

The Discussion on Possible
Sampling Bias in the Second
Lancet Study of Mortality in Iraq

Michael Spagat
Department of Economics Department,
Royal Holloway College
University of London

Presentation given at the third conference of
the Households in Conflict Network

IDS
University of Sussex
December 11, 2007
Updated on April 22, 2009
with a correction on page 7.

[Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey \(L2\)](#) by Burnham et al. published in *The Lancet* 2006.

“We estimate that as of July, 2006, there have been 654,965 (392,979 – 942,636) excess Iraqi deaths as a consequence of the war, which corresponds to 2.5% of the population in the study area. Of post-invasion deaths, 601,027 (426,369-793,663) were due to violence, the most common cause being gunfire.” (From the summary on the first page of L2).

The sheer size of these figures compared to other sources immediately raised some red flags. For example:

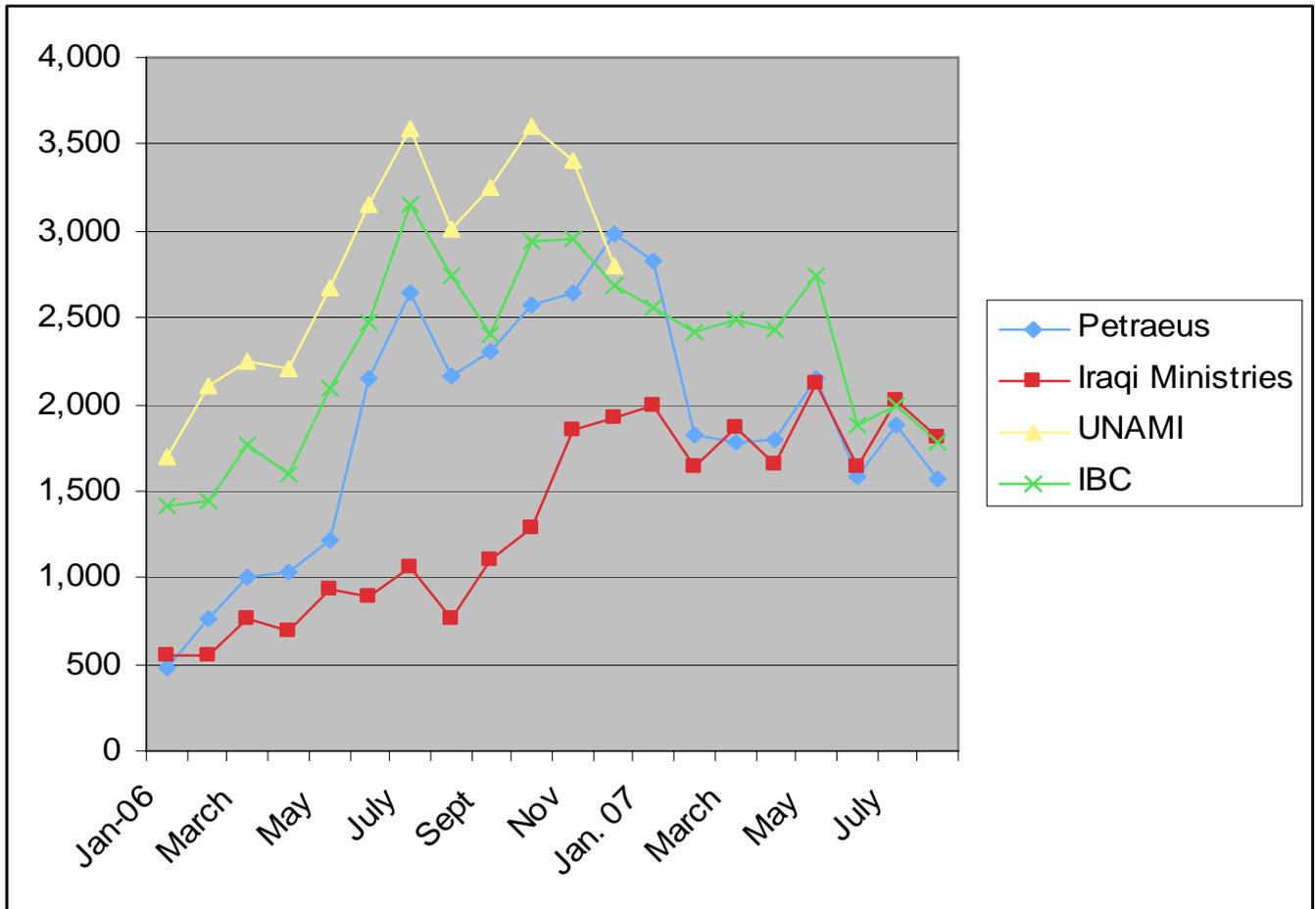
1. An L2 estimate for car-bombing deaths would be just below 80,000 compared to about 6,000 people killed in car bombs that were noticed by international media sources (source, Iraq Body Count Project).

a. Car bombs are big news of exactly the sort the media loves to cover.

b. Both insurgents and government official have strong reasons for making sure that car-bombings get noticed.

c. Car bombs are noticed by many people in their immediate vicinity, most of whom are carrying cell phones: the explosions go off in crowded areas, can be heard over a great distance and cause local chaos.

