The Lancet Surveys of Mortality in Iraq

David Kane*

First Draft: May 29, 2007
This Draft: June 20, 2007

Introduction

The Lancet published two controversial articles about mortality in Iraq: Roberts et al. (2004) and Burnham et al. (2006). This article and the accompanying **R** software package have four purposes. First, I provide data from Roberts et al. (2004).[1] Second, since the data from Burnham et al. (2006) has been made available to some researchers but not others, I provide a summary of the restricted data. Third, for those with access, I provide some basic functions for working with the data. Fourth, I provide some critical comments.

Data from Roberts et al. (2004)

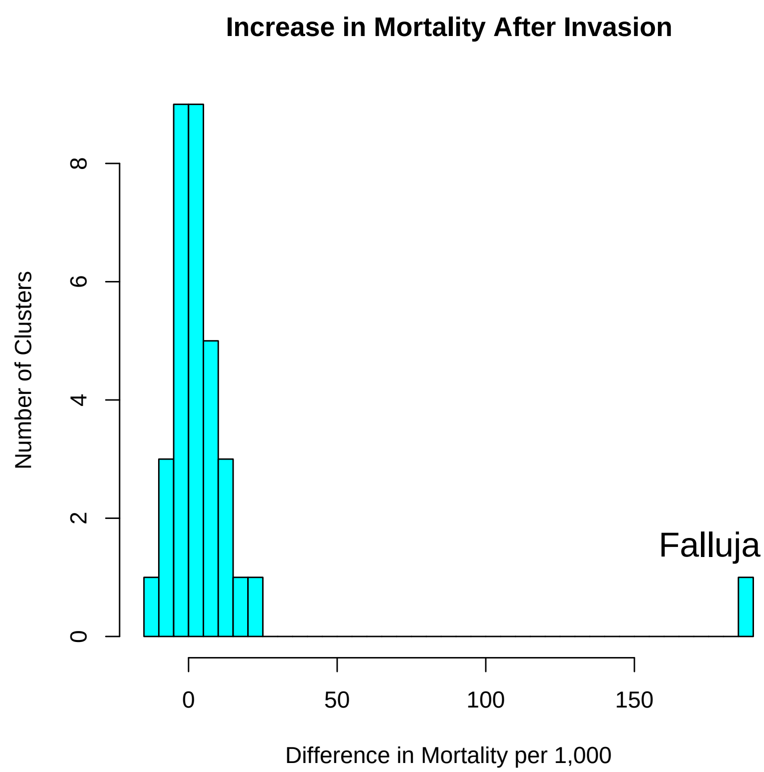
Typing `data(lancet1)` after installing the package loads the data frame. **lancet1** contains 20 variables and 33 rows, one for each of the clusters. The variables split broadly into two categories: pre-invasion and post-invasion.

*Institute Fellow, IQSS, Harvard University, Cambridge, MA 02138. dkane@iq.harvard.edu. Thanks to Arjun Ravi Narayan for excellent research assistance.

[1] I thank Tim Lambert of Deltoid for posting the summary data associated with Roberts et al. (2004). I thank the authors of Burnham et al. (2006) for making their data available to me. I urge them to make the data available to all. Shannon Doocy, Jeff Enos and Michael Spagat provided comments on an earlier draft. I also thank Shannon Doocy for patiently answering my questions about the data.

Statistics for each cluster include the number of births, deaths (infant and total), and persons alive (in different age groups).

The total person-months in each period is provided.[2] The mortality rate pre-invasion is calculated as $\frac{pre.deaths*12000}{pre.person.months}$. This same calculation is done for post-invasion mortality. The difference is used to calculate diff.mort.rate for each cluster.



Summary statistics for diff.mort.rate are:

```
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-14.5    -2.4     2.4     8.4     8.0   187.0
```

The Falluja cluster is the outlier, with an increase in mortality of 187 per 1,000. Removing this observation yields:

[2]The exact details are unavailable to us since the data does not include the timings of the births and deaths within each cluster.

