"Mortality after the 2003 invasion of Iraq: A cross-sectional cluster sample survey", by Burnham et al (2006, Lancet, www.thelancet.com): An Approximate Confidence Interval for Total Number of Violent Deaths in the Post Invasion Period

Mark J. van der Laan
Division of Biostatistics, University of California, Berkeley
laan@stat.berkeley.edu

October 26, 2006

Context:

The following confidence interval is relying on the assumption that the design succeeded in obtaining a random sample of 47 geographical clusters of 40 households among the total population of Iraq households. As pointed out in our article (van der Laan, de Winter, 2006), beyond the fact that the used design should have been informed by known violence levels in Iraq and could thereby be strongly improved, we are concerned about a number of potential biases in the actual obtained sample which will only widen the confidence interval, and might even make it impossible to draw any conclusions based on a statistical foundation. Addressing some of these concerns will require additional information beyond the information provided in the Lancet article.

To name a few of our concerns:

1. Given the first selected household in each cluster, are the remaining 39 households really a random sample only correlated by its geographical location?

2. Is the actual selection of the first household and thereby cluster of 40 housholds truly comparable with randomly drawing a cluster in a bucket of clusters of households making up the total Iraq population,

3. Should we not incorporate the fact that a very significant percentage (minimally 20%, but possibly larger, since this information was not provided) of the death certificates corresponding with reported violent deaths have not been provided?,

4. Should we not deal with the fact that the cluster size is a random variable ranging in this sample from 169 to 344, which might be possibly correlated with the violence level in the sampled geographical area, and seriously investigate how this might affect the reliability of the reported extrapolated total violent death counts?

Another issue is that according to the Lancet article, 80% of the sampled deaths could be verified with a death certificate. Under the assumption that a registered death certificate is issued by a Government organization which presumably keep track of copies or originals of these death certificates, this raises the question why it is not possible to actual count the registered deaths.

Confidence interval calculation based on the 47 cluster counts for violent deaths.

The design of the survey involves dividing up Iraq in 18 Governorates. In these 18 areas the researchers randomly projected 47 clusters. In every cluster 40 households were questioned. Every cluster is supposed to represent 577.000 people based upon an estimated population size of 27.139.583 persons. In every cluster, one aimed to question 40 households. In other words: every household represents on average 14.452 people. The survey does not mention the use of a standardized questionnaire.

As the authors of the Lancet article did, we will make the optimistic assumption that the 47 cluster counts of violent deaths can be viewed as an random sample of independent and identically distributed observations involving the random sampling of a cluster of a random number of around 40 households and counting the number of violent deaths in this cluster of households. The random variable corresponding with this experiment will be denoted with X and the random variables corresponding with 47 repeats of this experiment will be denoted with $X_1, \ldots, X_{47}$. The data provided to us representing the 47 outcomes of $X_1, \ldots, X_{47}$ are given by the following set of numbers which should be read by row to follow the labelling of the clusters used in the Lancet article:

17	15	0	0	3	4	0	0
5	1	0	2	2	6	0	9
3	0	1	5	7	12	22	0
0	0	1	25	2	24	35	9
6	5	4	4	6	1	3	1
3	5	2	0	25	9	18	

Let's denote these observed counts with $x_1, \ldots, x_{47}$. The sample mean $\bar{X} = 6.43$ and the sample variance is $S^2 = 69.16$. We are concerned with using this data set to obtain an estimate of its mean $\mu = EX$ and a corresponding confidence interval. This true mean can under certain additional assumptions be extrapolated towards the true number N of violent deaths in all of Iraq, and as a consequence the sample mean and a corresponding confidence interval for μ can be upscaled towards a confidence interval for the true number N of violent deaths in all of Iraq during the studied post-invasion period. We will use the same factor (being Total population size divided by average size of cluster) as used in the Lancet article which mapped $\bar{X} = 6.43$ into a total number of 600,000 violent deaths in Iraq during the studied post-invasion period. Thus the factor we will use to map an estimate and confidence interval for μ into an estimate and confidence interval for the true number N of violent deaths in Iraq in the studied post-invasion period is $600,000/6.43 = 93312.6$.

