



© 1997–2009, Millennium Mathematics Project, University of Cambridge.

Permission is granted to print and copy this page on paper for non-commercial use. For other uses, including electronic redistribution, please contact us.

23/10/2006

News

Body count



Although we are bombarded with statistics every day, we rarely hear about the methods used to compile them. Recently, though, a rare discussion of statistical methodology did spring up when US President George W. Bush deemed as "pretty well discredited" a study into the number of deaths in Iraq since the US led invasion in March 2003. The Iraqi government joined Bush in his criticism, calling the figures "inflated" and "far from the truth".

The study, which appeared in the medical journal *The Lancet*, was conducted by Gilbert Burnham and colleagues at the Johns Hopkins University in Baltimore, USA, and Al Mustansiriya University in Baghdad. It estimates that 655,000 deaths occurred as a result of the invasion, increasing the mortality rate of 5.5 per thousand inhabitants before the invasion to 13.2 per thousand inhabitants after the invasion. It also suggests that around 601,000 of the 655,000 "excess deaths" were violent, mainly due to gunfire, and that around 31% of the violent deaths could be attributed to coalition forces. The number of 655,000 excess deaths is far higher than previous estimates: the website [Iraq Body Count](#), for example, puts the number of civilian deaths between 44,000 and 49,000, while other sources put it no higher than 126,000. This study is a sequel to a similar one conducted by the same group in 2004, which also gave an unexpectedly high estimate and also sparked controversy.



The study's authors used a technique called "cluster sampling", which is often employed to estimate death tolls in war zones and the aftermath of disasters. They selected 47 geographical areas (clusters) within Iraq

Body count

and interviewed around 40 randomly chosen households in each cluster, giving a total of 1849 households comprising 12,801 individuals. Each household was asked about the number of deaths before and after the invasion and to provide death certificates where possible.

In total, the study recorded 629 deaths, 13% of which occurred in the 14 months before the invasion and 87% in the 40 months after the invasion. The figure of 655,000 excess deaths since the invasion arises from "extrapolating" these results to apply to the whole population of around 27 million.

The huge discrepancy between the new figures and previous estimates caused a barrage of criticism in the media and on the web. After examining the study more closely, many commentators felt that the figure of 655,000 was a cheat: the authors had in fact given a whole range, from 392,979 to 942,636, in which they estimate the true number of excess deaths to lie. As in the 2004 study, critics felt that the large size of this range indicates the authors' lack of confidence in their result, and that the number 655,000 was fed to the media to conceal this lack of certainty. So how should this range be interpreted and where does the number 655,000 come from?

The range from 392,979 to 942,636 is what statisticians call a *confidence interval*. Statistics operates on the notion that a random sample of a reasonable size is representative of a whole population. But of course, you can't expect that the results from a sample *exactly* reflect the whole population. There is even an outside possibility that, purely by chance, your sample is completely unrepresentative. Theoretically it is possible that by chance the researchers only picked households that had experienced more deaths and violence than all others. A confidence interval is one way of measuring this inherent uncertainty.

To understand this, think of a survey like the one done in Iraq as an experiment that can be repeated many times. Each time you take a random sample of the population and use this to estimate some value you are interested in, in this case the number of excess deaths in Iraq. Each repeat of the survey will give a different estimate and it's unlikely that any estimate will precisely equal the true value. How do you know how accurate your estimates are? Here probability theory comes to the rescue: it is possible to calculate which percentage of the estimates lies within a certain distance d of the true value. Conversely, you can start with a given percentage, say $p\%$, and calculate how big a number d you have to choose, so that $p\%$ of the estimates lie within d of the true value.



This theory gives you information about the level of uncertainty involved in your survey. You first set a *confidence level*, say $p\%$ and then calculate the corresponding number d . This tells you that if you repeat your survey many times, then in $p\%$ of all repeats the difference between the true value and your estimate is at most d . In other words, you can be $p\%$ confident that the *error* in your survey is less than d . The numbers that

Body count

lie within distance d of your estimate form an interval of length $2d$ this is the confidence interval.

For the Iraq study, a 95% confidence level gave the range from 392,979 to 942,636 excess deaths: the authors are 95% confident that the true number of excess deaths lies in this range. The theory also gives an indication as to which number in this range is *most likely* to equal the true value. In our case, this number is 654,965, hence the figure 655,000 that made the headlines.

