Comments on the Confidence Intervals of Roberts et al. (2004)

David Kane*

July 17, 2007

Abstract

Roberts et al. (2004) claim that the risk of death increased by 2.5-fold (95% CI 1.6-4.2) in Iraq after the US-led invasion.[1] I provide evidence that, given the other data presented in their paper, this confidence interval must be wrong. Comments and corrections are welcome.

*dkane@iq.harvard.edu. Institute Fellow, IQSS, Harvard University, Cambridge, MA 02138. Thanks to Arjun Ravi Narayan, Aaron Schwartz, Daniel Suo and Luyi Zhao for excellent research assistance.

[1] This work is part of a larger project critiquing Roberts et al. (2004) and Burnham et al. (2006). For simplicity I refer to Roberts et al. (2004) as L1 in this paper.

1 Introduction

The most important result from L1 is the first sentence of the Findings section.[2]

> "The risk of death was estimated to be 2.5-fold (95% CI 1.6 – 4.2) higher after the invasion when compared with the pre-invasion periods."

Unfortunately, if the other results presented in L1 are correct, this confidence interval is wrong. It is too narrow, especially at the lower end. *The Lancet* authors cannot reject the null hypothesis that mortality in Iraq is unchanged.[3]

Define terms used in L1 as follows: "rate of death," "risk of death" and "crude mortality rate" (CMR) are the number of deaths per thousand people per year; "relative risk" (RR) is the post-invasion crude mortality rate (CMR_{post}) divided by the pre-invasion crude mortality rate (CMR_{pre}).

[2] Entire Findings section:

> The risk of death was estimated to be 2.5-fold (95% CI 1.6 – 4.2) higher after the invasion when compared with the pre-invasion period. Two-thirds of all violent deaths were reported in one cluster in the city of Falluja. If we exclude the Falluja data, the risk of death is 1.5-fold (1.1 – 2.3) higher after the invasion. We estimate that 98,000 more deaths than expected (8,000 – 194,000) happened after the invasion outside of Falluja and far more if the outlier Falluja cluster is included. The major causes of death before the invasion were myocardial infarction, cerebrovascular accidents, and other chronic disorders whereas after the invasion violence was the primary cause of death. Violent deaths were widespread, reported in 15 of 33 clusters, and were mainly attributed to coalition forces. Most individuals reportedly killed by coalition forces were women and children. The risk of death from violence in the period after the invasion was 58 times higher (95% CI 8 – 419) than in the period before the war.

Note that the most widely quoted result from the study was the mean excess death estimate of 98,000 and its associated confidence interval of 8,000 to 194,000. The authors did not provide a confidence interval for excess mortality which included the data from Falluja and have declined my requests to do so.

[3] Although L1 presents its results within the frequentist paradigm, I prefer a Bayesian approach. The relative risk is an unknown quantity. Using data and models, we can estimate its value and provide confidence intervals for those estimates. Frequentists may supply their own translation.

$$\mathrm{RR} \equiv \frac{\mathrm{CMR}_{\mathrm{post}}}{\mathrm{CMR}_{\mathrm{pre}}}$$

If RR > 1, then the CMR has gone up. By itself, the L1 result of a RR of 2.5 (95% CI 1.6 – 4.2) seems plausible. However, this confidence interval is not consistent with the estimates presented for $\mathrm{CMR}_{\mathrm{pre}}$ and $\mathrm{CMR}_{\mathrm{post}}$.[4]

> $\mathrm{CMR}_{\mathrm{pre}}$ is 5.0 with a 95% confidence interval of 3.7 – 6.3.
>
> $\mathrm{CMR}_{\mathrm{post}}$ is 12.3 with a 95% confidence interval of 1.4 – 23.2.

Why is the confidence interval for $\mathrm{CMR}_{\mathrm{post}}$ more than *8 times* wider than that for $\mathrm{CMR}_{\mathrm{pre}}$ even though the sample sizes are almost exactly the same? Answer: Falluja. Consider See Figure 1.

The central finding of L1 includes the data from Falluja, so let us leave aside whether or not this cluster should be discarded as an outlier. When including Falluja, it is impossible to have a precise estimate of post-invasion mortality. The Falluja cluster by itself creates the large confidence interval for $\mathrm{CMR}_{\mathrm{post}}$. Lead author Les Roberts reports that:

> "There was one place, the city of Falluja that had just been devastated by shelling and bombing, and it was so far out of whack with all the others that it made our confidence intervals very, very wide." (Mares (2006))

In a presentation at MIT, author Gilbert Burnham went further.

