

# Conflict Deaths in Iraq: A Methodological Critique of the ORB Survey Estimate

Michael Spagat  
Royal Holloway, University of London

Josh Dougherty  
Iraq Body Count

In September of 2007 ORB, a British opinion polling firm, released an estimate that 1.2 million Iraqis had been killed in the conflict, subsequently lowering its estimate to 1 million. We compare three ORB polls and find important irregularities in ORB's mortality data in four central governorates of Iraq that account for more than 80% of the estimated deaths. These internal validity checks indicate that the ORB mortality data are not credible and would suggest a much lower estimate than ORB has published. We also analyze a number of specific error sources in the poll. Systematic errors, which include non-coverage and measurement errors, mostly point toward overestimation. Variable errors are also substantial but they are difficult to quantify in part due to incomplete disclosure of methodological details by ORB. External validity checks, including comparisons with two much larger and higher quality surveys, reinforce the conclusion that ORB has overestimated the number killed in Iraq by a wide margin. Thus, our paper answers a challenge facing the field of survey methodology, to explain how different surveys have produced such divergent mortality estimates for Iraq.

**Keywords:** Iraq, conflict mortality, survey quality, ORB poll, survey error

## 1 Introduction

In September of 2007 a British polling firm called Opinion Research Business (ORB) published an estimate of "1.2 million murders" in Iraq since the U.S. invasion of 2003 (ORB, 2007b). ORB subsequently revised this estimate down to just over 1 million in January of 2008 (ORB, 2008a). The polling data which formed the basis of this estimate were compiled by an Iraqi polling firm called IIACSS on behalf of ORB and represents the first ever attempt by ORB to use its polling to create a national mortality estimate, either for Iraq or for any other country.

All credible sources on conflict mortality since the 2003 invasion have shown a staggering level of human losses suffered by the people of Iraq, yet the ORB estimate is exceptionally high even within this universe of figures. ORB's estimate is comparable only to the Burnham et al. (2006) one published in the *Lancet* which was, in turn, much higher than any other estimate until the ORB estimate appeared almost a year later. ORB (2007b) leans on this *Lancet* estimate as support for its conclusions. Likewise, two authors of the *Lancet* study have regularly pointed to ORB as corroboration for their high 2006 estimate, with one going as far as to claim that ORB represents the "best estimate" of Iraqi deaths currently available (KUOW radio, 2008). Moreover, the Bloomberg School of Public Health of Johns Hopkins University, the parent organization behind the *Lancet* estimate, issued two separate official statements claiming the ORB poll

as important corroboration for the *Lancet* estimate.<sup>1</sup> However, the *Lancet* study itself has been widely questioned in articles such as Bohannon (2006), Dardagan et al. (2006), Daponte (2007), Guha-Sapir and Degomme (2007), Johnson et al. (2008), Laaksonen (2008), Munro and Canon (2008), Rosenblum and van der Laan (2009) and Spagat (2008 & 2010). Recently, the principal researcher of the *Lancet* study has been censured by the American Association for Public Opinion Research because he "repeatedly refused to make public essential facts about his research" and suspended by Johns Hopkins University from human subject research for protocol violations that compromised the safety of survey respondents (AAPOR, 2009 and Bloomberg School of Public Health, 2009).

However, aside from the authors of the *Lancet* estimate and Bloomberg officials, the ORB estimate has not been taken very seriously in the scientific community. For example, the Iraq Family Health Survey (IFHS) ignores the ORB estimate in its article on violent conflict mortality in Iraq that was published in the *New England Journal of Medicine* (IFHS, 2008a). Daponte (2007) also does not mention the ORB poll in its overview of conflict mortality estimates in Iraq. Tapp et al. (2008) mentions but does not discuss the ORB poll beyond repeating some information that appeared in ORB's first press release accompanying its poll (ORB, 2007b).<sup>2</sup> Thus, implicitly, the scientific literature seems to have dismissed the ORB poll without much discussion.

We argue, however, that for two main reasons the ORB

---

Contact information: Michael Spagat, Department of Economics, Royal Holloway, University of London, e-mail: M.Spagat@rhul.ac.uk

<sup>1</sup> One such Bloomberg statement is quoted in Munro and Canon (2008). Another is Bloomberg School of Public Health (2007). See also Burnham and Roberts (2007).

<sup>2</sup> The Tapp et al. (2008) article did not notice that ORB reduced its estimate by about 200,000 in January of 2008.

poll should not simply be dismissed a priori. First, ORB is a professional polling organization and, at least on the surface, appears to have applied some standard survey methods to arrive at its estimate. If ORB has overestimated violent deaths in the Iraq conflict by a wide margin, as we argue in this paper, then it is a challenge to the field of survey methodology to explain what went wrong. Conversely, if ORB's estimate is more or less accurate than two other very large and well-funded surveys of Iraq must have underestimated violent deaths by a wide margin (see section 6, table 2), in which case it would be a challenge to survey methodologists to explain what went wrong with these surveys. More generally, if survey methodologists are unable to separate bad conflict mortality surveys from good ones then the sample survey approach to determining conflict mortality must be considered unviable. Second, as noted above, the authors of the Lancet estimate and Bloomberg officials lean heavily on the ORB poll. Outside of academic circles, the ORB poll has been cited as important evidence in a range of public and media discussion, including Steele and Goldenberg (2008), National Public Radio (2007) and many blogs and websites such as Fair.org (McElwee, 2008). Thus, the ORB estimate has had an impact in much of the public discourse on mortality in the Iraq conflict and, therefore, should not simply be ignored.