With a confidence level of 95%, there still is, of course, a 5% "lack of confidence". It is possible to increase the confidence level to, say, 99%, but only at the cost of increasing the confidence interval also a higher level of certainty leads to a lower level of precision. Generally, the degree of certainty given by a 95% confidence level is deemed acceptable; most studies use this figure.

Thus, the large size of the confidence interval is not a measure of the authors' lack of confidence in their results. It is a measure of the level of precision they were able to attain while remaining 95% confident. The only way to make the interval smaller while retaining a 95% confidence level would be to increase the sample size. But, as the authors say, this would have been difficult in the Iraq study, as they had to "balance the need for robust data with the level of risk acceptable to field teams." They feel that their sample size of around 12,000 was sufficient for the degree of accuracy required. Provided you trust that the sample is unbiased, even a large confidence interval gives interesting information in this case, since its lower end, 392,979, is already way above all previous estimates of excess deaths.

The use of confidence intervals is nothing controversial. But what about cluster sampling, a technique that struck some critics as suspect? Selecting households from a series of clusters, rather than from the country as a whole, does not give a truly random sample. Deaths in one cluster could be correlated, since a single bomb may have been responsible for the majority of deaths within that cluster. Thus, deaths in any one cluster cannot be considered as independent events. Cluster sampling is used widely in situations where it is unfeasible to access all parts of a region, and statisticians have devised ways of adjusting for the correlations within particular clusters. The authors have used such techniques in their study.

Despite the criticism from some quarters, the study's methodology has been condoned by a host of distinguished scientists, including the four experts who cleared the study for publication in *The Lancet*. One expert, epidemiologist Ronald Waldman of Columbia University, told the *Washington Post* that its methods were "tried and true" and that "this is the best estimate of mortality we have."



Body count

Critics, on the other hand, say that while the tools used to evaluate the data may have been sound, the data itself the sample may have been unrepresentative. Indeed, in practice it is pretty much impossible to guarantee a completely unbiased sample. By the author's own admission, bias might have crept into the Iraq study in a variety of ways. The interviewers had some choice in picking the households to be interviewed and they might have been unconsciously drawn to those that appeared worst-affected by violence. According to their own political standpoint, the families interviewed might have misrepresented the number of deaths due to coalition forces. Although around 80% of all deaths were documented by certificates, some deaths might have been invented. Migration out of and within Iraq could also have an effect on the overall estimate.

While some of these aspects might have lead to an inflated estimate of the death rate, the authors point out that other sources of bias almost certainly lead to an under-estimate. Some deaths, particularly those of infants or combatants, might not have been reported at all. Whole families might have been wiped out and could therefore not have been interviewed, leading to a survivor bias that underestimates the death rate. In short, it is quite difficult to tally up the two kinds of bias.

So where does this leave us? Proponents of the study argue that, although the results might not be terribly accurate, the figures at least get us into "the right ballpark", as one of the authors put it. They also say that the wide gap between the new figures and previous estimates should not come as a surprise. Most previous estimates are based on *passive surveillance methods*, in other words, they rely on reports from the media, hospitals and morgues, rather than on active research. In times of conflict these reports are notoriously unreliable: "Aside from Bosnia," the authors say, "we can find no conflict situation where passive surveillance recorded more than 20% of the deaths measured by population-based methods."

Opponents of the study, including the Iraq Body Count (IBC), argue that the huge number of deaths indicated by the study could simply not have gone unnoticed. It would imply that a huge number of bodies had gone "missing", that many casualties did not receive hospital treatment and that a large number of death certificates were issued without being officially recorded. This, they say, could only happen if hospitals and ministries are massively incompetent, or even fraudulent. "In the light of such extreme and improbable implications," says the IBS, "a rational alternative conclusion to be considered is that the authors have drawn conclusions from unrepresentative data." In other words, the IBC suggest that the sample was biased.

Whether or not this is true is hard to determine for laypeople. For many, the decision on whether or not they accept the study probably comes down to who they want to believe. All we know for certain is that there is a gaping gap between the figures and that statistics and politics make for an explosive mix.

Further reading

- Read the study entitled [Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey](#);
- Read an [earlier study](#) into deaths in Iraq, which was conducted by the same authors in 2004 and sparked similar controversy.
- Read Paul Reynold's [analysis](#) on the BBC website;
- Read an article on the use of statistics in the study on [Stats.org](#).
- Read Iraq Body Count's [response](#) to the study.

Marianne Freiberger

Body count



Plus is part of the family of activities in the Millennium Mathematics Project, which also includes the NRICH and MOTIVATE sites.