The Reliability of Cluster Surveys of Conflict Mortality: Violent Deaths and Non-Violent Deaths

Presentation given at the conference:

"International Conference on Recording and Estimation of Casualties"

Carnegie Mellon University and University of , Pittsburgh - October 23 and 24, 2009

Michael Spagat – Royal Holloway, University of London

People contributing to the work include

Josip Dasovic
Simon Fraser University

Violetta Dalla,
&
Ana Gabriela Guerrero Serdán,
Royal Holloway, University of London

Mathew Harrison – Unaffiliated

Dueling Surveys

Recent surveys estimating violent conflict deaths have arrived at fundamentally incompatible conclusions.

For example, the Burnham et al. survey (*Lancet*) and the Iraq Family Health Survey (*New England Journal of Medicine*) differ in there violent death estimates for almost exactly the same period by a factor of 6.6 (or still by a factor of 4 if the comparison is done incorrectly as it usually is). Confidence intervals for these studies are nowhere near to overlapping.

In this talk I give some evidence that small surveys of violent conflict deaths are less reliable than is commonly thought. This may explain some discrepancies.

I also point to a source of systematic bias in surveys of violent conflict deaths that is potentially quite important.

We use data from the **Iraq Living Conditions Survey** 2004 (**ILCS**).

Except in Kurdistan all interviews were done between **March 22 and May 25, 2004**.

Interviews were done at **10 households** (with minor variation due to incompleteness) within each of **2,193 clusters** comprised of **70 to 200** households.

Thus, the ILCS was a **very large survey** in terms of both the number of clusters (psu's) and the number of households where interviews were conducted.

Moreover, **each cluster measurement** in the ILCS was **of just a small neighborhood**.

The ILCS recorded all household deaths: causes are classified as either: pregnancy/child birth, disease, traffic accident, war-related or "other (specify)".

"War-related deaths" and "violent deaths" should be essentially equivalent but I will use the ILCS term "war-related deaths" for these and call everything else "non-violent".

We have a simple two-column dataset consisting of **a list of (weighted) war-related deaths in every ILCS cluster and a list of (weighted) non-violent deaths in every ILCS cluster.**

Here are some interesting facts:

1. **Violence is punctuated**; Only **105 out of the 2193 (4.8%)** had positive war-related deaths, i.e., although Iraq suffered much violence during the ILCS coverage period the overwhelming majority of small neighborhoods of 70 to 200 households do not seem to have experienced any war-related deaths.

2. **Non-violent deaths are diffuse**; 902 out of the 2193 clusters had positive non-violent deaths.

3. **Violence is concentrated**; 80% (30%) of the clusters with violence had more than 10 (20) times the average number of war-related deaths.

4. **Non-violent deaths are not concentrated**; only 2.5% (0%) of the clusters had more than 10 (20) times the average number of non-violent deaths.

We study the small-sample properties of the most basic estimators of violent and non-violent conflict mortality by taking a large number of random draws of various sizes from the list of ILCS clusters following these procedures:

