver (2012).
13. Scholars use related techniques to estimate frequencies of extreme events in several fields. Earthquakes exceeding 7.0 on the Richter scale, for instance, are too rare to generate meaningful risk projections. In order to estimate the propensity for such extreme events, researchers rely on the notion that earthquakes are typically distributed according to power laws. Data on small earthquakes are thus used to project the potential incidence of large earthquakes. Rundle et al. (2012) explain how this technique can be applied to several phenomena; Clauset and Woodard (2013) use this approach to estimate the risk of large terrorist attacks. This article presents a similar estimation strategy, albeit in reverse: previous scholars have used power law relationships to estimate the potential incidence of extreme events that have yet to be observed. Using related logic, this article uses power law relationships to estimate the prior incidence of small-scale events that have gone unobserved.
14. See Prucha (1969) and Utley (1967, 1973).
15. This is especially because in some contexts where power laws appear, scholars believe there are theoretical reasons to expect deviation for small-scale events. Small earthquakes, for instance, are typically much less frequent than what power laws suggest. This may partly reflect undermeasurement, but there is also the possibility that if tectonic forces do not meet a certain minimum energy threshold, they may not release enough force to cause any earthquakes at all. Thus, there may be geologically determined constraints on how small an earthquake can be. In the context of armed conflict, there is no obvious theoretical reason why the size of events should be so constrained, but it is important to acknowledge that assuming the data will conform to the same distribution throughout the left-tail is something that is generally controversial in studying power laws.
16. Iraq data, drawn from <http://www.icasualties.org>, include information on 4,346 US fatalities that occurred during or as a result of serving in Iraq between the declared closing of major combat operations on May 2, 2003, and the formal withdrawal of US forces on December 18, 2011. Nine hundred nine entries were dropped because they were explicitly listed as occurring from noncombat related accidents or health conditions. Fatalities are considered to be part of the same "incident" if they occurred on the same day, in the same location, across members of the same unit, who are listed as being killed via the same incident type (e.g., improvised explosive device [IED] attack, indirect fire). For soldiers killed on the same day and in the same location but not listed as working for the same unit or being killed by the same incident type, cause of death was examined in the *Military Times*, which generally includes information on whether a particular soldier was killed alongside others. Soldiers listed as dying in military hospitals or outside Iraq were also cross-checked in this way. Data for Vietnam and Korea were drawn from the Defense Department's Defense Casualty Analysis System (DCAS:

<https://www.dmdc.osd.mil/dcas>). After removing fatalities explicitly listed as occurring due to noncombat-related accidents, health conditions, and suicides, there are 47,537 recorded US fatalities for Vietnam and 33,747 recorded US fatalities for Korea. DCAS data separate the date of death from the date of incidents that caused each death. If entries are listed as occurring for soldiers in the same unit, in the same location, from incidents occurring on the same date, then they are considered to have occurred as a result of the same “event” for the purposes of this analysis. Because it is not possible to combine multiunit incidents in the manner described for the Iraq data, these figures will undercount the large-scale events, and both distributions have been truncated at $x = 100$.

17. CSN provide code at <http://santafe.edu/~aaronc/powerlaws/>, accessed December 2012. Additional code developed for this project is provided with replication files.
18. This method considers each value of x to be a potential cutoff, calculates the maximum likelihood scaling parameter that fits a power law to the data from that point onward, and selects the pair (x_{\min}, α) that optimizes the Kolmogorov–Smirnov goodness-of-fit statistic over $x \geq x_{\min}$. This is the standard method scholars use to estimate power laws in armed conflict and other fields; for instance, see Bohorquez et al. (2009) and Bell et al. (2012).
19. If the data follow a power law with scaling parameter α over points $x \geq x_{\min}$, then the probability of a randomly chosen violent event causing x casualties is given by the function $f(x) = x^{-\alpha} \times \zeta(\alpha, x_{\min})^{-1}$, where ζ is the Hurwitz zeta function. To estimate the total number of violent events below the x_{\min} cutoff that we would expect to see if the data conformed to a power law throughout, multiply $f(x)$ by λ , where λ is the total number of observed events with magnitude $x \geq x_{\min}$. Used in this fashion, $f(x)$ no longer represents probability density, as its purpose is simply to establish proportionality between the frequency and severity of violent events. Note that this method can be implemented using any assumed probability distribution, as shown in the fourth section.
20. A Bayesian approach (suggested by Colin Gillespie in correspondence on this article) is to model missingness directly by assuming the data are distributed according to a power law and estimating simultaneously the scaling parameter along with the parameters of a function giving the probability that events of size x will be recorded. This approach explicitly accounts for the possibility that the data may be undermeasured at any value of x , though this approach is potentially sensitive to assumptions about the reporting function. Implementing this technique on an exploratory basis with the reporting function defined as a second-order polynomial resulted in estimates close to those presented here (e.g., an estimated 58,000 Native American casualties, with a standard error of roughly 7,000).
21. Gohdes and Price (2013) make this critique of the Peace Research Institute Oslo (PRIO) Battle Deaths data set; it is one reason why they and other authors recommend survey or multiple systems estimation approaches to estimating conflict size.
22. The histogram representing bootstrapped estimates of Native American casualties is truncated for clarity. There were eighty-six samples that produced casualty estimates greater than 100,000, with a maximum estimate of 236,475.
23. Each engagement was assigned to a region based on the “culture area” to which the tribes in the engagement belonged as recorded in the Smithsonian Institution’s

