

First Thoughts on the *National Journal* Articles on the *Lancet* Surveys of Iraq

By

Michael Spagat
Department of Economics
Royal Holloway College
University of London

In this short piece I focus on what is new in the *National Journal* (NJ) package of articles:

<http://news.nationaljournal.com/articles/databomb/index.htm>

(Make sure to click on the many sidebars and links within the main article.)

There are a few caveats. First, this is a preliminary reading and is work in progress. Comments are welcome. Second, I do not try to be exhaustive and list every little new thing. Rather, I try to select the most important new pieces of information. Third, one of the biggest contributions of the NJ package is to pull together and organize a very large amount of material, both new and old. Yet, even this effort is still far from comprehensive. I ask readers of this document to also consult two recent presentations of mine that further develop this story and also contain references to other good works on this subject:

http://personal.rhul.ac.uk/uhte/014/gmu_final.pdf

<http://personal.rhul.ac.uk/uhte/014/Households%20in%20Conflict%202007.pdf>

I will organize the following material by topic. Note that I will refer to the first and second *Lancet* surveys of Iraq as L1 and L2 respectively.

A. Ethics

This material is mostly contained in the [sidebar “Unscientific Methods”](#). In my morning perusals of blog activity on the NJ article I did not notice anyone picking up on this shocking aspect of the story.

1. NJ calls attention to the patient protection regulations of the US federal government, [provides a handy link](#) to these regulations, and makes a good case that these regulations have been violated by the L2 study. NJ gives the further information that anyone can make a complaint with the Office for Human Research Protections of the Health and Human Services Department.

2. NJ calls attention to the role of Johns Hopkins' [Institutional Review Board](#) (IRB) in protecting subjects of Hopkins research. The material on this Board is worth perusing. Several of the Board's own guidelines seem to have been violated by L2.

Gilbert Burnham talked a bit with the NJ about the ethical review and then refused to answer further questions. Hopkins as an institution stonewalled the NJ. Dean Michael Klag refused to comment on any aspect of the study and the University sent a terse, legalistic statement to the NJ on L2's ethics.

Hopkins non-cooperation went to the extreme of refusing to disclose even the data-entry form that was approved for the L2 study. Outrageously, the L2 team has refused to disclose both its questionnaire and its data-entry form. [No survey can have scientific standing if it holds back such critical information](#). Hopkins as an institution is now implicated in this significant breach of scientific standards.

Why would Hopkins refuse to release a blank L2 data-entry form? The NJ supplies enough material to form a good hypothesis on this question. It reports that Riyadh Lafta actually did hand over a data-entry form to the World Health Organization (WHO) which [the NJ helpfully posts on its web site](#). If this data-entry form was actually implemented in the field then the L2 field teams recorded names of heads of households plus names of household members who were born or died during the period covered by L2. In other words, if Lafta gave the true data-entry form to the WHO then the completed L2 data-entry forms will be filled with names. As I will show below, recording names in this way would cause significant ethical problems for the L2 study and for Hopkins.

Burnham made some damaging disclosures about the IRB process before he refused to answer further questions:

a. The reason why L2 was allowed to depart from standard procedures and obtain only oral consent of potential respondents, rather than the usual written consent, was "because of the fear that militias might capture the signatures from the surveyors and then attack the respondents." (The quote is from the NJ, not directly from Burnham.) This means that the IRB treated the studies' respondents, correctly, as a vulnerable population.

Several consequences flow from this designation of vulnerability to L2's study population but one stands out particularly. If the data-entry form that Lafta submitted to the WHO was approved by the IRB, then the IRB would have taken contradictory positions on the vulnerability issue. On the one hand, the IRB would have waived the requirement for written consent on the grounds that writing down names would have endangered human subjects. On the other hand, the IRB would have approved a data-entry form that required L2 field teams to write down at least one name for every household surveyed. There are other possibilities. One is that the L2 researchers switched data entry forms, abandoning the IRB-approved one in the field and substituting a new one requiring the collection of names. Another possibility is that Lafta submitted a false data-entry form to the WHO. This last scenario gains plausibility from the fact that the NJ obtained a [second data-entry form](#) (and also [a questionnaire](#)) that may have been

the ones actually used by the L2 study in the field. Burnham, Roberts and Hopkins all decline to either confirm or deny whether any of these forms were actually used or whether entirely different forms, that have not yet surfaced, might be the true ones.