1. Fix a sample size of 10 clusters.

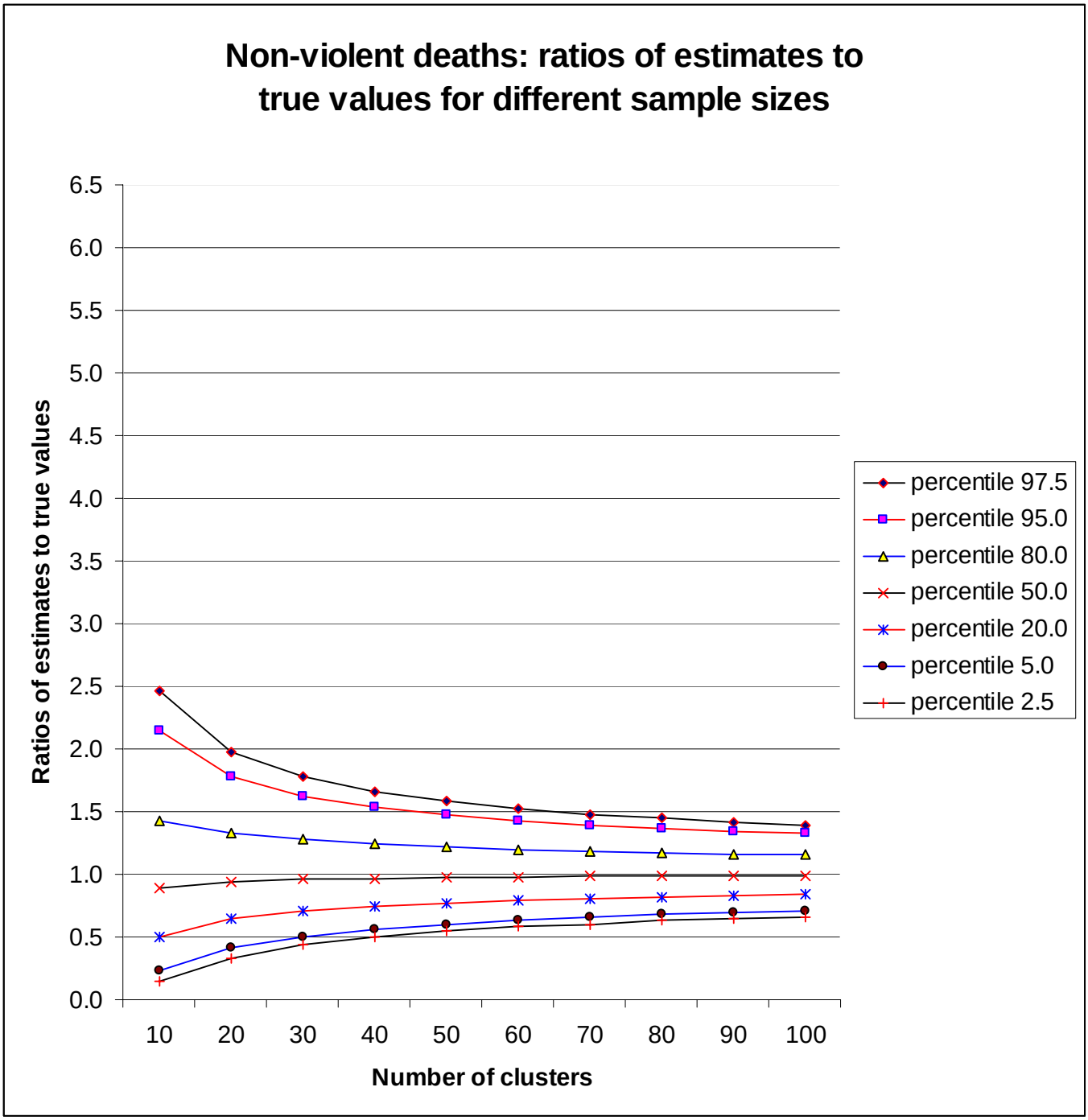
2. Draw 10,000 different samples of 10 clusters (with replacement) from the ILCS list of 2,193 clusters.