- Handbook of North American Indians*. The “Eastern region” in this analysis comprises the Northeast and Southeast culture areas, the “Plains region” comprises the Northern and Southern Plains culture areas, and the “Western region” comprises the Southwest, California, Great Basin, Plateau, and Northwest culture areas. In cases where the tribe involved in an engagement was unidentified, region was determined by the state or territory where fighting took place.
24. Population data are also from the Smithsonian Institution’s *Handbook of North American Indians*. For this analysis, a tribe’s population is coded based on estimates provided as close as possible to the initiation of contact between the tribe and the United States, as this is when the tribe would have entered the sample as having the potential to fight the US Army. In cases where the tribe participating in the engagement was unidentified, a population estimate was interpolated based on the average population of tribes that appeared in the data set within each particular culture area. All of these population estimates are inexact, which should add imprecision to the resulting estimates.
 25. See, for example, Burnham et al. (2006); Obermeyer, Murray, and Gakidou (2008); and Spagat et al. (2009).
 26. Each of the models here is estimated by optimizing the Kolmogorov–Smirnov statistic for data points above the $x_{\min} = 20$ cutoff, using CSN’s code.
 27. For the power law, it is 0.399 and 0.401 for the lognormal.
 28. Estimated via Vuong closeness test as recommended by CSN.
 29. Clodfelter (2008) gives the following totals: the War of Independence (6,824 battle deaths), the War of 1812 (2,260), the Mexican–American War (1,721), the Spanish–American War (385).
 30. See Fischer (2010, 9) for a tabulation of Iraq casualty estimates from a range of governmental and nongovernmental sources and Spagat (2010) for a critique of the *Lancet* article’s methodology.
 31. See Spagat et al. (2009) and Lacina and Gleditsch (2012) for responses to these critiques.

References

- Andreas, Peter, and Kelly Greenhill, eds. 2010. *Sex, Drugs and Body Counts: The Politics of Numbers in Global Crime and Conflict*. Ithaca, NY: Cornell University Press.
- Asher, Jana. 2013. “Using Surveys to Estimate Casualties Post-conflict.” In *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict*, edited by Taylor B. Seybolt, Jay D. Aronson, and Baruch Fischhoff, 97–122. New York: Oxford University Press.
- Axelrod, Alan. 1993. *Chronicle of the Indian Wars: From Colonial Times to Wounded Knee*. New York: Prentice Hall.
- Bell, Michael J., Colin S. Gillespie, Daniel Swan, and Phillip Lord. 2012. “An Approach to Describing and Analysing Bulk Biological Annotation Quality: A Case Study Using UniProtKB.” *Bioinformatics* 28 (18): i562–i568.
- Blackhawk, Ned. 2006. *Violence over the Land: Indians and Empires in the Early American West*. Cambridge, MA: Harvard University Press.