Since we are confronted with a small sample size for a very spread out probability distribution there is no reason that a standard model for the distribution of the count X is appropriate, quite on the contrary. In addition, a nonparametric bootstrap method will also be very limited in its performance due to the fact that such a small sample for such a diverse and large country w.r.t. violence will not be representative: that is, the empirical distribution putting probability $1/47$ on each of the 47 numbers $x_1, \ldots, x_{47}$ will be a very poor approximation of the true probability distribution of X. Therefore, we will focus on a method which aims to be appropriate in finite samples. It is based on Bernstein's inequality (e.g., van der Vaart, Wellner, Empirical Processs, Springer, 1996), which provides an upper bound for the tail probability of the deviation of the sample average from the true average, given an upper bound for the variance $\sigma^2 = \mathrm{VAR}(X)$ of the (assumed to be bounded) random variables and their maximal value:

$$P\left(\mid \sum_{i=1}^{47}(X_i-\mu)\mid > x\right) \leq 2\exp\left(-\frac{x^2}{2(v+Wx/3)}\right),$$

where W is the largest possible value for $(X_i-\mu)$ and v needs to be an upper bound for the true variance $n\sigma^2$ of $\sum_{i=1}^{47}(X_i-\mu)$. Applying this inequality at $\sqrt{n}x$ with $v=n\sigma^2$ yields

$$P(\sqrt{47}\mid \bar{X}-\mu\mid > x) \leq 2\exp\left(-\frac{x^2}{2(\sigma^2+\frac{Wx}{3\sqrt{47}})}\right).$$

Define $q=q(0.95,W,\sigma^2)$ as the solution of

$$2\exp\left(-\frac{q^2}{2(\sigma^2+\frac{Wq}{3\sqrt{47}})}\right)=0.05.$$

Then, it follows that

$$P\left(\bar{X}-\frac{q(0.95)}{\sqrt{47}} \leq \mu \leq \bar{X}+\frac{q(0.95)}{\sqrt{47}}\right) \geq 0.95,$$

and thus that $\bar{X}\pm q(0.95)/\sqrt{47}$ is a 0.95 finite sample confidence interval. We could now obtain an approximate 0.95 confidence interval by replacing in the expression for $q(0.95,W,\sigma^2)$ W by an appropriate upper bound for the maximal number of violent deaths in a cluster of 40 households, and σ^2 by its estimate $S^2=69.16$. Off course, with large probability (e.g. >0.5) S^2 is not an upper bound for σ^2, since it is in fact an unbiased estimate of σ^2. It is considered common practice to accept this kind of approximation of a wished confidence interval. However, the standard error of S^2 will be muchlarger than the standard error in $\bar{X}$, where the latter is already large in this very variable experiment. Therefore, in this application we consider this a serious concern. Fortunately, we can investigate the sensitivity of our results w.r.t. this choice. Similarly, W should at minimal be larger than the maximum of all observed counts in this data set (35) minus μ, and the maximum of all observed counts in the previous 2004 study (52, in the Falluja cluster which is included in the 2006 study) minus μ. Therefore, $W=50$ is not too large, and seems very reasonable choice. Again, it is easy to evaluate the sensitivity of the reported confidence interval w.r.t. to these choices for σ^2 and W with the R-code provided below.

Calculation of $q(0.95)$:

Standard algebra shows that $q = q(0.95)$ solves the following quadratic equation (take the log, note $\log(0.025) = -\log(40)$)

$$q^2 + bq + cq = 0,$$

where

$$\begin{aligned} b &\equiv -\frac{2W\log(40)}{3\sqrt{47}} \\ c &\equiv -2\sigma^2\log(40). \end{aligned}$$

The two possible solutions are $\frac{-b\pm\sqrt{b^2-4c}}{2}$. Since the solution needs to be a positive number it follows that

$$q(0.95, W, \sigma^2) = \frac{W\log(40)}{3\sqrt{47}} + \frac{1}{2}\sqrt{b^2 + 8\sigma^2\log(40)}.$$