[4] Consider these numbers as presented in L1:

> During the period before the invasion, from Jan 1, 2002, to March 18, 2003, the interviewed households had 275 births and 46 deaths. The crude mortality rate was 5.0 per 1,000 people per year (95% CI 3.7 – 6.3; design effect of cluster survey = 0.81). Of the deaths, eight were infant deaths (29 deaths per 1,000 livebirths [95% CI 0 – 64]). After the invasion, from March 19, 2003, to mid-September, 2004, in the interviewed households there were 366 births and 142 deaths – 21 deaths were children younger than 1 year. The crude mortality rate during the period of war and occupation was 12.3 per 1,000 people per year (95% CI 1.4 – 23.2; design effect = 29.3) and the estimated infant mortality was 57 deaths per 1,000 livebirths (95% CI 30 – 85).

Given this presentation (and the software used), it is almost certain that the confidence intervals for estimates of the crude mortality rates are normally distributed.

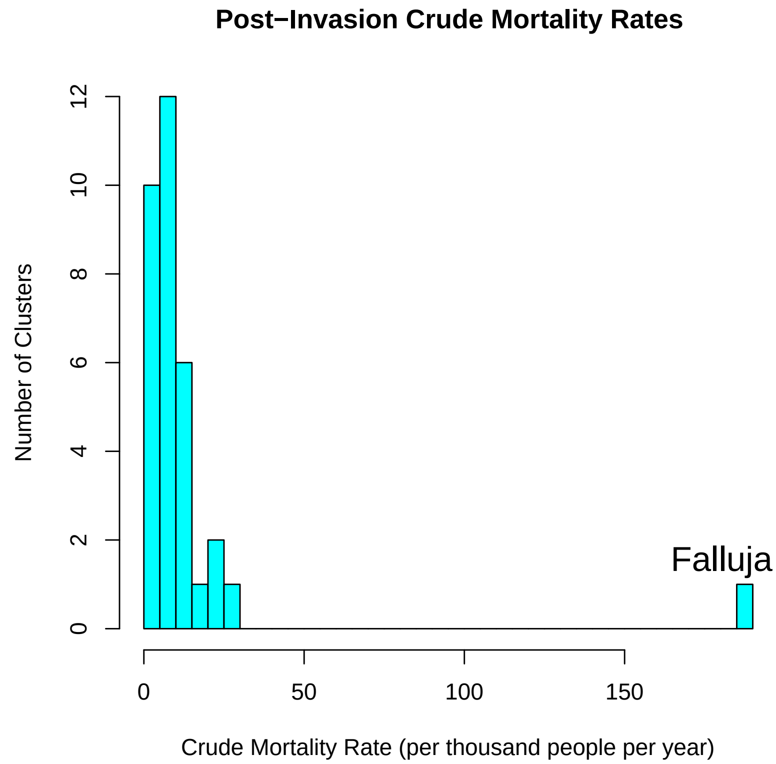

Figure 1: Histogram of CMR_{post} for the 33 clusters sampled in L1. The results for Falluja are unlike those for any other cluster but are not, according to the authors, unreasonable. They report that "in Falluja, the team noted that vast areas of the city had been devastated to an equal or worse degree than the area they had randomly chosen to survey. We suspect that a random sample of 33 Iraqi locations is likely to encounter one or a couple of particularly devastated areas." Falluja is a legitimate data point and should not be removed from the default statistical analysis.

> Now this is what the confidence intervals would look like. There is a 10% probability that it was less than 44,000 and only a 2.5% chance that it was less than 8,000. If we put Falluja into it, the top end of the confidence interval would be infinity. It really skewed things so badly that we decided that we should just leave

it out and be conservative." (Burnham (2007))

First, any empirical researcher is vaguely suspicious of a result which just barely rejects the primary null hypothesis, in this case, that mortality in Iraq is unchanged. Given this testimony from Roberts and Burnham, isn't it likely that a small change in the model specification would lead to a confidence interval which includes zero? Since the authors refuse to provide *anyone* with the underlying data (or even a precise description of the actual methodology), there is no way for outsiders to know for sure. Second, almost all readers of L1 would conclude that excluding Falluja was "conservative" because the result would certainly be more statistically (and substantively) significant if the Falluja data is included. Or so these readers would naively assume.