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -14.5    -2.5     2.4     2.8     7.8    23.2
```

Since all the (public) data is available in the package, I do not provide further analysis here.[3] However, I am concerned that the presentation of the confidence interval in Roberts et al. (2004) is misleading. The authors claim:

> 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.

First, any empirical researcher is vaguely suspicious of a result which just barely rejects the primary null hypothesis, in this case, that the war has not increased mortality in Iraq. 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 the article would conclude that excluding Falluja was "conservative" because the result would certainly be more statistically (and substantively) significant if the Falluja data is included.

Yet it is not clear that excluding Falluja is "conservative." In fact, I *think* that including this cluster — i.e., using all the available data — generates a result with such a wide confidence interval that we can not even reject the primary null hypothesis that the war has not increased Iraqi mortality.

Consider Roberts (2006).

> There was one place, the city of Falluja which had just been devastated by shelling and bombing. It was so far out of whack with all the other [clusters] that it just made our confidence intervals very, very wide.

[3] Note that the authors have refused to provide data at any level more detailed than cluster aggregates. There is no way to perform a thorough comparison of the results of Roberts et al. (2004) and Burnham et al. (2006) without such access. Why do the authors provide (most of whom overlap between the two surveys) refuse to do so? I do not know.

Burnham (2007) goes further.

> We got a huge amount of criticisms for these confidence intervals, and I'll come to this confidence interval in just a bit. But we had a confidence interval at the low end of 8,000 and at the high end of 194,000. ...
>
> 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.

It is not "conservative" to leave out data when, if that data were included, your primary finding would disappear. Moreover, it is even worse not to tell your readers about this fact.

What is the intuition for how including a cluster with such a large increase in mortality could change the result from statistically significant to statistically insignificant? Consider the figure. The reason that cluster sampling is, sometimes, less accurate the fully random sampling is that the effect of interest may be "clumped." Imagine that all the war-deaths were in one city. It would be easy for a cluster survey to report no increase in mortality if, by chance, that unlucky city were not randomly selected. The phrasing may be awkward, but cluster sampling works less well when deaths are clustered.

The reason that the inclusion of Falluja explodes the confidence interval is that the pattern of mortality *across clusters* allows us to calculate how clustered death is in post-invasion Iraq. If you ignore Falluja, then the clumping of excess mortality does not look too problematic. Deaths are not concentrated in any one cluster. If, having looked at 32 clusters, you don't find any with especially big increases (or decreases) in mortality, you can conclude that your cluster sample is almost as good as a fully random sample. But, if you look at all 33 clusters, you see that there is evidence of significant clumping, that at least some cities are *dramatically* more violent than others.

The problem for Roberts et al. (2004) is that the existence of Falluja means that there might also be a city which featured a *decrease* in mortality which was just as dramatic as the increase in Falluja. Evidence of clumping in one direction means that there might be clumping in the other direction as well, that there might be a city in Iraq that was incredibly violent during the last year of Saddam Hussein's rule but which is now much more peaceful. The fact that Roberts et al. (2004) did not find such a city does not mean that it does not exist.

Of course, we have all sorts of prior evidence for doubting this scenario, for not believing that there was any place in Iraq that was as violent in 2002 as Falluja was in 2004. Yet the formulas for cluster modeling do not include this sort of prior information by default. The Falluja "outlier" is clear evidence that outliers exist and that, therefore, confidence intervals need to be very wide. How wide would the confidence interval be for Roberts et al. (2004) if the data from Falluja were included? Without more details on the precise methodology used, it is hard to say. Roberts et al. (2004) do admit that:

> 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).

Note the huge design effect (29.3) in the results for post-invasion mortality when Falluja is included. In other words, the authors have almost no idea what post-invasion mortality in Iraq actually is! It could be as low as 1.4 per 1,000 per year (making Iraq the healthiest country on Earth) or as high as 23.2 per 1,000. The dramatic clumping of death within clusters (caused by Falluja) precludes any firm inferences about post-invasion mortality.

The authors get around this problem by ignoring it.

> We tabulated data and calculated the number of births, deaths, and person-months associated with every cluster. For every period of analysis, crude mortality, expressed as deaths per 1,000 people per year, was defined as: (number of deaths recorded/number

