

Has War Really Declined since World War 2?

by

Michael Spagat

Department of Economics

Royal Holloway University of London

m.spagat@rhul.ac.uk

and

Stijn van Weezel

Unaffiliated

Presentation made by Michael Spagat in Oslo, December 2018

Orientation

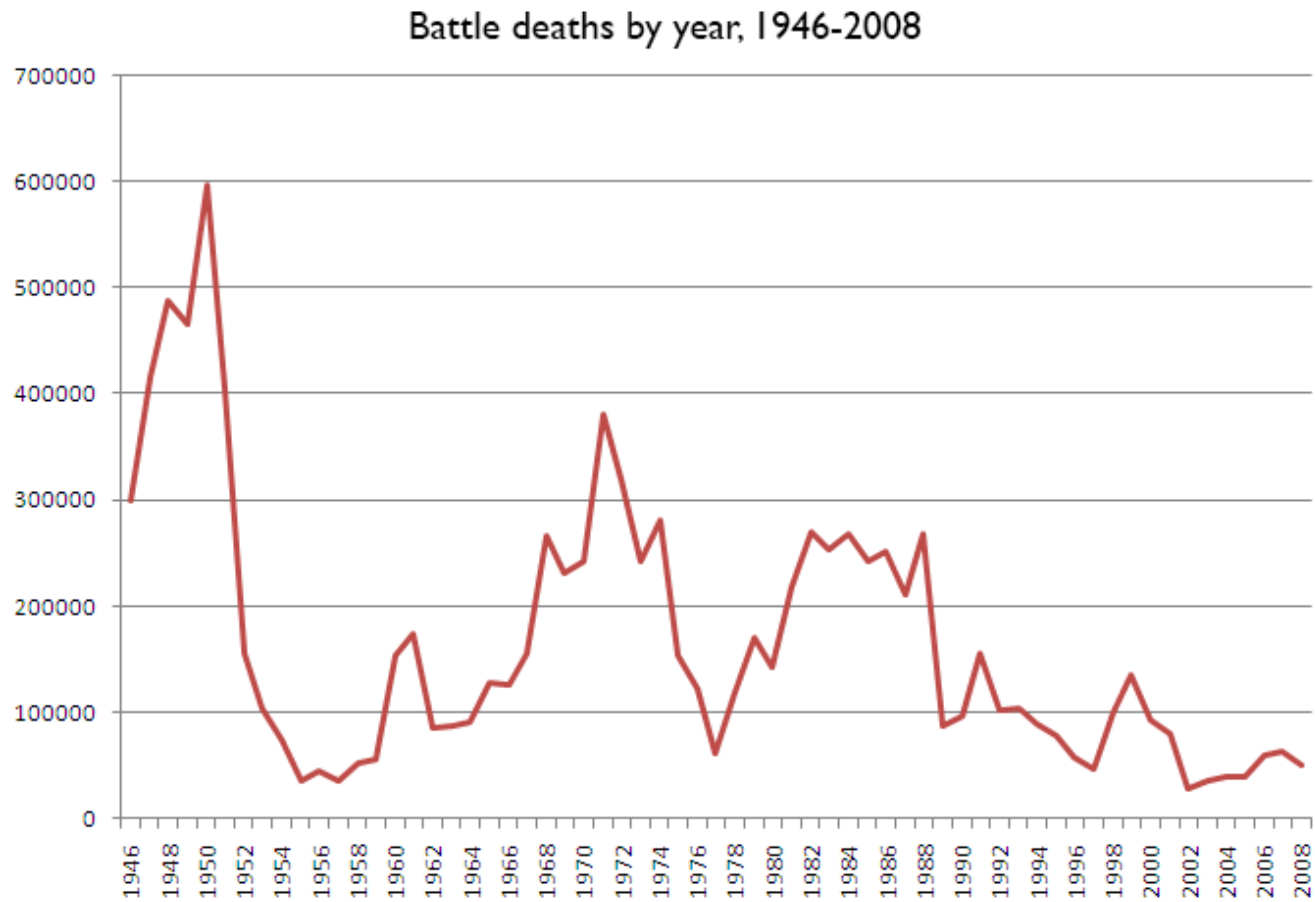
Numerous scholars have argued that war is in decline.