It is easy to get lost on the basics here. No legitimate survey researcher will refuse to show his questionnaire or data-entry form when asked. Yet we now have a preposterous situation with multiple forms floating around and neither the L2 authors nor a Hopkins official will stand up and say:

“Here is the questionnaire. Here is the data-entry form.”

What are they hiding?

b. Burnham told the NJ that he submitted an English-language consent form to the IRB but the IRB did not check the teams’ Arab and Kurdish translations of these consent forms. If so, then this appears to be in flagrant violation of [the informed consent regulations stated clearly on the IRB website of Hopkins](#). Burnham is quoted by NJ as saying that the IRB did not ask to see L2’s translated oral consent statements since such translations are only required when “risks to participants ... are substantially above what they regularly experience.” If so, then the IRB waived the normal requirement for written consent on the ground that collecting signatures would endanger the lives of the L2 respondents and then turned around and waived the requirement to examine translations of consent forms on the ground that these same respondents were not facing unusual risks.

Note that the mention of a Kurdish translation of the consent statement is surprising on its own. According to the L2 paper, field-team members were fluent in English and Arabic. There is no mention of Kurdish. The field teams might have shown a Kurdish translation to their Kurdish-speaking respondents but it is unclear how they could have answered questions that might have arisen about the risks of participating in the survey or even how the field teams could have conducted interviews with Kurdish speakers. [Apparently, few people under 30 years old in the Kurdish zone of Iraq speak even rudimentary Arabic.](#)

There is another issue here not mentioned by the NJ. The L2 researchers have still not disclosed any consent forms in any language. Researchers, such as Dr. Madelyn Hicks of the Institute of Psychiatry in London, have asked to see these forms and been turned down.

c. Burnham disclosed that the field procedures actually implemented by L2 differed in at least one major way from those that were approved by the IRB: the use of neighborhood children to spread news of the survey. Burnham says “That’s actually what happened; that wasn’t part of the study design.”

The [IRB document on the responsibilities of Principle Investigators](#) states directly a responsibility of:

“Submitting to IRB all changes in protocol(s) previously approved by IRB, and ensuring that changes in approved research are not initiated without prior IRB approval.”

If this use of local children, highly unusual to put it mildly, was not approved by the IRB then it was a significant departure from IRB regulations and must be viewed as a major violation of the ethical oversight process. Surely Hopkins cannot ignore this.

d. Burnham makes the startling disclosure that interviews were conducted on the *doorsteps* of respondents. (Until the NJ article, these doorstep interviews were just an internet rumor.) Moreover, Burnham suggests that the IRB actually approved doorstep interviewing. If so, then the IRB and Hopkins are party to an egregious breach of the privacy of L2’s respondents that would have inflicted elevated risk upon them. According to Burnham, L2 interviewers wore white coats and, hence, would have been highly visible conducting interviews on doorsteps. The fact that violence interviews were conducted would be common knowledge in every neighborhood visited by the L2 project. Local militias, in many cases the perpetrators of the violence that L2 was endeavoring to discover, would have been completely aware of these interviews. (I am preparing another document that elaborates on these risks that I can send upon request.) This breach of confidentiality alone could easily explain why Hopkins stonewalled the NJ on the IRB.

3. The NJ calls attention to possibility that the IRB may have exempted L2 from federal regulations under a rule allowing such exemptions for surveys that do not collect names. Yet, as noted above, L2 may have actually collected names and the IRB may have approved a data-collection form that mandated the collection of names.

4. Roberts seems to concede that it is likely that his study has incited hatred and violence:

“When *NJ* asked Roberts about the risk that his estimate would incite more violence, his confidence seemed to waver for the only time during the interview.” This area of study is a minefield,” he said. “The people you are talking about are the same kind of people who deny the Holocaust.” Does it give him qualms that some of those people use his study to recruit suicide bombers? “It does,” he replied after a pause. “My guess is that I’ve provided data that can be narrowly cited to incite hatred. On the other hand, I think it’s worse to have our leaders downplaying the level of violence.””