3. For each of these 10,000 samples calculate the average number of war-related deaths in this sample of 10 clusters.

4. Repeat the above steps for samples of 20, 30,…, 100, 200, 300,…2,000 clusters.

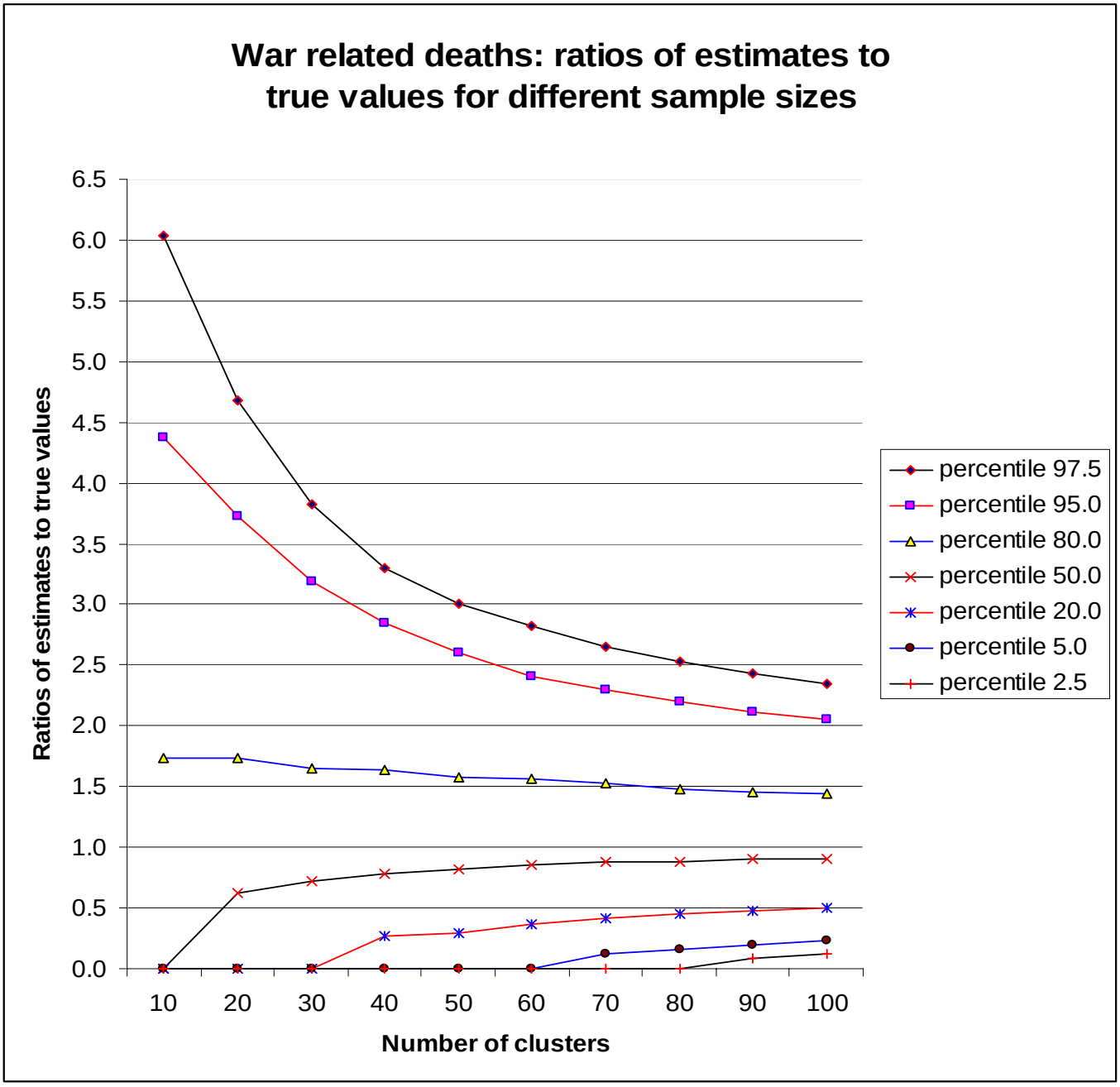
5. Repeat all of the above steps for non-violent deaths.

The next five slides present the results of these Monte Carlo simulations for clusters between the sizes of 10 and 100.