Yet excluding Falluja is not "conservative." In fact, including this cluster — i.e., using all the available data — generates a result with such a wide confidence interval that the reported increase in Iraqi mortality becomes statistically insignificant.

L1 estimates a CMR_{pre} of 5.0 per thousand. We can translate the confidence interval given by the relative risk estimate into a CMR_{post} of 8.0 – 21 by multiplying CMR_{pre} by relative risk. (This ignores the uncertainty of the pre-invasion estimate.) Yet this result contradicts the direct estimate of post-war mortality which the authors provide. See Table 1.

In other words, their direct measure of the confidence interval for CMR_{post} is so wide that there is no way that their confidence interval for the relative risk can be correct. Note that the two results match fairly well for the upper bound of the confidence interval (e.g., 12.3 versus 12.0 and 23.2 versus 21) but not for the lower bound (8.0 versus 1.4). Furthermore, the more imprecise their measure of CMR_{pre}, the worse this conflict becomes.

The 98,000 excess deaths is the most reported statistic from L1. Consider how the authors calculate this number.

> "We estimated the death toll associated with the conflict by subtracting pre-invasion mortality from post-invasion mortality, and multiplying that rate by the estimated population of Iraq (assumed 24.4 million at the onset of the conflict) and by 17.8 months, the average period between the invasion and the survey."

Relative Risk and Crude Mortality Rate Estimates and Confidence Intervals						
	With Falluja			Without Falluja		
	2.5%	Est.	97.5%	2.5%	Est.	97.5%
Relative Risk (RR)	1.6	2.4	4.2	1.1	1.5	2.3
Post-invasion CMR	1.4	12.3	23.2	5.6	7.9	10.2
Post-invasion CMR*	8.0	12.0	21.0	5.5	7.5	11.5

Table 1: The first two rows are the RR and CMR_{post} estimates and 95% confidence intervals as reported in L1. The third, starred, row is my attempt to match the CMR_{post} by multiplying the CMR_{pre} estimate of 5.0 by the RR estimates and confidence interval. Note that the CMR_{post} confidence interval reported by L1 matches up well, with one notable exception: the lower bound of the CMR_{post} confidence interval when Falluja is included. Why is 8.0 so much greater than 1.4?

It is interesting to note that L1 only reports an excess death confidence interval without Falluja. Their formula seems to be:

$$\text{Excess Death} = (\text{CMR}_{\text{post}} - 5.0) * \left(\frac{\text{population}}{1,000}\right) * (\text{survey length})$$

Let us calculate an excess death confidence interval with Falluja. Use the CMR_{post} estimate and bounds of the confidence interval reported by L1 and the entire population of Iraq. We use each of these three values of CMR_{post} in the above formula to calculate corresponding estimate and confidence interval bounds for excess deaths. The calculations without Falluja are done with the population of Anbar (the governorate represented by Falluja, subtracted and using the post-invasion confidence interval that ignores Falluja.

See Table 2.

Figure 2 presents the distributions of CMR_{pre} and CMR_{post}.

From Figure 2, we can see how little we know about CMR_{post} because interval is so large. Visually, it is clear that there is a non-trivial probability that CMR has actually gone down. The basic intuition is obvious: a significant amount of the mass of the probability distribution for the estimate

Post-invasion Excess Deaths						
	With Falluja			Without Falluja		
	2.5%	Est.	97.5%	2.5%	Est.	97.5%
L1 Report	NR	298,000	NR	8,000	98,000	194,000
Calculated	-130,000	264,000	659,000	21,000	100,000	178,000

Table 2: The first row shows the estimate and confidence interval for excess deaths as reported in L1. Note that the confidence interval for the excess deaths including Falluja are not reported in L1. However, L1 mentions that the estimate would increase by 200,000 if Falluja were included, suggesting the excess deaths estimate with Falluja was 298,000. The second row, Calculated, presents my estimates for the death toll using L1's confidence intervals and formula. The data that L1 reports, that is the excess deaths excluding Falluja, very closely tracks the row that I calculate. Note how much wider the confidence interval becomes when Falluja is included, and that the lower bound is negative.

of CMR_{post} is less than the lower tail for the distribution of the estimate of CMR_{pre}.