B. Data Integrity

1. NJ managed to squeeze out some information on the enigmatic Riyadh Lafta.

Finding articles written by Lafta during the Saddam Hussein era is a big coup. NJ has a short discussion of the article by Lafta and others entitled [“Risk Factor of Death Among <2y Children A Hospital Based Study”](#). Here is the final paragraph of the article:

“So we can conclude from results that the most important and wide spread underlying causes of the deterioration of child health standards in Iraq is the long term impact of the non-humanized economic sanction imposed through united nation resolution.”

No such conclusion is possible from the data presented. The study looks at the records of all children admitted to a hospital in Baghdad over a six-month period. Since there is no information on how such a sample would be generated there is no way to extrapolate the data collected to the national level as the authors do in their conclusion. Also, the data cannot be used to study the dynamics of child health, i.e., deterioration over time, since it is essentially a snapshot of a single point in time. If the data of the study are legitimate then about half of the children who died in the sample were malnourished. There is no way to know whether malnourishment was “the main underlying risk factor” in these deaths as claimed by the study or whether such malnourishment-associated deaths had become more frequent than they had been before economic sanctions. The paper also claims to have done a survey of the mothers of the children. It does not present any of the data from this survey directly but does draw conclusions about how socio-economic status, presumably as measured by this survey, relates to child health. The paper exhibits a lack of understanding of the difference between correlation and causation.

I had problems downloading the second Lafta document and have contacted the NJ about this.

An important contribution of the NJ article is that it does not just give the L2 researchers a free ride on Lafta’s non-presence. Of course, people very familiar with the L2 discussion know that Lafta had sole responsibility for the field work, that he is unavailable to answer any questions about this field work and that the US-based L2 people know little about the field work and have frequently changed their stories about this work. So this is not completely new. Still, most journalists have simply accepted that Lafta was under wraps and have acquiesced in only talking with Burnham and Les Roberts. Yet, as the NJ article makes clear, the whole L2 edifice rests squarely on Lafta’s invisible shoulders.

We learn from the NJ that Lafta worked in the Ministry of Health under Saddam Hussein on trying to convince the world that sanctions were killing lots of Iraqi children. Lafta’s article, discussed above, was part of this effort.

Another interesting Lafta angle is that Burnham and Roberts profess to know nothing about Lafta’s publications. Either they are embarrassed to admit to knowing about his work under Saddam Hussein or they recklessly gave him full control over their L2 field operations despite knowing little about him.

Richard Garfield is the person who connected the L2 team with Lafta, the one who knows Lafta the best and the one who seems to vouch for Lafta the most in the article:

"I've known him for years," Garfield told *NJ*. "I used to work with his boss in 2003, studying how Saddam had pilfered cash [intended] for the health care system. He's thoughtful, careful, and we became friends."

Yet Garfield says that he removed his own name from the L2 project because Lafta was left to work on his own without supervision:

"Thus, most of the oversight for *Lancet I* -- and all of it for *Lancet II* -- was done long-distance. For this reason, although he defends the methodology, Garfield took his name off *Lancet II*. "The study in 2006 suffered because Les was running for Congress and wasn't directly supervising the work as he had done in 2004," Garfield told *NJ*."

2. There is new material on the political motives behind the study and its publication. Some of the NJ contribution here is just pulling together already-known material into one spot but there is fresh material as well.

a. There is the [Horton video](#) which has been known for a long time but apparently not to everyone who wants to know about it, judging from what I have seen on the blogs.

b. NJ does a good job of documenting the political timing of the publication of both L1 and L2 and how Roberts (for L1) and Burnham (for L2) demanded pre-election publication as a condition for submission. Roberts had already admitted publicly that these two studies were meant to influence US elections. (This was in an AP story that can be found as the October 12, 2007 entry of [David Kane's blog](#)).