Numerous scholars have also challenged this decline-of-war thesis.

I can't survey this whole discussion here.

But I can give you the basic set-up for an interesting emerging debate that happens to also be good background for part of the PhD dissertation of Céline Cunen.

Here is a (locally produced) time series on battle deaths in war since the end of World War 2 – the numbers come from the PRIO battle deaths dataset.



It's a rather jagged curve but it does seem, broadly, to decline over time.

The decline would look rather sharp if we added in World War 1 and World War 2.

The above curve is just one piece of evidence meant to provide a flavour of the decline-of-war thesis – for more you could consult books by [Steven Pinker](#) and [Joshua Goldstein](#) or various instalments of the [Human Security Report](#).

We move now to the new critique of the decline-of-war thesis.

This recent debate flows out of an old insight of the great Lewis Fry Richardson - ***the distribution of war sizes has a fat tail.***

This means that, although bigger wars are less common than smaller ones, the gradient along which war frequencies decline as war sizes increase is not nearly as steep as it would be if war sizes followed, e.g., a normal distribution.

More important for today is an apparent implication of this fat-tail discovery for the frequency of what I will call truly huge wars, i.e., wars at least as big as World War 1; ***truly huge wars are really rare but not really really really rare.***

Suppose that every time a war breaks out there's a random draw determining the war's size and, further, that each draw is independent of the others.

Also suppose that the probability of truly huge war is 0.006 on each draw.

Then there is about a $1/3$ chance that at least 1 war out of 200 will turn out to be truly huge.

And, of course, there is a $2/3$ chance that we will suffer through 200 wars without a truly huge one erupting.

So we can view this probability of 0.006 as simultaneously not too big and not too small, kind of in between, just like the porridge of Goldilocks.



0.006 is not so small that the risk of a truly huge war can be safely ignored – a 1/3 chance of a truly huge one in 200 draws on average is, in fact, pretty alarming.

But 0.006 is also not so big that the seriousness of the risk will necessarily be immediately visible – indeed, there is a 2/3 probability that we can run through 200 wars while the risk of a truly huge one just lurks quietly below the surface.



Note, moreover, that the numbers in the above example are not completely arbitrary.

In the [war dataset](#) of Kristian Gleditsch there are 362 wars from the beginning of the 19th century through World War 2 and $2/362$ is roughly equal to 0.006.

Also, this dataset contains 212 wars after World War 2, i.e., roughly 200.

So using the pre-World War 2 prevalence of truly huge wars to estimate post-World War 2 prevalence of truly huge wars suggests that we may, in fact, be in **bad Goldilocks range**, as described in the previous slide.

This bad Goldilocks property captures the essence of the new critique of the decline-of-war thesis made by Cirrillo and Taleb (2016) and Clauset (2018).

Informally, the essential claim is that, because of the fat-tail property of war sizes, it is too soon to say that the risk of truly huge wars has disappeared, or even decreased a lot, since the end of World War 2.

More formally, Clauset tested the hypothesis that war sizes after World War 2 are generated by the same mechanism that generated war sizes before World War 2 and fails to reject what I will call this “no-change hypothesis”.

Let's continue, now, to focus on tail probabilities for the distribution of war sizes, but let's consider lots of tails, i.e, all the wars above size X for all X .

Over the next bunch of slides I'll make four changes to Clauset's assumptions that, together, are sufficient to largely overturn Clauset's results, that is, I'll reject most of my no-change hypotheses.

I'll then assess the plausibility of these changes and the importance of each one in moving the results from non-rejection toward rejection.