R-code

We now used the following trivial R-code to compute this 0.95 quantile and the corresponding confidence interval for μ. Here the function *leftlim* maps a value for W and σ^2 and sample size n (here $n = 47$) into the left-limit of the confidence interval for μ and the left-limit for the confidence interval for the true total count N of violent deaths. The function *confintdata* takes the data set $(x_1, \ldots, x_{47})$ and W and maps it into the approximate confidence intervals for μ and N by using as upper bound for σ^2 the sample variance S^2.

```
q095<-function(w,sigma2,n)
{ b<- w*log(40)/(3*sqrt(n))
b1<-2*w*log(40)/(3*sqrt(n))
b+0.5*sqrt(b1^2+8*sigma2*log(40))
}

leftlim<-function(w,sigma2,n)
{ a<-q095(w,sigma2,n)/sqrt(n)
leftcount<-6.43-a
```

```
lefttot<-leftcount*93312.6
c(leftcount,lefttot)
}

confintdata<-function(w,x)
{
n<-length(x)
sigma2<-var(x)
mu<-mean(x)
a<-q095(w,sigma2,n)/sqrt(n)
leftcount<-mu-a
rightcount<-mu+a
lefttot<-leftcount*93312.6
righttot<-rightcount*93312.6
c(leftcount,rightcount,lefttot,righttot)
}

confintdata(40,x)
[1] 1.921892e+00 1.092917e+01 1.793368e+05 1.019829e+06
 confintdata(50,x)
[1] 1.572296e+00 1.127877e+01 1.467150e+05 1.052451e+06
```

Using $W = 50$ (the Falluja cluster had a count of 52 violent deaths in the 2004 sample)and $\sigma^2 = S^2$ yields a confidence interval for N given by the approximate range $[150,000, 1,000,000]$.

Sensitivity of confidence interval to choice of guessed upper bound for σ^2:

One can investigate the sensitivity of this confidence interval by using other choices for the guessed upper bound of σ^2 than the known to be poor choice S^2, which will be smaller than σ^2 with probability possibly close or larger than 0.5. Therefore, we investigate what happens to the left-limit of the confidence interval if we replace S^2 by S^2 plus an estimate of its standard error, and S^2 plus twice this estimate of its standard error. One would expect that σ^2 is with high probability smaller than the latter upper bound so that this represents a particularly reliable confidence interval. The standard error of S^2 is based on the fact that S^2 is an asymptotically linear estimator of

σ^2 with influence curve $IC(X) = X^2 - EX^2 - 2EX(X - EX)$, so that the variance of S^2 can be approximated in first order with $\mathrm{VAR}(IC(X))/47$. We found that the standard error in S^2 is in this manner estimated to be 20.5. The two corresponding left-limits of the confidence interval using as upper bounds for σ^2 $S^2 + 20.5$ and $S^2 + 41$ now reduce to 107000 and $70,000$, respectively, as shown in the code below.

```
Function calculating estimate of standard error of $S^2$:

S2se<-function(x) { n<-length(x) barx<-mean(x) x2<-x^2
xc<-x-rep(barx,n) barx2<-mean(x2) x2c<-x2-rep(barx2,n)
ic<-x2c-2*barx*xc s2ic<-var(ic) s2<-s2ic/n sqrt(s2) }

Standard error of S^2:

>S2se(x)
 [1] 20.46923

Variance of data vector X:

>var(x)
 [1] 69.16281

>q095(50,69.16,47)
[1] 33.27168

Margin of error for choice W=50 and var upper bound 69.16

>q095(50,69.16,47)/sqrt(47)
 [1] 4.853174

Margin of error in average cluster count for choice $W=50 and var
upper bound 89.16 (S^2 plus 1 standard error)

>q095(50,89.16,47)/sqrt(47)
[1] 5.27131

mean(x) [1] 6.425532
```

```
Left-limit of confidence interval for average cluster count:

 mean(x)-5.27$ [1]
1.155532

Extrapolation factor used in Lancet article:
> F
[1] 93312.6

Left limit of confidence interval for total Iraq violent death count
using W=50 and var upper bound 89.16 (S^2 plus 1 standard error)

> F*1.155
[1] 107776.1

Margin of error in average cluster count for choice W=50 and var
upper bound 110 (S^2 plus 2 times standard error)

 q095(50,110,47)/sqrt(47)
 [1] 5.664517

Left limit of confidence interval for average cluster count

> mean(x)-5.66
[1] 0.7655319

Left limit of confidence interval for total Iraq violent death count
using W=50 and var upper bound 110 (S^2 plus 2 times standard error)

> 0.76*F
[1] 70917.58
```