But Burnham had previously denied the obvious truth that L2 was politically timed:

<http://www.washingtonpost.com/wp-dyn/content/discussion/2006/10/18/DI2006101801279.html>

"Portland, Ore.: Thank you for taking my question, which is in two parts. First, could you briefly identify potential sources of sampling and nonsampling error in your study?

Second, it has been implied by sources associated with the administration that the timing of this study was politically timed. But didn't the editorial board of the *Lancet* decide when to run this study? How long did this study undergo peer review by the *Lancet* until the decision was taken to publish it?

Gilbert Burnham: There are a variety of potential sample pitfalls to be wary of. Since you seem to have some knowledge in this area, I would refer you to the discussion in the *Lancet* article for our discussion of these. We have tried to prevent these from occurring, and I think we were successful.

This was not politically timed. We wanted it out much earlier for exactly this

confusion risk. We started working on it in November 2005, and hoped to have it out in July or August, but many delays happened. The *Lancet* had a very detailed peer review, and then we worked closely with an editor to get the paper in its final form. This followed a standard track through the *Lancet* process--something over which no author has control--although we wish some times!"

Yet in the NJ article we see:

"Roberts conceded that he opposed the Iraq invasion from the outset, and -- in a much more troubling admission -- said that he had e-mailed the first study to *The Lancet* on September 30, 2004, "under the condition that it come out before the election." Burnham admitted that he set the same condition for *Lancet* II. "We wanted to get the survey out before the election, if at all possible," he said."

"...the *Lancet* editor who agreed to rush their study into print, with an expedited peer-review process..."

Evidently Burnham lied to the guy from Portland Oregon in an amiable and folksy style.

3. The funding from George Soros is clearly new. I think that hidden Soros funding is indeed a legitimate concern. Why has this funding been concealed for more than a year? Nevertheless, as Richard Garfield pointed out to the NJ, it is possible to have political motivations and still do solid work. For example, a scientist may be funded by the oil industry and still conduct good research on climate change.

In my opinion, the right response to the obvious political motivations of many of the key L2 researchers is to review the study with particular care. But this prescription takes us right back to the central problem of L2: its extreme lack of transparency. The mixture of "black-box data" with political motivation, often concealed, is a recipe for disaster. Analogously, it is normal to be wary of the motives of someone trying to sell you a used car. But only a fool would actually buy a used car from a seller who insists that the buyer may not look under the hood.

4. The story of how God drove Lafta to go to Falluja and from which he returned with unbelievable results is an old one. But NJ usefully presents this within the context of the implications for L2 of what Lafta has done in the past when left to his own devices in the field. One could say that L2 allowed Lafta do his Falluja thing on a national scale.

Another fact that has been out there but not really noticed is that L1 implied 15,000 people killed by US military vehicles in just a year and a half, i.e., about 30 per day for an extended period of time. This ludicrous finding constitutes fully 15% of the 100,000 excess deaths claimed in L1.

5. This oddly naïve quote from Richard Horton clarifies that the *Lancet* has jettisoned normal defense mechanisms against fraud and placed itself in a uniquely vulnerable position among journals:

“*Lancet* Editor **Richard Horton** shares this fundamental faith in scientists. He told *NJ* that scientists, including Lafta, can be trusted because "science is a global culture that operates by a set of norms and standards that are truly international, that do not vary by culture or religion. That's one of the beautiful aspects of science -- it unifies cultures, not divides them.””

Horton seems to say that if research is submitted to him a scientist, any scientist, Horton will rule out the possibility of fraud.

Horton goes on to argue, in effect, that he is running a business that would fail if he added fraud-prevention to the *Lancet's* peer reviewing practices:

"if for every paper we published we had to think, 'Is this fraud?' ... honestly, we would fold tomorrow."

The L2 paper is riddled with a large number of errors and false statements and ignores or does not properly address a large amount of contrary evidence. Indeed, L2 contradicted all existing evidence at the time of its publication. Yet, even in this situation, the *Lancet* and its peer reviewers never considered the possibility of fraud. Such practices make the *Lancet* into a sitting duck for fraud.