2. L2 implies an average of about 30,000 violent deaths per month during the first half of 2006, more than ten times the rates recorded by a wide variety of different monitoring mechanisms:



(See my recent lecture at George Mason University for some details on these sources.)

What can explain these very high numbers?

Together with three physicists and a statistician I explored the possibility that sampling bias might have contributed to some, possibly substantial, overestimation (now accepted subject to minor revisions at the *Journal of Peace Research*.)

"The third stage consisted of random selection of a main street within the administrative unit from a list of all main streets. A residential street was then randomly selected from a list of residential streets crossing the main street." (From L2)

The field teams would then select a household on this residential cross street to a main street and do interviews at 40 *contiguous* households, according to the sampling methodology published in *the Lancet*.

Our group argued that residential cross streets to main streets *would suffer from higher-than-average violence* within the context of the Iraq war because:

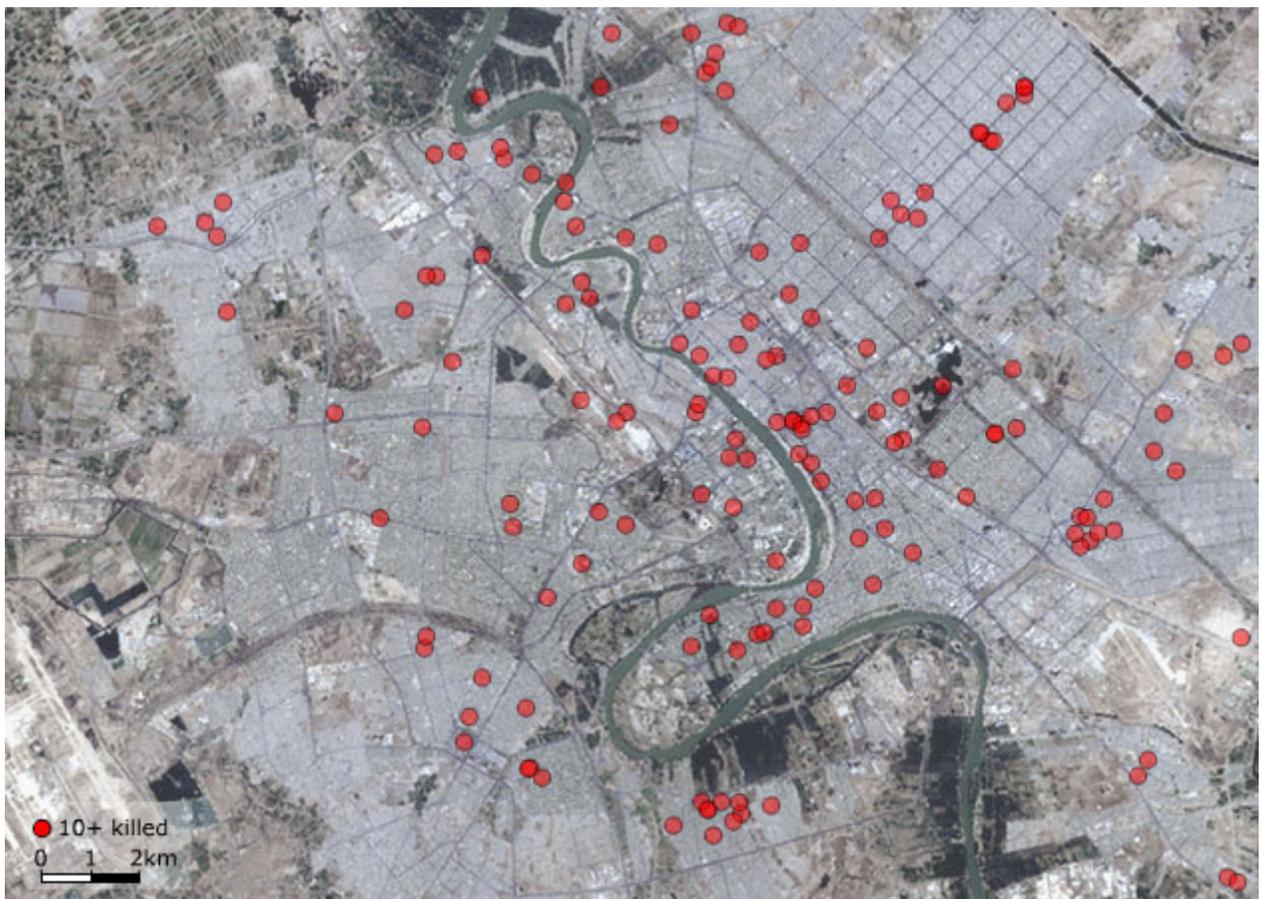
a. Crowded markets, cafes restaurants and other attractions will be on such streets.