Along the way I'll show you a number of pictures that are constructed as follows.

1. Pick a cut-off year. These will always be either Clauset's choice of 1945 or 1950, a date inspired by Cunen, Hjort and Nygard (2018).
2. For every war size X we compute the fraction of wars before the cut-off year that are of size X or greater – call this fraction Y .
3. We also compute the fraction of wars after the cut-off year that are of size X or greater – call this fraction Z – and the number of wars after the cut-off year – call this N .
4. Then we test the hypothesis that there are Z or fewer wars of size X or greater in N independent draws given a probability of Y on each draw that war size will exceed X .

For our first and, we think, most illuminating exercise we make the following changes compared to relative to Clauset's set-up.

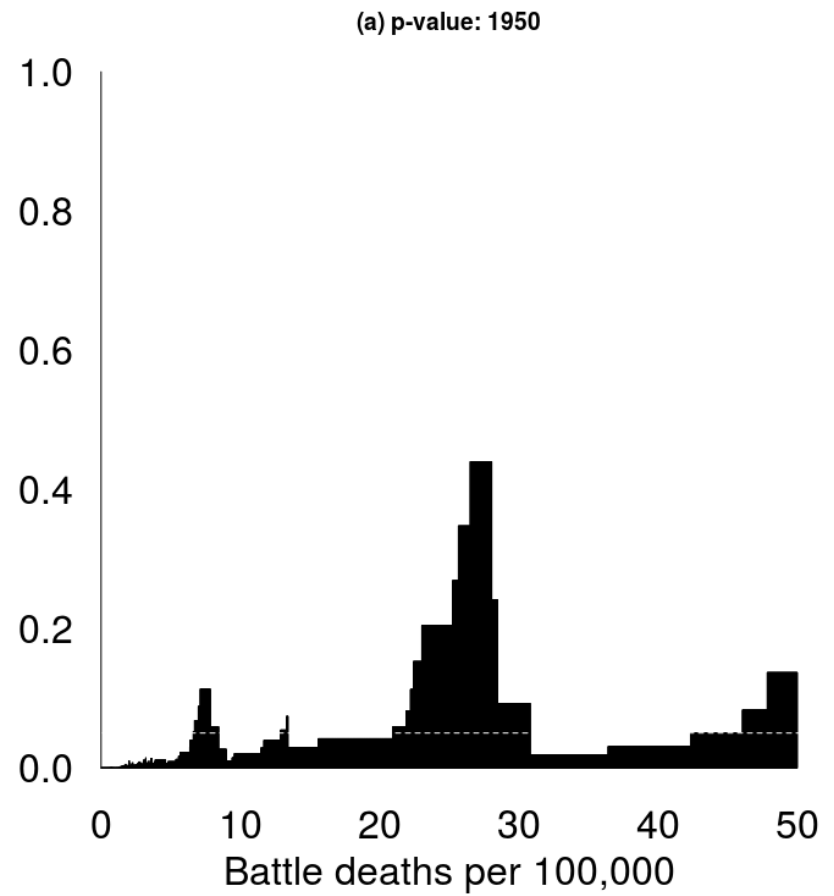
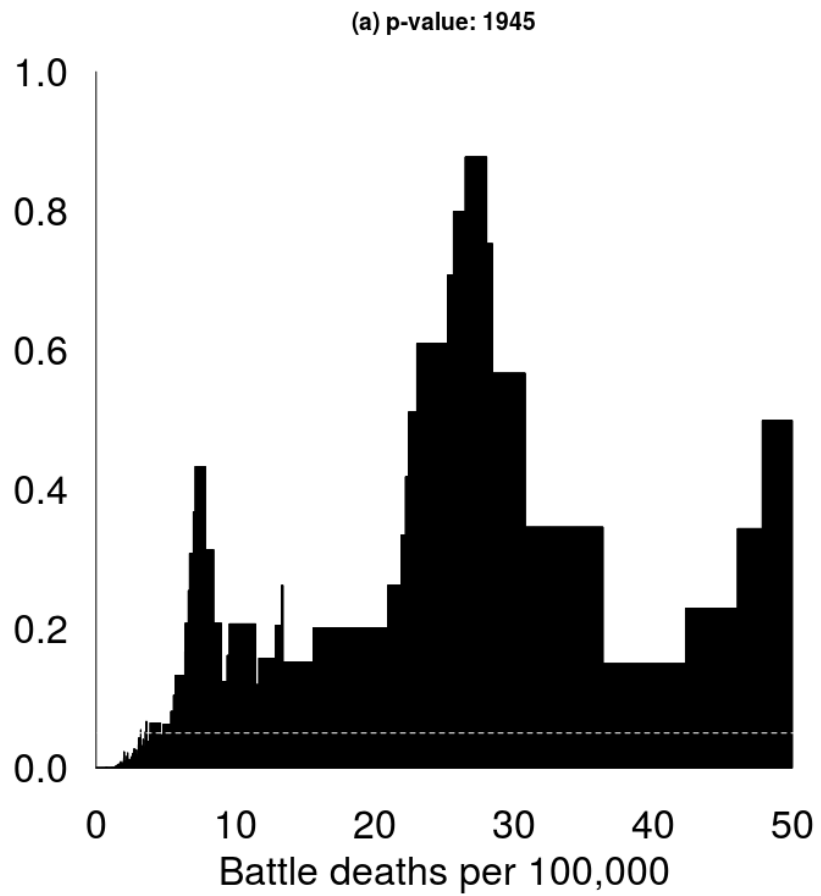
1. We divide war sizes by contemporaneous world population numbers rather than using raw war sizes as Clauset does.