> of person-months lived in the interviewed households) x 12 x 1,000. We estimated the infant mortality rate as the ratio of infant deaths to livebirths in each study period and presented this rate as deaths per 1,000 livebirths. Mortality rates from survey data were analysed by software designed for Save the Children by Mark Myatt (Institute of Ophthalmology, UCL, London, UK), which takes into account the design effect associated with cluster surveys, and reconfirmed with EpiInfo 6.0. We estimated relative and attributable rates with generalised linear models in STATA (release 8.0). To estimate the relative risk, we assumed a log-linear regression in which every cluster was allowed to have a separate baseline rate of mortality that was increased by a cluster-specific relative risk after the war. We estimated the average relative rate with a conditional maximum likelihood method that conditions on the total number of events over the pre-war and post-war periods, the sufficient statistic for the baseline rate. We accounted for the variation in relative rates by allowing for overdispersion in the regression. As a check, we also used bootstrapping to obtain a non-parametric confidence interval under the assumption that the clusters were exchangeable. The confidence intervals reported are those obtained by bootstrapping. [Footnotes omitted.]

There is a fair amount going on in this description and, without more detail and access to the underlying data, no outsider can reproduce the reported results. Yet I *think* that the key is that the authors report confidence intervals for the relative risk using a bootstrap procedure. Even though they have no good estimate of post-invasion mortality (when Falluja is included), they can make this uncertainly "disappear" with the bootstrap.

A standard cluster sample confidence interval calculation (Groves et al. (2004)) would force them to include the uncertainty associated with the huge design effect, would ensure that they allow for the *possibility* of a cluster with a decrease in mortality similar to the increase measured in Falluja. Such a confidence interval would, in all likelihood, include zero.

One way to *guess* at the correct confidence interval is to note that the design effect in the data when Falluja is included is almost 15 times greater than when it is excluded.

> More than a third of reported post-attack deaths (n = 53), and

> two-thirds of violent deaths (n = 52) happened in the Falluja cluster. This extreme statistical outlier has created a very broad confidence estimate around the mortality measure and is cause for concern about the precision of the overall finding. If the Falluja cluster is excluded, the post-attack mortality is 7.9 per 1,000 people per year (95% CI 5.6 – 10.2; design effect = 2.0).

The design effect is the ratio of the sampling variance for a statistic computed under the cluster plan to the sampling variance for the same statistic computed with a simple random sample. Since the denominator for the design effect with Falluja included and excluded is the same, the variance for estimates of statistics like the mortality rate should be approximately 15 times larger if the Falluja data is included. If the confidence interval without Falluja is 8,000 to 194,000 excess deaths then it *may be* that the confidence interval including Falluja is -400,000 to 1,000,000.[4]

Data from Burnham et al. (2006)

I have performed a series of manipulations to the raw data as distributed by the authors. See the functions `prep.deaths()` and `prep.houses()` for details. Each function produces a data frame. Call them **deaths** and **houses**.

deaths

With 629 rows and 14 variables, **deaths** includes a row for every death recorded in the study. The variables are:

id A number for each household interviewed. Multiple deaths from the same household share the same id. There is no unique id in the **deaths** data frame.

[4]This calculation is very rough. I assume that the confidence interval should increase in proportion to the square root of the increase in the design effect: $(194,000 - 8,000) * sqrt(29.3/2.0) \approx 700,000$. This new confidence interval would be centered around 300,000 since the authors report that the estimate for Falluja "indicates a point estimate of about 200,000 excess deaths in the 3% of Iraq represented by this cluster" and $98,000+200,000 \approx 300,000$. Again, if the authors will release the underlying data and computer code, we can determine the correct confidence interval. Comments on this calculation are welcome.

governorate The governorate in which the household of the deceased is located. There are 18 governorates in Iraq, but "miscommunication" led to only 16 being sampled by the survey team.

cluster The cluster in which the household of the deceased was located. There are 47 clusters. Clusters are numbered from 1 to 52. Numbers 17, 19, 29, 50 and 52 are missing. The authors sampled 52 clusters, but five were discarded. Falluja was purposely (Burnham (2007) and private communication) oversampled (3 clusters instead of 1) to check on the mortality estimate from Roberts et al. (2004). One of the three was selected at random (51) for inclusion while the other two (50 and 52) were discarded. Three other clusters (17, 19 and 29) were sampled by mistake. Removing the three mistakes and the two extra Falluja clusters leaves a final sample of 47.

date The death month. Although the actual date of death was recorded by the interviewers, the authors transform all dates to the first day of the respective months to protect participant identity. There are 57 missing dates.

year The year of death. There are no missing values. The number of deaths by year are:

```
2002 2003 2004 2005 2006
  64   80  141  194  150
```

The survey ended in July 2006, so mortality has increased every year.

invasion A binary variable (with values "pre" and "post") indicating whether the death occurred before or after the invasion in March 2003. This is provided even when there is no date of death. There are 547 deaths post-invasion and 82 deaths pre-invasion.

impute A TRUE/FALSE flag indicating whether an imputation was done to fill in the year and invasion variables even if the date of death was missing. There are 18 TRUE values for impute. In each of these cases, date is NA. There are a total of 57 NA dates. For the 39 observations with NA date and impute equal to FALSE, the interviewers were able to establish the year and whether or not the death happened after the invasion, so no imputation was necessary.

sex There were 485 male deaths and 144 female deaths.

age The age of the deceased person. Summary statistics are:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0	24	40	42	61	100	19

age.group The age group of the deceased. Following Table 2 in the article, categories are child (0-14 years), adult (15-59 years), and elderly (60+).

sex \ age.group	child	adult	elderly
male	49	292	127
female	31	60	51

This ignores the 19 deaths with NA age.

cause.summary A brief summary that describes the cause of death using one of 114 different categories. The five most common are:

gunshot from unknown	MI	CVA	bullet by USA army	exploded vehicle
91	61	45	44	38

death.nature A binary variable indicating whether the death was due to violent or non-violent causes. There are 302 violent deaths and 327 non-violent deaths.

	child	adult	elderly
violent	26	251	12
non-violent	54	101	166

This ignores the 19 deaths with NA age.

cause.category The authors aggregate cause.summary into 13 higher level categories.

gunshot	carbomb
169	38
other explosion	air strike
43	40
violent, unknown	old age
6	27
accident	cancer
36	48
heart disease or stroke	chronic illness
122	43
infection disease	infant death
4	40
non-violent, other	
13	

certificate A variable with one of three values: "yes" if a death certificate was available, "no" if a death certificate was not available, or "forgot" if the interviewers failed to ask for one. Burnham (2007) notes:

> So this time [Burnham et al. (2006) rather than Roberts et al. (2004)] our intent was to ask every household where a death was reported for a death certificate. Now, if you have done survey work you know you don't always get all the answers you want. And, in this case, in 13% of cases the interviewers forgot to ask for the death certificate.

Forgetfullness is not evenly distributed across clusters.

	Death Certificate Status		
Governorate	no	yes	forgot
Anbar	0	57	6
Basrah	1	51	1
Qadissiya	0	15	0
Sulaymaniyah	2	6	0
Babylon	0	28	0
Baghdad	0	100	60
Thi-Qar	11	19	0
Diyala	3	69	3