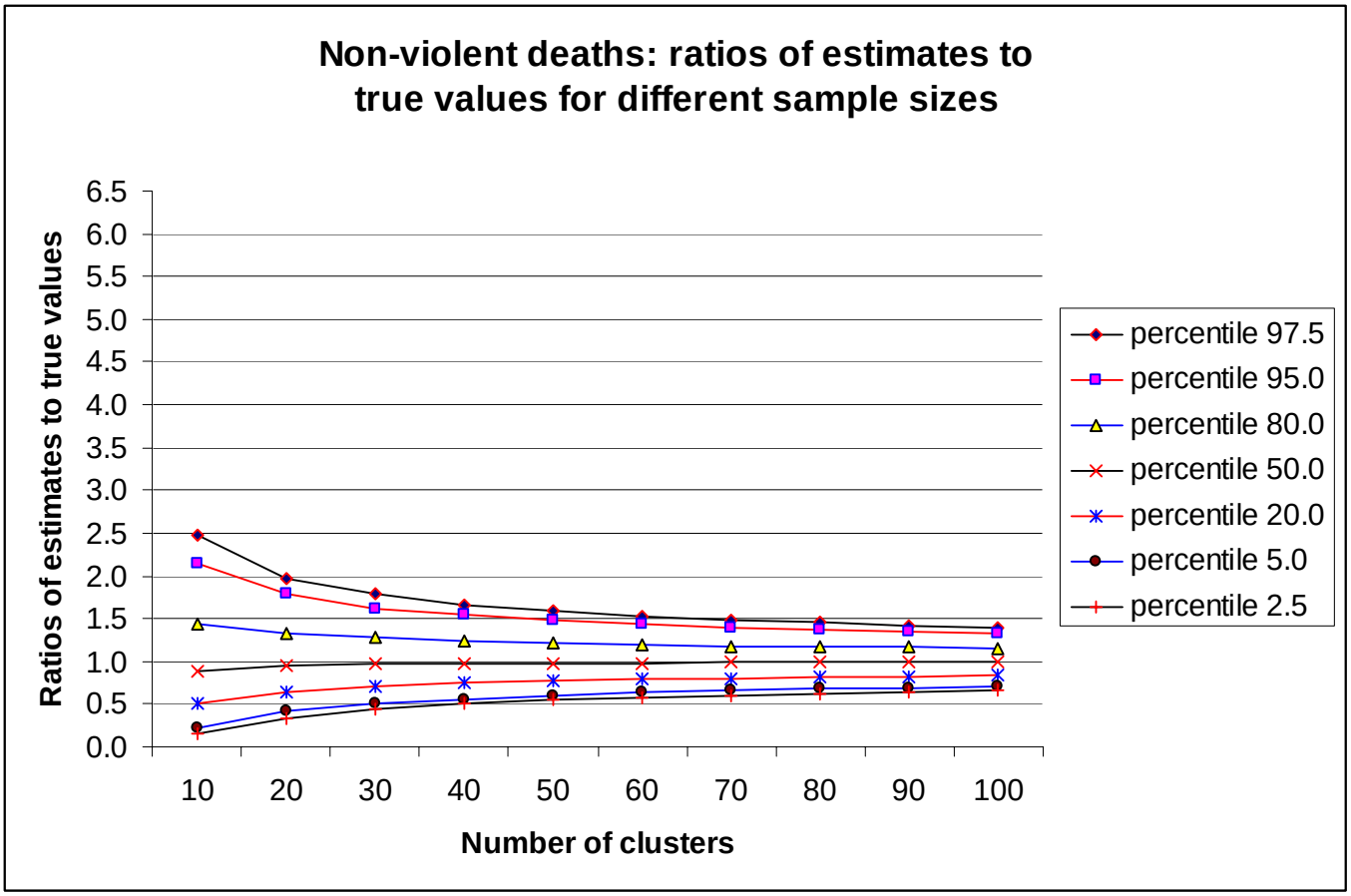
With 30 (50) clusters 60% of the estimates are within 30% (20%) of the true value.

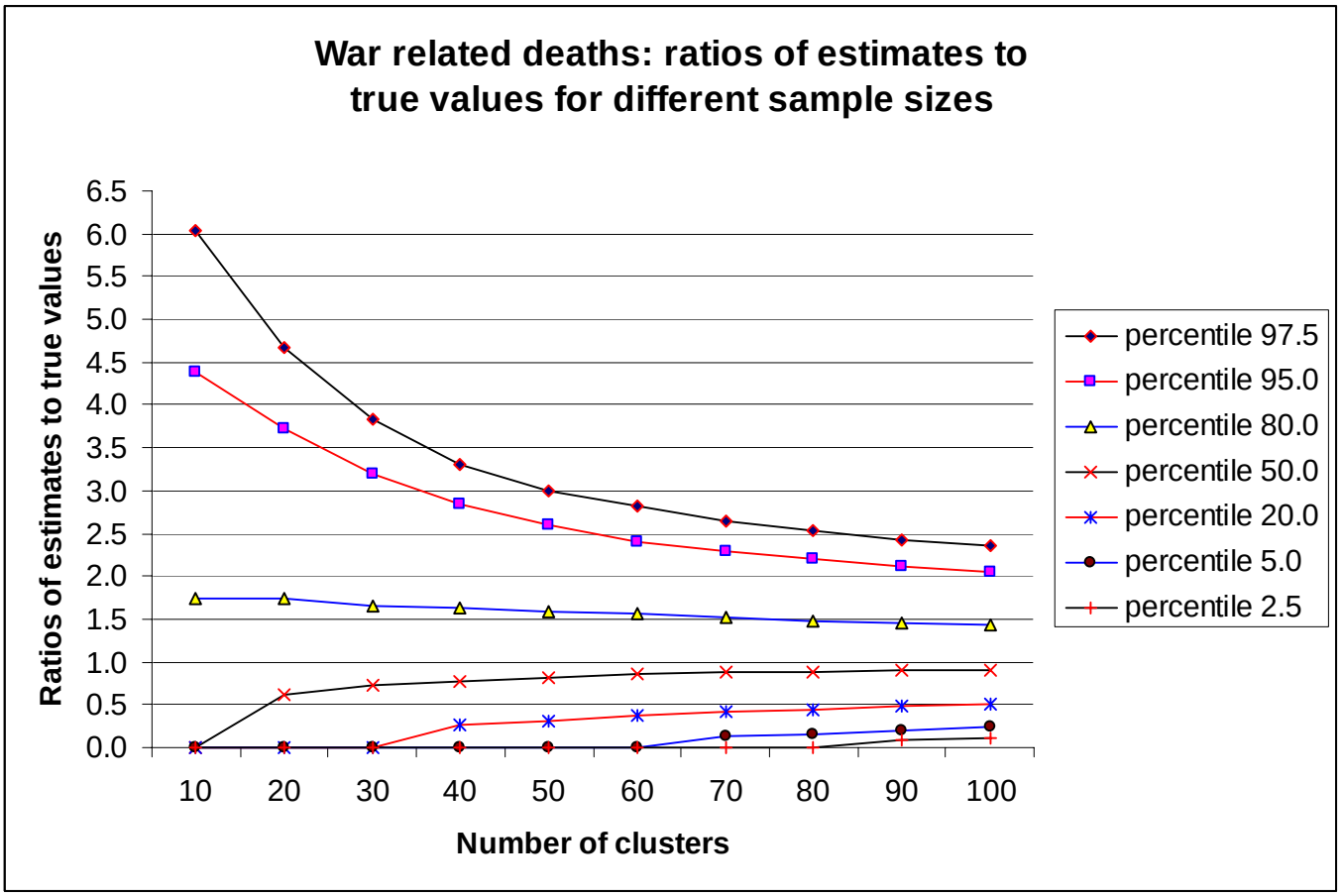
With 50 (100) clusters there is less than a 5% chance (virtually no chance) of deviating from the true value by more than 50%.