This choice follows much of the decline-of-war literature which stresses the probability of an average person getting killed in a war rather than absolute violence levels.

2. We consider 1950 as a possible break point rather than only 1945 as Clauset does.

3. We include intra-state wars in addition to just inter-state wars as Clauset does.

4. We use Kristian Gleditsch's data rather than the [Correlates of War](#) dataset that Clauset uses.



The pictures only show the numbers for tails starting at 50 battle deaths per 100,000 and below.

There are just four wars with more than 50 battle deaths per 100,000 of world population.

World 2 (size 781) and World War 1 (size 499) are so massive that the picture would become almost unreadable if we included them.

The American Civil War (52) and the Chinese Civil War (51) are just barely above 50 so it seemed most elegant to start at 50.

Let's walk through the left graph from right to left i.e, moving down toward smaller and smaller values on the X axis.

Moving to the left, the graph steps down when we hit a war that started before 1945.

Intuitively, switching a pre-1945 war into the pre-1945 right-tail by lowering the threshold for tail membership raises our estimated probability that a randomly war starting after 1945 will qualify for the post-1945 right-hand tail –

Thus, the p value for the no-change hypothesis for this newer, wider tail (lower threshold) must be below the p value for the older, thinner (higher threshold) tail.

Moving to the left, the graph steps up when we hit a war that started after 1945.

Intuitively, switching a post-1945 war into the post-1945 right-tail by lowering the threshold for tail membership increases the number of realizations of tail events without changing the probability of such events under the no-change hypothesis.

Thus, the p value must increase.

It's worth taking a few seconds to identify the wars that move the curves.

The third Sino Japanese War, the Russian Civil and the Pakistan-Bengal War, all starting before 1945, pull p down below 0.2.

But then the Korean War, Vietnam Phase 2, the Iran-Iraq War and the Second South Sudan war pull it way up above 0.8 – in fact the tails starting from around the 26-28 range on up is fatter post-World War 2 compared to pre-World War 2

I'd venture that there has been something of a shift toward smaller wars since 1945 although we don't come remotely close to rejecting any no-change hypothesis at anything like a standard level.

However, the picture looks very different if we look at the picture on the right which considers 1950 as the break point.