```
Erbil          1  21      2
Kerbala        1   8      0
Tameem         0   1      4
Missan         0  10      0
Ninewa        24  64      7
Wassit         0   5      0
Najaf          2  14      0
Saleh Al-Din   0  33      0
```

It is not clear why interviewers were so much more likely to forget to ask for death certificates in Baghdad.

houses

The second data frame, **houses**, has 1849 rows and 16 variables, one of which (id) can be used for linking to the **deaths** data frame. The period covers January 1, 2002 through the date of the survey, which ranges from May 20, 2006 through July 10, 2006.

id, governorate, cluster The same as in the **deaths** data frame.

size The number of household members at the time of survey.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
      1       5       6       7       9      26      13
```

males, females The number of males/females in the household. There are 55 NA values.

births The number of births in the household.

```
  0   1   2   3   4   5   7
924 531 284  78  23   8   1
```

deaths The number of deaths in the household.

```
   0    1    2    3    4    7
1323  440   74   10    1    1
```

immigration The number of immigrants to the household. There are 1720 houses with zero immigrants. The summary statistics of the remaining households are:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.0	1.0	3.0	4.2	6.0	21.0

emigration The number of emigrants from the household. There are 1697 houses with zero emigrants. The summary statistics of the remaining households are:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1	1	2	3	4	13

size.2002 The number of household members on January 1, 2002. This is the number of members at the time of the survey (size) minus births plus deaths. As with size, there are 13 NA values, consistent with those for size.

mid.2002 The average of size and size.2002. There are 13 NA observations, consistent with the NAs in size and size.2002.

has.death A binary variable indicating if there were any deaths in the household. There are 526 households with at least one death.

deaths.violent The number of violent deaths per household.

0	1	2	3	7
1598	208	39	3	1

deaths.nonviolent The number of non-violent deaths per household.

0	1	2	3	4
1558	262	23	5	1

Comments

This section provides a guided tour of some of the more interesting features of the data.

Clusters

The clusters in the data set are labeled from 1 to 51, with missing numbers at 17, 19, 29 and 50.[5] In general, cluster numbers are grouped within governorates. For example, the three Thi-Qar clusters are 11, 12 and 13. The five in Ninewa are 34 – 38. The two in Saleh Al-Din are 47 and 48. A prominent exception to this pattern is Baghdad, which includes clusters 14 – 24 (with 17 and 19 missing), 33, 39 and 40.

Burnham et al. (2006) state:

> Only 47 of the sought 50 clusters were included in this analysis. On two occasions, miscommunication resulted in clusters not being visited in Muthanna and Dahuk, and instead being included in other Governorates. In Wassit, insecurity caused the team to choose the next nearest population area, in accordance with the study protocol. Later it was discovered that this second site was actually across the boundary in Baghdad Governorate. These three misattributed clusters were therefore excluded, leaving a final sample of 1849 households in 47 randomly selected clusters.

Cluster 51 was in Falluja. The other two Anbar clusters are 30 and 31. Burnham (2007) mentions that they sampled three clusters in Falluja (which is in the Anbar province) even though the plan called for only one cluster. They did this because the Falluja data from Roberts et al. (2004) was such an outlier that they wanted a better estimate for this violent city. Having interviewed in three clusters, the authors then selected one of the three randomly. The selected cluster was the least violent of the three. This was cluster 51. The other two clusters in Falluja were numbered 50 and 52. The authors have declined to release the data for these two clusters. Rather than picking one of the three clusters to use, averaging over all three would generate a more precise estimate of the change in mortality.

Clusters 17, 19 and 29 were excluded from the analysis. The authors have refused to release the data for these clusters as well. For 2 of the three, it is unclear how the authors knew to exclude these specific clusters.[6] As long as these three clusters are similar to the 47 included in the analysis, there

[5] As discussed above, 52 is also missing.

[6] It is easy to determine that the Wassit cluster which was (mistakenly) moved to Baghdad should be excluded. But how did the authors decide which other two clusters to ignore? My *understanding* is that the authors wanted to have at least one cluster

should not be much of a problem. But, if those clusters are very different, either much more or less violent, then there might be a problem.

Interview Procedure

The paper states that the interviewers went to houses in each cluster until they completed 40 interviews.

> Empty houses or those that refused to participate were passed over until 40 households had been interviewed in all locations.

However, the data shows that only 29 of 47 clusters featured exactly 40 interviews. The following table shows the number of clusters for each total number of houses interviewed:

33	36	38	39	40	41
1	2	5	8	29	2

The clusters with 41 households were 23 (Baghdad) and 37 (Ninewa). Those with fewer than 38 were 3 (Najaf), 9 (Babylon) and 26 (Sulaymaniyah). This variation is troubling. If clusters in violent regions had more interviews than those in less violent regions, the calculated mortality rate *might* be too high. This problem may have occurred. For example, Sulaymaniyah (in the Kurdish north) should have featured 120 interviews, 40 for each of the three clusters. Instead, only 113 households were interviewed. Since Sulaymaniyah featured *no* violent deaths, interviewing fewer households there inflates the estimate of post-invasion violent mortality in Iraq.[7]

Yet the claim that interviews were conducted until 40 households agreed to participate is contradicted elsewhere in the article.