b. Military patrols focus on such streets. In fact, many military vehicles can only go down the larger streets.

c. Abductions and mass shootings will also tend to be on such streets. For example, Sunnis would not travel deep into Shiite territory, abduct some people and make a long drive to reach safe territory. Rather, they will make a quick foray in and out of enemy territory, perhaps just crossing over a main street that divides the two areas, just into a residential area.

This map seems to suggest that large attacks in Baghdad could be biased toward residential cross-streets to main streets

Baghdad: Mapping the violence
Attacks since May 2003 in which more than 10 people were killed.



Note that incidents of this size almost certainly cover under half of all deaths.

Here is a map pinpointing one particular bombing that killed 34 children who were gathering candy that had been tossed out by soldiers from a US Humvee.



This attack seems to have taken place precisely on a residential cross street to a main street.

This picture illustrates the types of areas that will be missed by a methodology of doing interviews at 40 contiguous households beginning at a household on a residential cross street to a main street.



Scope is limited for reaching areas not on residential cross streets to main streets.



This picture suggests that the *areas across the country that would have been outside the survey space might have been very large.*

Anyone can use (the amazing) Google Earth facility to evaluate this for his or her self.

But in the end *much hinges on what which streets are actually treated as main streets in the sampling methodology.*

"The third stage consisted of random selection of a main street within the administrative unit from a list of all main streets." (Emphasis added).

These lists of main streets at the core of the stated sampling procedures are crucial for evaluating the extent of possible bias in L2.

Unfortunately, the L2 authors have thwarted analysis of which areas are inside and which areas are outside the survey space by refusing to release their lists of main streets or even say where these lists came from.

What about information on the cross streets?

“The interviewers wrote the principal streets in a cluster on pieces of paper and randomly selected one. They walked down that street, wrote down the surrounding residential streets and randomly picked one. Finally, they walked down the selected street, numbered the houses and used a random number table to pick one. That was our starting house, and the interviewers knocked on doors until they’d surveyed 40

households.... The team took care to
destroy the pieces of paper
which could have identified households if
interviewers were searched at checkpoints.”
(Emphasis added.) [Interview with Gilbert
Burnham in the New Scientist, April 2007](#)

In other words, the L2 authors *have destroyed much of the evidence of how they sampled.*

Note that pieces of paper with lists of main streets and cross streets would be of no use for identifying households included in the survey.

Even lists of all houses and flats on a street would not be usable for identifying the households that were actually interviewed.

On the other hand, the L2 data-entry form that L2 author Riyadh Lafta submitted to the WHO contains spaces for listing the *name* of each head of households in addition to *names* of people who died or were born during the L2 sampling period:

Governorate	Cluster No.	House No.	Name of householder	
No. of family members	Males	Females		
No. of live births since 2002:	Name	sex	Date of birth	
1.	
2.	
3.	
No. of deaths since 2002				
Name	Sex	Age	Date of death	cause (in details):
1.
2.
3.
Presence of death certificates:	Yes	No		
Hospitalization due to violence:	Age	Sex	Date	cause
In-migration	out-migration (during that period)			

If the field teams could travel around with all these names they did not have to destroy pieces of paper with street names written on them.

The circulation of people between inside and outside the survey space is an important factor.

If there were perfect circulation of people between areas inside the survey space and areas outside the survey space then there would be no sampling bias, even if violence levels are higher inside the survey space than they are outside.

Of course, people living inside the survey space will tend to be disproportionately represented inside the survey zone and vice versa although the precise size of this home bias is hard to determine with great accuracy.

However, the following slide suggests that this home bias could be quite strong (if we are to believe in this highly dubious piece of the L2 dataset).

24 deaths out of the 66 occurring in the incident actually appear in the L2 dataset!

Baghdad market blast kills scores

A huge explosion has ripped through a busy Baghdad market, killing at least 66 people, officials say.

[OPEN](#)



About 100 others were injured in the car bomb attack in Sadr City, a Shia area frequently targeted by insurgents.

The explosion left a scene of carnage and devastation, with the dead and injured lying amid the wreckage of cars, shops and market stalls.

The new government has been battling to improve Baghdad security, and last weekend unveiled a national unity plan.

[Enlarge Image](#)

But the attack was the worst incident in the capital for weeks.