6. Fritz Scheuren, a former president of the American Statistical Association (ASA) and author of the ASA's book [What is a Survey?](#), cuts to the heart of the matter. L2 did not employ any of the normal defense mechanisms against fraud that are routinely used in surveys. I have a separate document that goes into detail on what these mechanisms are and their absence in L2 that I would be happy to supply upon request. I also recommend the ASA document [“Interviewer Falsification in Survey Research”](#)

One aspect of the abdication of any defense against fraud, stressed by Mathew Warshaw of D3 Systems, was the failure of L2 to collect demographic data to check the veracity of the data. That is just one defense mechanism but an important one. [Steven Moore made this point in a Wall-Street Journal Op Ed more than a year ago. Roberts responded by claiming that Moore had invented the obvious falsehood that L2 had not collected demographic data](#) (scroll to the bottom). The basis of Roberts' charge is the curious argument that Moore should have learned from reading *L1* that L2 had actually collected demographic data. The fact that Moore had not consulted a 2004 study to discern the methodology of a 2006 one revealed him to be in bad faith. It turns out that Moore was right.

7. The material on death certificates that I provided to NJ with David Kane is pretty much new to the public discussion although I did present it earlier at a seminar that is posted on my web site:

http://personal.rhul.ac.uk/uhte/014/gmu_final.pdf

This contains a lot more material on data integrity that did not make it into the NJ article. In my opinion, some of this material is more damaging than what was used by the NJ.

8. The NJ discussion of cluster 33 is not really new. The problem was discovered by Olivier Degomme and Deberati Guha-Sapir of CRED in Belgium some months ago. But these issues have not been properly appreciated and it is good that the NJ article draws attention to them.

9. The NJ quotes Roberts reiterating his old defense on the timing of the publication of L1 and applying it to L2 as well. He states plainly that the survey teams would have been murdered if they had been perceived (by whom?) as sitting on their results until after the US election:

“If this study was finished in September and not published until after the November elections -- and it was perceived that we were sitting on the results -- my Iraqi colleagues would have been killed.”

NJ responds with simple common sense. Perhaps this is true but, if so, this is a pretty big reason to distrust both the L1 and the L2 data. If Lafta and his field teams would have been killed over the timing of their publications then they might also been killed for publishing estimates perceived to be too low by the same unidentified and intimidating forces menacing the teams over their publication calendar. Perhaps such intimidation explains why Lafta insisted on going to Falluja, against the advice of Roberts and returned with such extraordinarily high numbers. Intimidation might also have been a driving force behind the whole L2 project.

10. This is not a ringing endorsement of the quality of the L2 data:

“Burnham also paused when asked whether Iraqi factions manipulated him and his colleagues and then replied, "We're reasonably confident that we were not manipulated.”

11. This is not a ringing endorsement of the quality of the L2 data from L1-co-author Richard Garfield, placing the L2 estimate too high by a factor of 4:

“Garfield told *National Journal* that he guesses that 250,000 Iraqis had died by late 2007. That total requires an underlying casualty rate only one-quarter of that offered by *Lancet II*.”

12. And this is not a ringing endorsement from Richard Horton:

“"Anything [the authors] can do to strengthen the credibility of the *Lancet* paper," Horton told *NJ*, "would be very welcome." If clear evidence of misconduct is presented to *The Lancet*, "we would be happy to go ask the authors and the institution for an official inquiry, and we would then abide by the conclusion of that inquiry.”

His solution of asking the authors themselves and the (at present) stonewalling Johns Hopkins to have a look at L2 is not a promising one. The *Lancet* has already been presented with evidence of misconduct and they have not reacted. Surely the *National Journal* article itself should be enough to trigger an investigation. Previous articles in [Science](#), [Science](#) (again), [Nature](#), [Slate](#) (particularly [this disturbing bit](#)), and [The Times](#) should already have been enough.

One final quote:

“Burnham says, for instance, that Lafta offered to take reporters to visit some of the neighborhoods used in the clusters, although he declined to say whether the reporters would be allowed to visit the surveyed households or to pick the clusters to see.”

Someone should volunteer to take this tour.