---

in each governorate. Unfortunately, the Iraqi survey teams failed to follow the correct procedure. Instead, they allocated clusters to governorates in proportion to population, without ensuring that each governorate received at least one cluster. Such mistakes happen in field work all the time. It is fair to say that two of the clusters should have been in Muthanna and Dahuk. But how can the authors determine which of the 50 clusters were the "mistakes," were the clusters that would not have been sampled had the survey teams followed the correct procedure?

[7]This would depend on the precise methodology used to calculate excess mortality. If the calculation were based on estimating mortality within each cluster and then aggregating these cluster estimates, it would not bias the results if less violent clusters had more interviews. In private communication, Shannon Doocy reports that just such a cluster-based estimation strategy was used. Yet the paper reports:

> In 16 (0.9%) dwellings, residents were absent; 15 (0.8%) households refused to participate. In the few apartment houses visited, the team progressed to the nearest households within the building. One team could typically complete a cluster of 40 households in 1 day.

If 40 households had been interviewed in each of the 47 retained clusters, there would be 1,880 rows in **houses**. The claim that 31 households were either absent or refused to participate is consistent with the 1849 observations actually present. It is unclear how one can reconcile this description (absences/refusals were not replaced) with the previous claim that "40 households had been interviewed in all locations."

Death Certificates

The authors claim that missing death certificates, whether caused by the failure of the interviewer to ask or by the inability of the interviewee to produce one, are not a problem because there is no correlation between this missingness and other variables.

> Data entry and analysis was done with Microsoft Excel, SPSS version 12.0, and STATA version 8. Period mortality rates were calculated on the basis of the mid-interval population and with regression models. Mortality rates and relative risks of mortality were estimated with log-linear regression models in STATA. To estimate the relative risk, we used a model that allowed for a baseline rate of mortality and a distinct relative rate for three 14-month intervals post-invasion — March, 2003 – April, 2004, May, 2004 – May, 2005, and June, 2005 – June, 2006. The SE for mortality rates were calculated with robust variance estimation that took into account the correlation between rates of death within the same cluster over time. The log-linear regression model assumed that the variation in mortality rates across clusters is proportional to the average mortality rate; to assess the effect of this assumption we also obtained non-parametric CIs by use of bootstrapping. As an additional sensitivity analysis, we assessed the effect of differences across clusters by extending models to allow the baseline mortality rate to vary by cluster. [Footnotes omitted.]

In general, a "robust variance estimation" would make use of sample size information. A cluster with 1,000 observations provides more precise estimates than one with 10. Again, without access to the details of the exact models (ideally the precise commands used in Stata and SPSS), it is impossible for any outsider to know whether or not varying sample sizes across clusters, correlated as they are with cluster mortality, influences the results.

> Survey teams asked for death certificates in 545 (87%) reported deaths and these were present in 501 cases. The pattern of deaths in households without death certificates was no different from those with certificates.

There were 501 deaths with death certificates, 45 without and 83 observations where the interviewer forgot to ask for one. The two locations with the largest number of "forgot" values were both in Baghdad: cluster 33 (24 deaths) and cluster 24 (10 deaths).

There were 35 deaths for which a certificate was provided, but no date of death is listed. This is interesting as one would expect that death certificates provide the date of death. There were 14 deaths for which the death certificate was not asked for and the date of death is NA. It would seem especially important to ask for a death certificate when the interviewee can not recall the date of death.