This change flips the Chinese Civil War and the Korean War, of the biggest ever, into the post-break-point category.

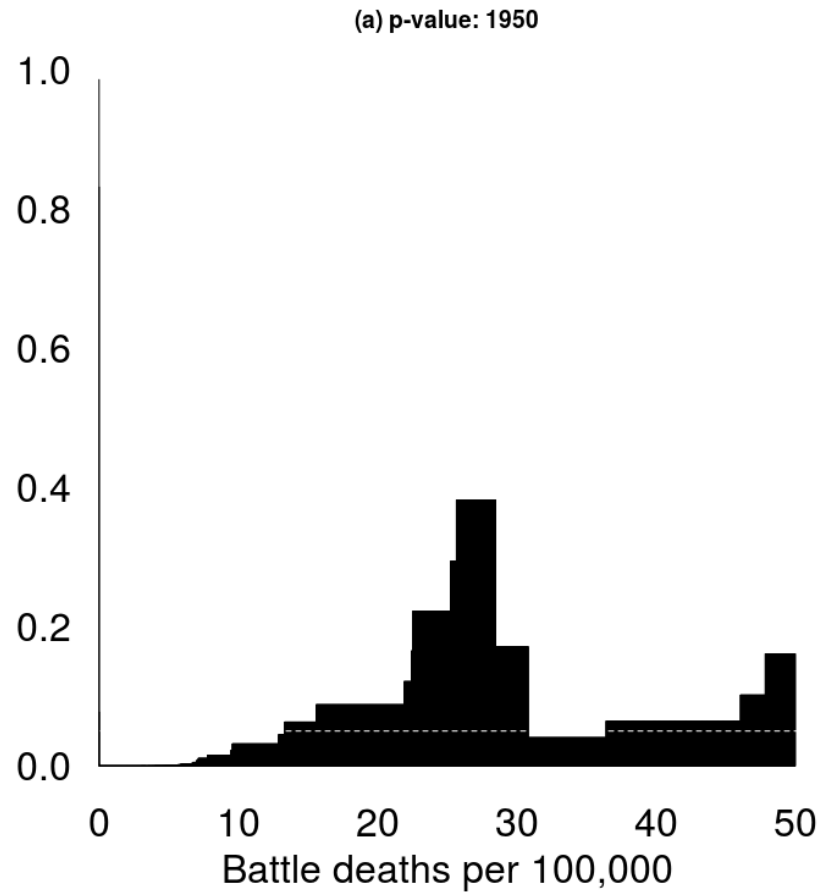
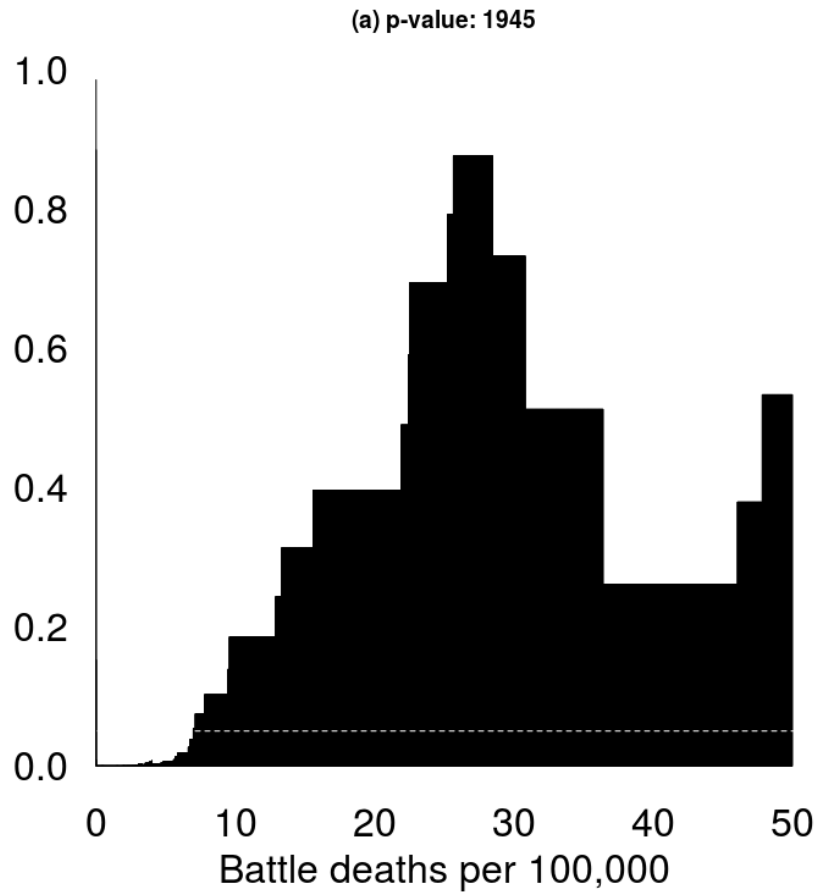
Now we get strong rejections of the no-change hypothesis as soon as we move much below 50.