The following sections prove that there is approximately a 10% chance that CMR_{post} is *lower* than CMR_{pre} using two different methods. If CMR_{post} is less than CMR_{pre}, then $\text{RR} < 1$. This means that the lower bound of the RR confidence interval, 1.6 is much to high and thus contradicts the RR confidence interval reported L1.

1.0.1 Proof Assuming Normal Distribution

Assume that uncertainty about the estimates for both CMR_{pre} and CMR_{post} is normally distributed.

$$\text{CMR}_{\text{pre}} \sim N(\mu_{\text{pre}}, \sigma^2_{\text{pre}})$$
$$\text{CMR}_{\text{post}} \sim N(\mu_{\text{post}}, \sigma^2_{\text{post}})$$

Consider the distribution of the difference between CMR_{pre} and CMR_{post} since it provides the probability that CMR_{post} is less than the CMR_{pre} and, therefore, the probability that $\text{RR} < 1$. If there is a significant probability that $\text{RR} < 1$, it then follows that the confidence interval for RR is far too narrow. Denote the distribution of the difference in CMRs as ΔCMR.

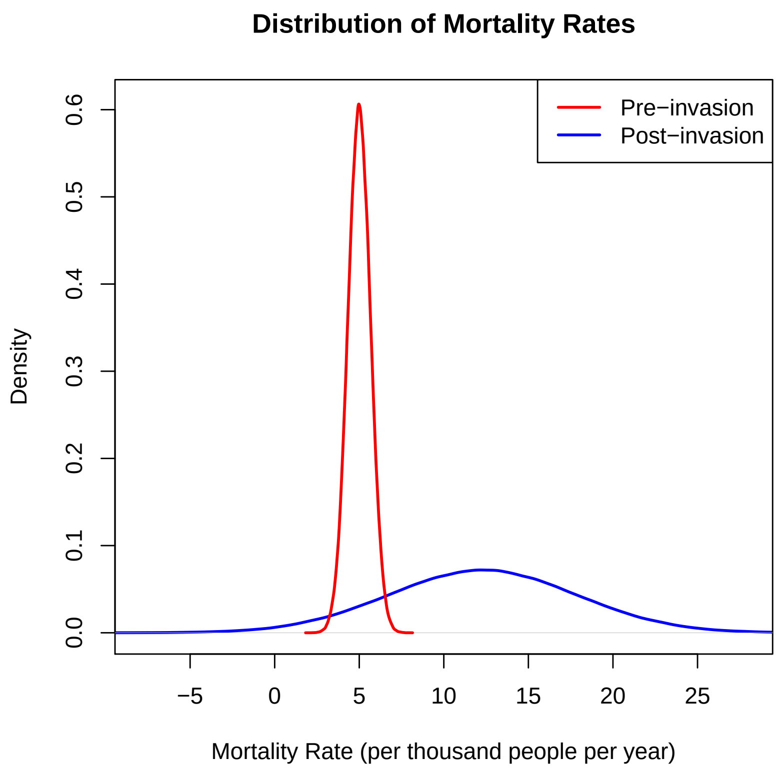

Figure 2: Assume a normal distribution for the probability densities for estimates of CMR_{pre} and CMR_{post}. The variance, σ, for CMR_{pre} and CMR_{post} is calculated from the L1 confidence intervals and corresponding z-scores: $\text{CMR}_{\text{pre}} \sim N\left(\mu = 5, \sigma^2 = \left(\frac{5-3.7}{1.96}\right)\right)$ and $\text{CMR}_{\text{post}} \sim N\left(\mu = 12.3, \sigma^2 = \left(\frac{12.3-1.4}{1.96}\right)^2\right)$. The density plots above were then created from simulated draws from the those distributions. This approach is not entirely correct since CMRs cannot be negative. However, calculations using a truncated normal distribution produce similar results.

$$\Delta\text{CMR} \sim N(\mu_{\text{post}} - \mu_{\text{pre}}, \sigma^2_{\text{post}} + \sigma^2_{\text{pre}} + 2\sigma^2_{\text{pre post}})$$