The asking rate for death certificates is correlated with the year of the death — the later the year, the higher the likelihood of the interviewers not asking for a death certificate. It is unclear why interviewers would be more likely to "forget" to ask for a certificate if the death occurred in 2006 rather than 2002.

```
                    Year
Death Certificate Status 2002 2003 2004 2005 2006
                  no        4    6    4   20   11
                  yes      58   65  124  151  103
                  forgot    2    9   13   23   36
```

Of the 64 deaths in 2002, the interviewers forgot to ask for a death certificate 3% of the time. In 2006, it hits 24%. Comparing pre-invasion and post-invasion forgetfullness yields the same result.

```
                    Invasion
Death Certificate Status pre post
                  no        6   39
                  yes      71  430
                  forgot    5   78
```

Why are interviewers more than twice as likely (14% versus 6%) to forget to ask for a death certificate if the death occured after the invasion? A similar issue arises with the nature of the death.

Death Certificate Status	Nature of Death: violent	non-violent
no	22	23
yes	210	291
forgot	70	13

It is unclear why the nature of a death would make an interviewer more likely to forget to ask for a death certificate. In the case of non-violent deaths, the interviewers forget to ask 4% of the time. For violent deaths, the forget-to-ask rate is 23%.[8]

Cluster 33

Cluster 33 has a total of 25 deaths. All 24 of the deaths where the interviewers failed to ask for a death certificate had the same cause ("exploded vehicle") and occurred in the same month (July 2006). Since data collection ended on July 10th, these deaths must have happened prior to that date. It is not clear why death certificates were not asked for. Since the survey started on May 20th, the interviewers, by this point, had 6 weeks to learn to remember to ask for death certificates. It is also not clear how they remembered to ask for a death certificate in the one other death in this cluster, a non-violent heart attack.

Besides the issue of missing death certificates, there are two other problems with this cluster. First, the authors report that "deaths from car explosions have increased since late 2005." This is true but misleading. Consider total mortality from car bombs in 6 month intervals.

pre-2004	Jan-Jun 2004	Jul-Dec 2004	Jan-Jun 2005	Jul-Dec 2005	Jan-Jun 2006	Jul 2006
0	2	4	3	3	2	24

[8]In private correspondence, the authors note that this sort of analysis ought to be conducted at the level of households rather than individual deaths since deaths without certificates are often clustered (?) within households. The authors also note that these patterns may have been driven by variations in forgetfulness which are correlated with violence levels across governorates. Interviewers forgot to ask for death certificates much more often in governorates with moderate levels of post-invasion violence. Of the 83 forgot-to-ask deaths, 60 were in Baghdad with 24 of these in the last few weeks of the survey. I hope to do further analysis on these topics in the next draft.

There were no deaths from car bombs prior to 2004, but mortality was largely constant for the 2.5 years from January 2004 through June 2006. There were 6 deaths in 2004, 6 in 2005 and 2 in the first six months of 2006. The entire rise in car bombs deaths comes from the data for one month in one cluster.

The second problem with this cluster is having so many deaths from car bombs (or, more likely, a single car bomb) centered on a single neighborhood. Johnson et al. (2006) argue that the methodology of Burnham et al. (2006) generated a "main street bias" because interviewers were more likely to select households near main streets where violence is more common. This cluster would seem to provide evidence that main street bias might be a concern.[9]

Conclusion

I hope that this brief description of the data is useful to those who are interested in *The Lancet* articles but do not have access to the data and that the tools provided here are also useful to those who have such access. I hope to extend this article in the near future to include more analysis. Please contact me with suggestions.

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.

[9]The authors argue that a weakness of the main street bias argument is that requires that deaths occur at or close to home. If this car bomb exploded in a market which served the neighborhood in which these victims lived, it is likely to have injured people from throughout the neighborhood, whether they lived on a main street, a side street or an out-of-the-way alley. Main street bias is only a concern if the deaths caused by this car bomb were from households on the street on which the car bomb exploded. The authors argue that most deaths occur away from home and that, therefore, main street bias is not a concern.

R. M. Groves, J. F. Floyd J., M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau. *Survey Methodology*. John Wiley & Sons, 2004.

N. F. Johnson, M. Spagat, S. Gourley, J.-P. Onnela, and G. Reinert. Bias in epidemiological studies of conflict mortality, 2006.

L. Roberts. Interview on The National Interest, November 2006. URL `http://www.youtube.com/watch?v=3tU6xoQd_jo`. [Online; accessed 14-May-2007].

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.