Summary

Putting aside our mentioned reservations and concerns about the data, we constructed a finite sample 0.95 confidence interval for the true violent death count in the post invasion period studied in the Lancet article (39 months). This 0.95 confidence interval requires the specification of an upper bound for the maximal possible value W for $X - EX$ so that $Pr(| X - EX |< W) = 1$, and an upper bound for the true variance $\sigma^2 = \text{VAR}(X)$. We selected $W = 50$ because the previous study drew a cluster with a violent death count of 52 (Falluja cluster), but other choices can be tried out. Given this choice of $W = 50$, we consider three possible guessed upper bounds. If we use for the upper bound of σ^2 the sample variance S^2 (known to be a wrong upper bound with probability around 0.5), then we obtain the confidence interval

$$[150,000, 1000000].$$

If we use for the upper bound of σ^2 S^2 plus an estimate of its standard error, then we obtain as 0.95 confidence interval

$$[107000, 1090824]$$

. If we use for the upper bound of σ^2 S^2 plus 2 times this estimate of its standard error (this is normally a reliable upper bound for σ^2), then we obtain

$$[71000, 1127216].$$

Thus, under the assumption that the 47 cluster counts are a true random sample and the extrapolation factor is valid (relying on cluster size not being informative of violence in the area sampled), then one can state with confidence 0.95 that the violent death count in Iraq in the post invasion period is larger than 71000.

It will also be of interest to run some simulation studies to determine the coverage probabilities of bootstrap based confidence intervals and Bernstein's inequality based confidence intervals in a small sample size setting (relative to noise level), as in this particular study in which the standard error of the estimate of σ^2 is large.

0.1 Generalization of formula for confidence interval:

Suppose now that one would divide up Iraq in K areas. For each area we sample n_k clusters of households yielding a sample $X_{ki}, \ldots, X_{kn_k}$, $k = 1, \ldots, K$. Let μ_k be the mean count for the k-th area, and let $w(k)$ be a weight proportional to the population size of the k-th area, $k = 1, \ldots, K$, so that $\mu = \sum_k w(k)\mu(k)$ is the parameter of interest. Let $\bar{X} = \sum_k w(k)\bar{X}_k$, where $\bar{X}_k = \frac{1}{n_k}\sum_{i=1}^{n_k} X_{ki}$. We are interested in constructing a confidence interval for μ based on the estimate $\bar{X}$.

We have

$$\begin{aligned} P(\mid \bar{X} - \mu \mid > q) &= P(\mid \sum_k w(k)\bar{X}_k - \sum_k w(k)\mu(k) \mid > q) \\ &= P(\mid \sum_k \sum_{i=1}^{n_k} \frac{w(k)}{n(k)}(X_{ki} - \mu(k)) \mid > q). \end{aligned}$$

We can now apply Bernstein's inequality to $P(\mid \sum_k \sum_i Z(k,i) \mid > q)$ with $Z(k,i) \equiv \frac{w(k)}{n(k)}(X_{ki} - \mu(k))$, $i = 1, \ldots, n(k)$, $k = 1, \ldots, K$. Suppose that W is such that $P(\mid Z(k,i) \mid < W) = 1$. Let $\sigma^2(k) = \mathrm{VAR} X_{ki}$, $i = 1, \ldots, n(k)$, and let $\sigma^{2*}(k)$ be a known/guessed upper bound of $\sigma^2(k)$, so that $v \equiv \sum_k \frac{w(k)^2}{n(k)}\sigma^{2*}(k)$ is an upper bound of $\mathrm{VAR}\sum_k \sum_i Z(k,i)$.

By Bernstein's inequality we have

$$P(\mid \bar{X} - \mu \mid > q) \;\; \leq \;\; 2\exp\left(\frac{-q^2}{2\left(v + Wq/3\right)}\right).$$

Let $q(1-\alpha) = q(1-\alpha, v, W)$ be the solutionof setting the right-hand side equal to $1-\alpha$. It follows that

$$q(0.95) = \frac{1}{3}W\log(2/\alpha) + \frac{1}{2}\sqrt{\frac{4}{9}W^2\log^2(2/\alpha) + 8v\log(2/\alpha)}.$$

Thus, a finite sample 0.95 confidence interval is given by

$$\bar{X} \pm q(0.95, v, W).$$