Plugging in the appropriate data and assuming a zero covariance yields:

$$\Delta\text{CMR} \sim N(7.3, 5.6^2).$$

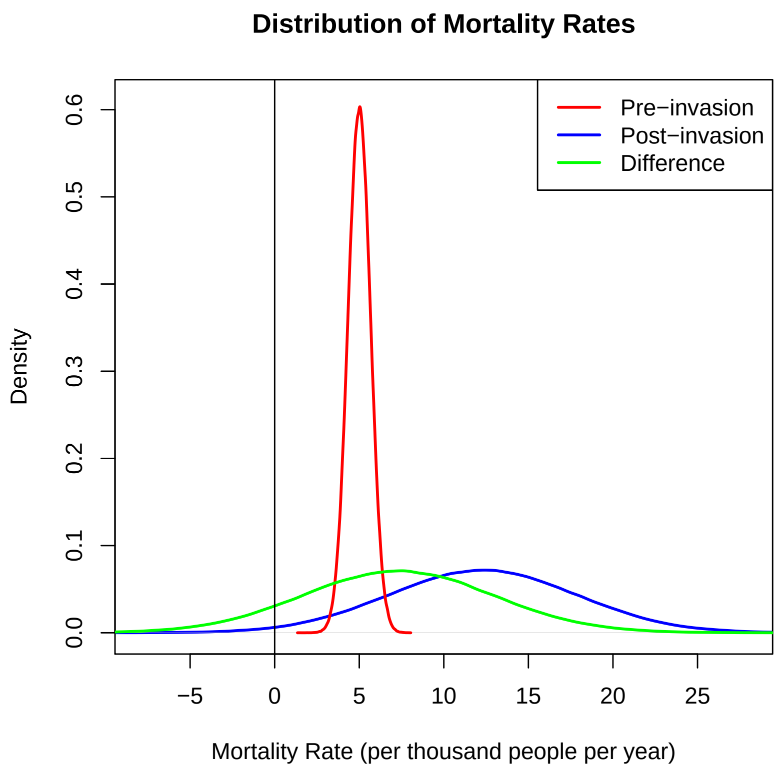

Figure 3: Simulated distributions for CMR_{pre}, CMR_{post} and ΔCMR. The covariance between CMR_{pre} and CMR_{post} is set at zero.

See Figure 3. If ΔCMR is less than 0, then the RR is less than 1 because CMR_{post} is less than CMR_{pre}. Given the data, there is a 10% chance that $\Delta CMR < 0$, i.e., that $RR < 1$. A simple simulation, assuming a zero correlation between estimates of CMR_{pre} and CMR_{post}, confirms these results.[5]

Not only is the probability of lower CMR_{post} being less than CMR_{pre} around 10%, but the lower bound of the RR confidence interval reported in L1, 1.6, must be far from the 2.5th percentile. In fact, it is at least at the 20th percentile, depending on the assumed correlation. See Figure 4 for the graphical evidence. The lower bound of the confidence interval for the

[5]The results do not change significantly for any other value of the correlation.