Reports say the car bomb was detonated as a police patrol passed, causing both police and civilian casualties.

The bomb was clearly aimed at causing the maximum possible casualties, says the BBC's Jim Muir in Baghdad.

It was detonated at one of the busiest times of the day in a popular market in a densely-populated area.



[In pictures: Baghdad blast](#)

The appearance of so many of the deaths from this one event in the L2 dataset suggests either very substantial sampling bias or that these deaths were added in improperly into the L2 sample.

The burden is on the L2 authors to evaluate the extent of possible bias in their sample.

There are two main arguments the L2 authors have used to suggest that sampling bias is not a problem for L2.

1. They claim that the results in L2 are very similar to the results they obtained in a [previous survey](#) also published in the *Lancet* in 2004 (L1) for the period covered by L1.

[Elsewhere we have dealt with this claim extensively.](#)

As a quick summary I note that L2's central estimate for violent deaths during the L1 sampling period exceeds L1's central estimate for violent deaths by more than a factor of two.

So this comparison hardly disposes of the possibility of sampling bias in L2.

2. The L2 authors have often defended themselves by stating that they did not actually follow the sampling procedures that they claimed to have followed in their *Lancet* publication.

For example, Les Roberts and Gilbert Burnham wrote in a [letter to *Science*](#) that peer reviewers and editorial staff at the *Lancet* had gotten them to remove the following sentence from their description of their sampling methodology:

"As far as selection of the start houses, in areas where there were residential streets that did not cross the main avenues in the area selected, these were included in the random street selection process, in an effort to reduce the selection bias that more busy streets would have."

Thus, this part of the description of sampling methodology should have read:

"The third stage consisted of random selection of a main street within the administrative unit from a list of all main streets. A residential street was then randomly selected from a list of residential streets crossing the main street. *As far as selection of the start houses, in areas where there were residential streets that did not cross the main avenues in the area selected, these were included in the random street selection process, in an effort to reduce the selection bias that more busy streets would have.*" (new text italicized)

Combining this with Gilbert Burnham's New Scientist interview already quoted, we have (apparently):

1. Field teams wrote names of main streets on pieces a paper and selected one at random.

2. They walked down this street writing down names of cross streets and then selected one of these.

- 3? The field teams then somehow became aware of other streets in the area that did not cross the main avenues. The teams may have selected some of these according to an as-yet undisclosed procedure.

Key Points

1. If the description of the sampling methodology published in the *Lancet* is incorrect then the *Lancet* should publish a correction.
2. The extra sentence actually concedes that the published procedures are biased toward violent areas and that the modification to these procedures, whatever it was, only reduced this bias.

3. The pieces-of-paper description of Burnham may be biased but it is, at least, an operational procedure that can be followed.

If other types of streets, beyond those covered by the published methodology, were included then we need to know *how* these streets were actually included.

Even more basic, how did the field teams even discover the existence of these streets that could not be seen by walking down main streets?

Clearly the teams did not have detailed street maps before they entered each area: otherwise it would not have been necessary to walk down selected streets writing down names of surrounding streets on pieces of paper.

We can also rule out the possibility that the teams completely canvassed entire neighborhoods, building up detailed street maps from scratch; this would have taken a very long time and the teams had to follow an extremely compressed schedule that required them to perform forty interviews in a day.

In [an article in Nature](#) the L2 authors floated one possible explanation on how the field teams had augmented their street lists but it was promptly rejected by an actual field-team member.

“But again, details are unclear. Roberts and Gilbert Burnham, also at Johns Hopkins, say local people were asked to identify pockets of homes away from the centre; the Iraqi interviewer says the team never worked with locals on this issue.”

Even if locals had identified such “pockets of homes away from the centre” we would still have needed to know how these were included in the randomization procedures.

To summarize, appending the new sentence to the published version of the sampling methodology converts the study from one *with* a specified sampling methodology to one *without* a specified sampling methodology.

A sample survey cannot be taken seriously if its authors do not specify its sampling methodology.

Further Contradictions

"The sites were selected entirely at random, so all households had an equal chance of being included."

([The Human Cost of War](#), emphasis added)

"Our study team worked very hard to ensure that our sample households were selected at random. We set up rigorous guidelines and methods so that any street block within our chosen village had an equal chance of being

selected." ([Exchange of letters in Slate](#), emphasis added)

"... we had an equal chance of picking a main street as a back street." ([Les Roberts on ABC Radio International](#))

These statements contradict each other and the methodology published in the *Lancet*.

Some streets are much longer than others.

Some streets are much more densely populated than others.

Such varied units cannot all have equal probability of selection.

Conclusions

The sampling procedures published in the *Lancet* may contain substantial upward bias.

If these are not the true procedures for the study then, if the study is to have scientific standing, we need to learn what the true sampling procedures were.