Then run of post-1950 wars with sizes between 26 and 28, now 3 rather than 4, does pull the p values back up but even at the peak we can still say that the no-change hypothesis is probably false.

Now let's evaluate the importance of the 4 changes that we made simultaneously to arrive at the picture on slide 14.

The next slide shows that switching from the Correlates of War (COW) data to the Gleditsch data doesn't matter.

I prefer the Gleditsch data because it includes a number of wars that COW passes over due to excessive Eurocentrism – but this turns out to not matter for these exercises.



Normalizing for population does matter hugely for our approach – without this we get nowhere.

In absolute terms 7 out of the biggest 10 wars and 14 out of the 20 biggest wars start before 1950.

In per capita terms 8 out of the biggest 10 wars and 17 out of the 20 biggest wars start before 1950.

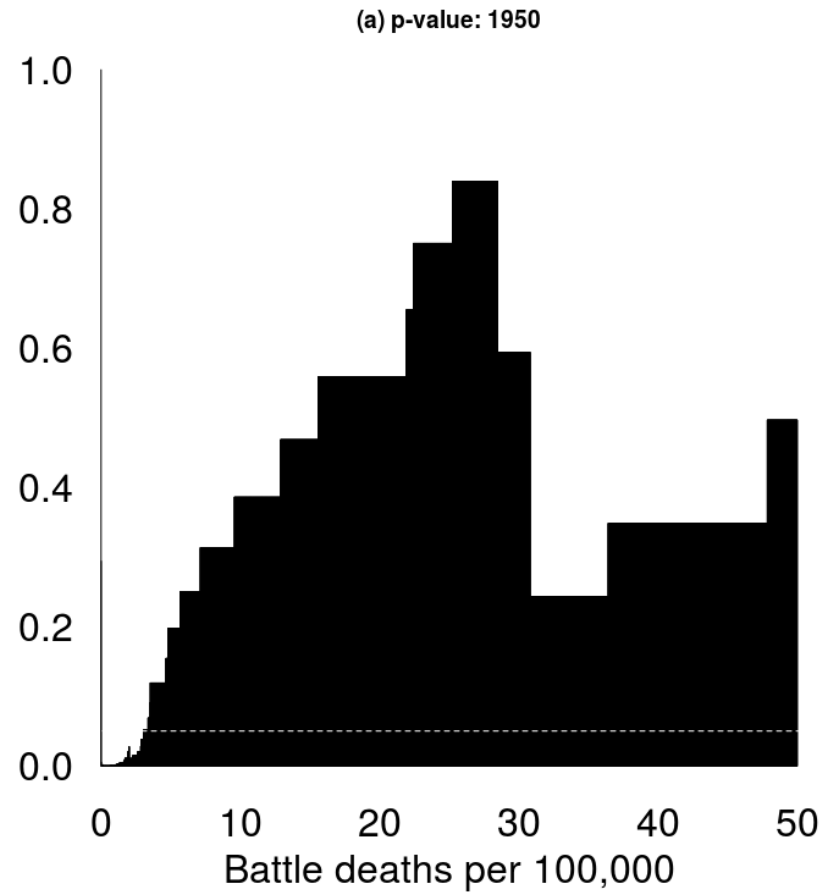
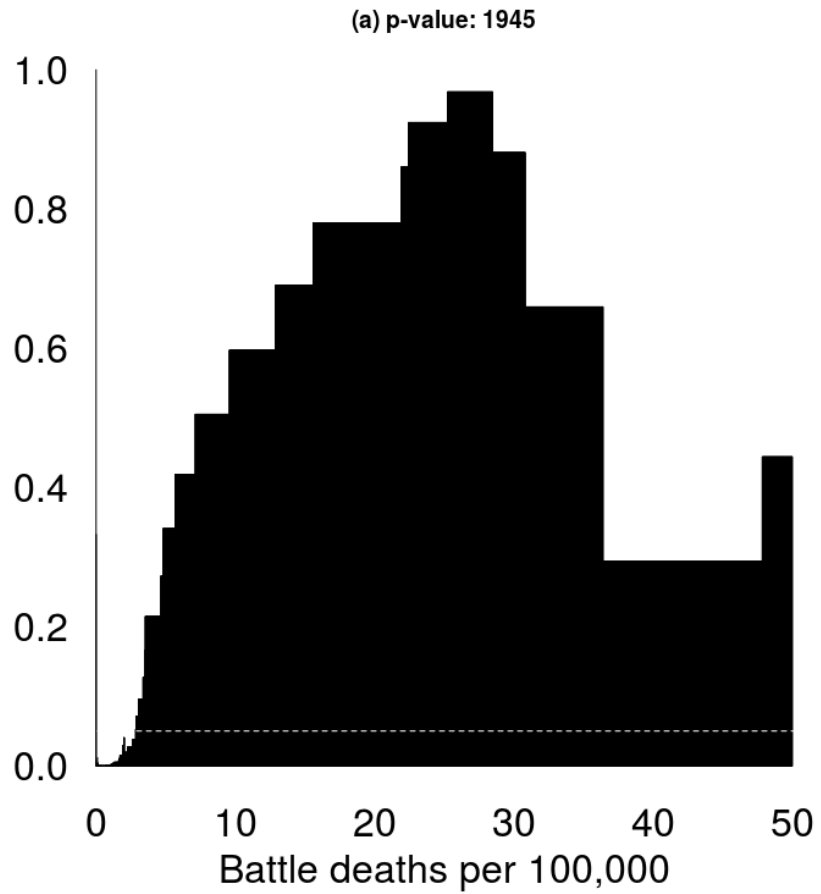
That said, other approaches, such as that of Cunen, Hjort and Nygard (2018), find breaks, including at 1950 without normalizing for population.

Finally, recall that all the pictures so far have included intra-state wars in addition to the inter-state wars that Clauset includes.

We think that it's best to include intra-state wars since these can be very blood indeed – for example, the US Civil War and the Chinese Civil War are, respectively, the 3rd and 4th biggest wars in the dataset in per capita terms.

So we think that including intra-state wars is the correct decision.

The next picture shows that this decision matters – our results are much weaker without intra-state wars.



So far we have only looked at these two potential break points, 1945 and 1950 and you can see that the change point matters.

The next picture considers all possible break points and graphs the average p below 50.

It turns out that the best one by this criterion is 1950.

Average p-value for different break points



Conclusions

First, a caveat – we've looked only at the time series of war sizes (population adjusted) but we think that a full analysis should look at a broad array of evidence of the sort you see in the decline of war literature

Nevertheless, we think the decline-of-war thesis is still alive and kicking.