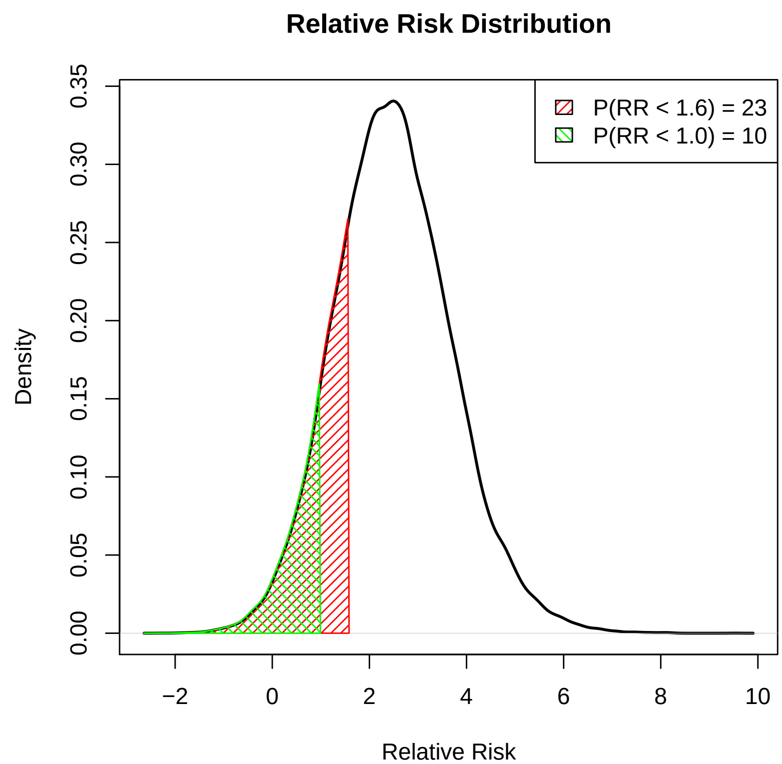

Figure 4: The distribution of relative risk assuming a correlation of zero between estimates of CMR_{pre} and CMR_{post}. RR values were generated by dividing normal draws from CMR_{post} by draws from CMR_{pre}. The probabilities that RR < 1 and RR <1.6 are also calculated and the corresponding area shaded.

relative risk must be much lower than 1.6.

1.0.2 Proof Assuming Unimodal Distribution

Relax the assumption that the uncertainty of the estimates of CMR_{pre} and CMR_{post} is normally distributed. Instead, just assume that it is unimodal unimodal.[6]. The strategy of the proof is the same as in the above: show that there is a non-trivial probability that CMR_{post} is less than CMR_{pre} and that,

[6]Thanks to Michael Spagat for outlining the mathematics of this argument

therefore, the lower bound of the RR confidence interval given in L1 is too high.

$$P(\text{CMR}_{\text{pre}} > 3.7) = 0.975$$

The probability that CMR_{post} is less than 3.7 can be broken up into two parts, the probability that it is less than 1.4 plus the probability that it is between 1.4 and 3.7. We choose 1.4 because, as reported in L1, this is the lower bound of the 95% confidence interval of CMR_{post}. Note that we cannot calculate the *total* probability that CMR_{post} is less than CMR_{pre} because we don't know the full details of the CMR distributions. Instead, we just calculate the joint probability that $\text{CMR}_{\text{post}} < 3.7$ and $\text{CMR}_{\text{pre}} > 3.7$. This probability is so high that it, alone, is inconsistent with the lower bound of the 95% confidence interval for relative risk as reported in L1.

$$\begin{aligned} P(\text{CMR}_{\text{post}} < 3.7) = &\ P(\text{CMR}_{\text{post}} < 1.4) \\ &+ P(1.4 < \text{CMR}_{\text{post}} < 3.7) \end{aligned}$$

Because we know that 1.4 is the lower bound of CMR_{post} 95% confidence interval, the probability that CMR_{post} is below 1.4 is 2.5%, so

$$P(\text{CMR}_{\text{post}} < 3.7) = 0.025 + X$$

where $X = P(1.4 < \text{CMR}_{\text{post}} < 3.7)$. Because we do not assume that CMR_{post} is normally distributed, X cannot be determined. However, because the distribution is increasing towards the mean, we know that

$$X > 0.025$$

because 0.025 is the mass on the interval from 0 to 1.4, whereas X is the mass on the longer interval from 1.4 to 3.7.

Let $d_{1.4}$ be the average density of the CMR_{post} probability distribution between 0 and 1.4 and let $d_{3.7}$ be the average density of the CMR_{post} probability distribution between 1.4 and 3.7. By definition,

$$\begin{aligned} 0.025 &= d_{1.4} * 1.4 \\ X &= d_{3.7} * (3.7 - 1.4) \end{aligned}$$