- Bohorquez, Juan Camilo, Sean Gourley, Alexander R. Dixon, Michael Spagat, and Neil F. Johnson. 2009. "Common Ecology Quantifies Human Insurgency." *Nature* 462: 911-14.
- 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.
- Calloway, Colin G. 2003. *One Vast Winter Count: The Native American West before Lewis and Clark*. Lincoln: University of Nebraska Press.
- Cederman, Lars-Erik. 2003. "Modeling the Size of Wars: From Billiard Balls to Sandpiles." *American Political Science Review* 97 (1): 135-50.
- Clauset, Aaron, Cosma R. Shalizi, and Mark E. J. Newman. 2009. "Power-law Distributions in Empirical Data." *SIAM Review* 51 (4): 661-703.
- Clauset, Aaron, and Ryan Woodard. 2013. "Estimating the Historical and Future Probabilities of Large Terrorist Events." *Annals of Applied Statistics* 7 (4): 1838-65.
- Clauset, Aaron, Maxwell Young, and Kristian Skrede Gleditsch. 2007. "On the Frequency of Severe Terrorist Events." *Journal of Conflict Resolution* 51 (1): 58-89.
- Clodfelter, Micheal. 2008. *Warfare and Armed Conflicts: A Statistical Encyclopedia of Casualty and Other Figures, 1494-2007*. Jefferson, NC: McFarland.
- Delay, Brian. 2008. *War of a Thousand Deserts: Indian Raids and the U.S.-Mexican War*. New Haven, CT: Yale University Press.
- Dykstra, Robert R. 2003. "Body Counts and Murder Rates: The Contested Statistics of Western Violence." *Reviews in American History* 31 (4): 554-63.
- Fischer, Hannah. 2010. *Iraq Casualties: U.S. Military Forces and Iraqi Civilians, Police, and Security Forces*. Washington, DC: Congressional Research Service.
- Gohdes, Anita, and Megan Price. 2013. "First Things First: Assessing Data Quality before Model Quality." *Journal of Conflict Resolution* 57 (6): 1090-108.
- Hämäläinen, Pekka. 2008. *The Comanche Empire*. New Haven, CT: Yale University Press.
- Hayes, Brian. 2002. "Statistics of Deadly Quarrels." *American Scientist* 90 (1): 10-15.
- Jewell, Nicholas P., Michael Spagat, and Britta L. Jewel. 2013. "MSE and Casualty Counts." In *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict*, edited by Taylor B. Seybolt, Jay D. Aronson, and Baruch Fischhoff, 185-212. New York: Oxford University Press.
- Klingner, Jeff, and Romesh Silva. 2013. "Combining Found Data and Surveys to Measure Conflict Mortality." In *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict*, edited by Taylor B. Seybolt, Jay D. Aronson, and Baruch Fischhoff, 147-164. New York: Oxford.
- Lacina, Bethany, and Nils Petter Gleditsch. 2005. "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths." *European Journal of Population* 21 (2-3): 145-66.
- Lacina, Bethany, and Nils Petter Gleditsch. 2012. "The Waning of War Is Real: A Response to Gohdes and Price." *Journal of Conflict Resolution* 57 (6): 1109-27.
- Lum, Kristian, Megan Price, Tamy Guberek, and Patrick 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): 1-26.
- Manrique-Vallier, Daniel, Megan E. Price, and Anita Gohdes. 2013. “Multiple Systems Estimation Techniques for Estimating Casualties in Armed Conflicts.” In *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict*, edited by Taylor B. Seybolt, Jay D. Aronson, and Baruch Fischhoff, 165-182. New York: Oxford University Press.
- Michno, Gregory F. 2003. *Encyclopedia of Indian Wars: Western Battles and Skirmishes, 1850-1890*. Missoula, MT: Mountain Press.
- Michno, Gregory F., and Susan J. Michno. 2008. *Forgotten Fights: Little-known Raids and Skirmishes on the Frontier, 1823 to 1890*. Missoula, MT: Mountain Press.
- Newman, M. E. J. 2005. “Power Laws, Pareto Distributions and Zipf’s Law.” *Contemporary Physics* 46 (5): 323-51.
- Nunnally, Michael L. 2007. *American Indian Wars: A Chronology of Confrontations between Peoples and Settlers and the United States Military, 1500-1901*. Jefferson, NC: McFarland.
- 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 Medical Journal* 336. Published electronically June 26, 2008. Accessed March 26, 2014. <http://www.bmj.com/content/336/7659/1482>.
- Pinker, Steven. 2011. *The Better Angels of Our Nature: Why Violence Has Declined*. New York: Viking.
- Prucha, Francis Paul. 1969. *The Sword of the Republic: The United States Army on the Frontier, 1783-1846*. New York: Macmillan.
- Ratjar, Steve. 1999. *Indian War Sites: A Guidebook to Battlefields, Monuments, and Memorials*. Jefferson, NC: McFarland.
- Reddy, Marlita A. 1993. *Statistical Record of Native North Americans*. Detroit, MI: Gale.
- Richardson, Lewis F. 1948. “Variation of the Frequency of Fatal Quarrels with Magnitude.” *Journal of the American Statistical Association* 43 (244): 523-46.
- Richter, Daniel K. 2001. *Facing East from Indian Country: A Native History of Early America*. Cambridge, MA: Harvard University Press.
- Rundle, John B., James R. Holliday, William R. Graves, Donald L. Turcotte, Kristy F. Tiampo, and William Klein. 2012. “Probabilities for Large Events in Driven Threshold Systems.” *Physical Review E* 86 (2): 021106.
- Seybolt, Taylor B., Jay D. Aronson, and Baruch Fischhoff, eds. 2013. *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflict*. New York: Oxford University Press.
- Silver, Nate. 2012. *The Signal and the Noise: Why Most Predictions Fail—But Some Don’t*. New York: Penguin.
- Spagat, Michael. 2010. “Ethical and Data-integrity Problems in the Second Lancet Survey of Mortality in Iraq.” *Defence and Peace Economics* 21 (1): 1-41.

- Spagat, Michael, Andrew Mack, Tara Cooper, and Joakim Kreutz. 2009. "Estimating War Deaths: An Arena of Contestation." *Journal of Conflict Resolution* 53 (6): 934-50.
- Sturtevant, William, and gen. ed. 1978-. *Handbook of North American Indians*, 20 vols. Washington, DC: Smithsonian Institution.
- Utley, Robert M. 1967. *Frontiersmen in Blue: The United States Army and the Indian 1848-1865*. New York: Macmillan.
- Utley, Robert M. 1973. *Frontier Regulars: The United States Army and the Indian 1866-1891*. New York: Macmillan.
- Webb, George W. 1939. *Chronological List of Engagements between the Regular Army of the United States and Various Tribes of Hostile Indians which Occurred During the Years 1790 to 1898, Inclusive*. New York: AMS Press.