With 30 (50) clusters more than 5% of the estimates of war-related deaths are more than triple (2.5 times) the true value and more than 20% (5%) do not detect any deaths at all.

With 50 (100) clusters estimates are within 50% of the true value 46% (60%) of the time.


Non-violent deaths: ratios of estimates to true values for different sample sizes
Ratios of estimates to true values
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
10
20
30
40
50
60
70
80
90
100
Number of clusters
percentile 97.5
percentile 95.0
percentile 80.0
percentile 50.0
percentile 20.0
percentile 5.0
percentile 2.5


War related deaths: ratios of estimates to true values for different sample sizes
Ratios of estimates to true values
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
10
20
30
40
50
60
70
80
90
100
Number of clusters
percentile 97.5
percentile 95.0
percentile 80.0
percentile 50.0
percentile 20.0
percentile 5.0
percentile 2.5

Summery

Non-violent deaths are estimated much more precisely than war-related deaths.

Small samples, such as the widely used 30 or 50, perform quite badly for war-related deaths; they can easily fail to detect any deaths or, on the other hand, overestimate by a factor of 3.

Notice that the ***median*** **estimates for war-related deaths are well below the true values in small samples,** i.e, underestimation is more likely than overestimation; the median estimate for a sample of 30 (50) is 30% (20%) below the true value.

These simulation procedures are unbiased by design. Therefore, **overestimation tends to be larger when it occurs than is underestimation when it occurs.**

In other words, in small samples you are more likely to underestimate than overestimate but when you overestimate you are likely be farther from the true value than you are when you underestimate.

Bias

Standard techniques for estimating the prevalence of rare events tend to overestimate (Hemenway, 1997).

Favorite survey-based “factoid” of the National Rifle Association - 2.5 million annual uses of guns in self defense in America.

This implies, for example, that:

> Guns are used in more than 100% of burglaries for which victims are home and have a gun, even though in most of these cases victims were sleeping. Other research puts this percent closer to 2%.
>
> Victims of robberies and rapes use guns against assailants more frequently than their assailants use guns against victims.

Explanation: *Classification Error* – particularly important for rare events like self-defense gun uses.

Suppose

1. 1% of households have a violent death.
2. 1% of the time when households do not have a violent that a (spurious) violent death is recorded anyway.
3. 10% of the time when households do have a violent death none are recorded.

Then a survey would expect to record 1.89% of households as having a violent death, i.e.,

$$99 \times 0.01 + 1 \times 0.9 = 1.89$$

So even though, by assumption, the survey is 10 times more likely to miss violent deaths that have really occurred than to record violent deaths that have not occurred, *there is still upward bias by nearly a factor of 2!*

If we change the prevalence of violent deaths to 0.1% than overestimation is by nearly a factor of 11, i.e., 1.089% versus 0.1%.

This is a systematic bias. Getting larger samples does not help.

It comes from the fact that the phenomenon measured is rare.