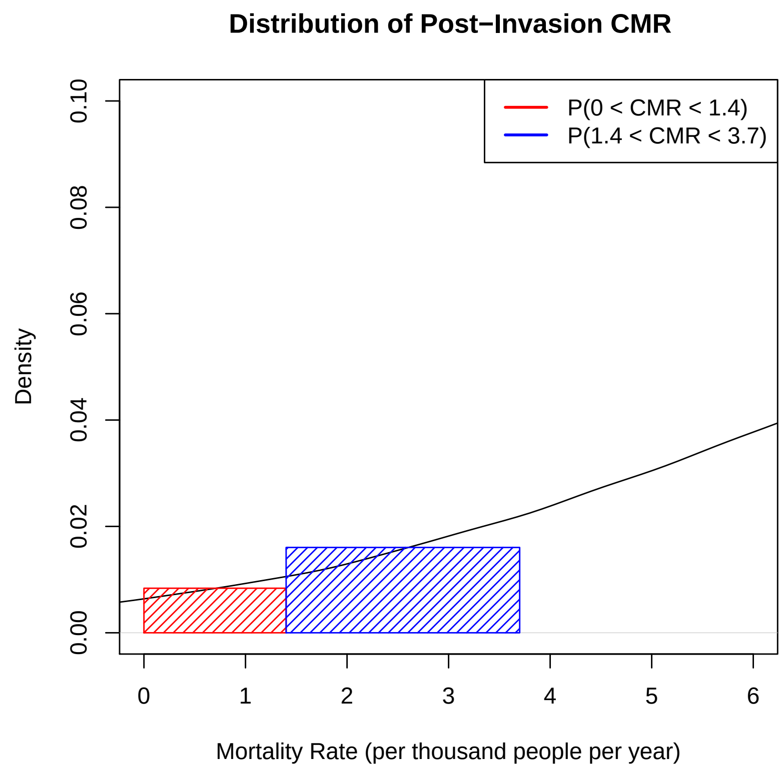

Figure 5: This figure clearly shows the basis of the average density argument. It is clear that the interval in question multiplied by the average density in that interval equals the area under the density curve in that same interval. Note that the total mass (area) between 1.4 and 3.7 is greater than the mass between 0 and 1.4. When assuming a distribution increasing towards the mean, even the worst case scenario, a uniform distribution, the mass between 1.4 and 3.7 will be larger.

Figure 5 provides a graphical representation of these average densities.

Because going from extreme values of CMR_{post} towards the central estimate should give increasing density, we know that $d_{3.7} > d_{1.4}$. Recall that $X = (d_{3.7}) * (2.3)$. Thus:

$$X > d_{1.4} * 2.3$$
$$X > 0.025 * \frac{2.3}{1.4}$$
$$X > 0.041$$

The joint probability that CMR_{pre} is greater than 3.7 *and* that CMR_{post} is less than 3.7 is the product of their individual probabilities multiplied together, assuming independence. Recall that:

$$P(\text{CMR}_{\text{pre}} > 3.7) = 0.975$$

and given that $X > 0.041$,

$$P(\text{CMR}_{\text{post}} < 3.7) = 0.025 + X$$
$$P(\text{CMR}_{\text{post}} < 3.7) > 0.066$$

we get

$$P(\text{CMR}_{\text{pre}} > 3.7 \text{ and } \text{CMR}_{\text{post}} < 3.7) > 0.975 * 0.066$$
$$P(\text{CMR}_{\text{pre}} > 3.7 \text{ and } \text{CMR}_{\text{post}} < 3.7) > 0.064$$

The probability that CMR_{pre} is greater than 3.7 and CMR_{post} is less than 3.7 is 0.064. As noted before, this probability is calculated around only a single point, 3.7, and is not even the total probability that CMR_{post} is lower than CMR_{pre}. Notice that L1 reports that there is only a 2.5% probability that CMR_{post} is less than 1.6 times greater than CMR_{pre}. This said, we have just shown that there is a 6.4% probability that CMR_{post} is less than 1 times greater than CMR_{pre}. Again, this is considering only one point, and not calculated over the entirety of both CMR_{pre} and CMR_{post} distributions.

From the calculations above, it is impossible to be 95% confident that there was an *increase* in mortality. The lower bound of the confidence interval for the relative risk can not be 1.6, as reported in L1. It must be much lower.

References

G. Burnham. Presentation at MIT, February 2007. URL http://web.mit.edu/webcast/tac/2007/mit-tac-wgbh-e51345-27feb2007-220k.%ram. [Online; accessed 14-May-2007].

G. Burnham, R. Lafta, S. Doocy, and L. Roberts. Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey. *The Lancet*, 368: 1421–1428, October 2006.

P. Mares. Counting the dead in Iraq, 2006. URL http://www.abc.net.au/rn/nationalinterest/stories/2006/1778810.htm.

L. Roberts, R. Lafta, R. Garfield, J. Khudhairi, and G. Burnham. Mortality before and after the 2003 invasion of Iraq: cluster sample survey. *The Lancet*, 364:1857–1864, October 2004.