In this paper we use the perspective of the survey quality literature (Biemer and Lyberg, 2003) to examine ORB's estimate for killings in the Iraq war. In section 2 we briefly examine the sample survey methodology for estimating war deaths and then present what has been disclosed of ORB's particular implementation of this methodology. In section 3 we examine sampling error, concluding that ORB's published "ranges" substantially underestimate the true sampling error of its estimate, but that ORB has not released enough information about its methodology to allow for a proper assessment of sampling error.

In section 4 we present an internal validity check of ORB's data that includes two additional ORB polls conducted just before and just after the one that we focus on in this paper. We find internal contradictions indicative of compromised data collection practices which greatly exaggerate the resulting estimate. In particular, in a contiguous region of four governorates accounting for more than 80% of ORB's estimate, a higher percentage of respondents report deaths of household members than the percentage of respondents reporting deaths of extended family members in an ORB poll conducted in February, 2007. ORB also conducted a third poll in March, 2008 in which it reports switching back to its extended-family question but with only a slight increase in the percent of respondents reporting deaths. Yet in southern Iraq ORB reports that the percent of households reporting deaths did, as expected given Iraq's large extended family networks, drop sharply when ORB switched from an extended family question to a household question. In light of these problems it is clear that ORB's data cannot support an estimate of 1 million deaths as claimed, but might support a very rough estimate of 300,000.

In sections 5 and 6 we identify further error sources, fo-

cusings mainly on systematic errors in the form of errors of measurement, non-coverage and non-response. These suggest that even a revised estimate of 300,000 is likely to be biased upward. In section 7 we compare ORB's estimate with other sources, including two larger and much higher quality mortality surveys. These external validity checks suggest that the ORB estimates greatly exceed those that derive from more credible methodologies and stand in contradiction to a wide range of evidence.

## 2 Methodology and Disclosure

In recent years there has been growing use of sample survey methodology to estimate deaths linked to violent conflicts. The most prominent examples of surveys that have produced national estimates have been Spiegel and Salama (2000) for Kosovo, a series of surveys done in the Democratic Republic of Congo (DRC) by the International Rescue Committee (IRC) (e.g., Coghlan et al., 2006) and five surveys in Iraq that we will discuss in the present paper.

The basic ideas of the survey approach are straightforward for survey professionals. A random sample of households is drawn, normally a cluster sample. Questions are asked aimed at discovering, among other things, the number of people living and dying in each household. There are questions about causes of death that distinguish, at a minimum, between violent deaths and non-violent deaths. Researchers estimate the number of violent deaths during the survey's recall period. In some cases, researchers also make estimates for what are known as "excess deaths". These are deaths above and beyond a baseline rate thought to be the one that would have prevailed if there had not been a war. In practice, the baseline mortality rate is generally defined as a pre-war rate and this is often measured within the survey itself by extending the recall period back to before the war began.

ORB has not disclosed many details about its methodology but it is clear that this methodology was roughly along the lines described above, although ORB does not give an excess death estimate. ORB's estimate for violent deaths in Iraq first appeared in a September 2007 press release published on the ORB website entitled "More than 1,000,000 Iraqis murdered":

"... this data suggests a total of 1,220,580 deaths since the invasion in 2003. Calculating the affect from the margin of error we believe that the range is a minimum of 733,158 to a maximum of 1,446,063." ORB, (2007b)

ORB's official description of its methodology is contained within six bullet points in its press release:

- "Results are based on face-to-face interviews amongst a nationally representative sample of 1,720 adults aged 18+ throughout Iraq (1,499 agreed to answer the question on household deaths)

- The standard margin of error on the sample who answered (1,499) is +2.5%
- The methodology uses multi-stage random probability sampling and covers fifteen of the eighteen governorates within Iraq. For security reasons Karbala and Al Anbar were not included. Irbil was excluded as the authorities refused our field team a permit.
- Interviews conducted August 12th-19th 2007.
- Full results and data tabulations are available at [www.opinion.co.uk/newsroom.aspx](http://www.opinion.co.uk/newsroom.aspx)
- ORB is a full member of the British Polling Council and abides by its rules.” ORB, (2007b)

The only other methodological information provided is an English language version of the questions. ORB declines to disclose the field versions of its questionnaire in Arabic and Kurdish, on the ground that such disclosures would lead to what ORB believes would be unproductive discussions of the quality of its translations.<sup>3</sup>

After its initial release ORB did further rural sampling<sup>4</sup> and in January of 2008 lowered its estimate by almost 200,000 deaths:

“... we now estimate that the death toll between March 2003 and August 2007 is likely to have been of the order of 1,033,000. If one takes into account the margin of error associated with survey data of this nature then the estimated range is between 946,000 and 1,120,000.” ORB (2008a)

ORB (2008a, press release)<sup>5</sup> provides the additional methodological information that there were “112 unique sampling points” and that in the end it conducted 2,163 successful interviews out of 2,414 attempts.

ORB does not disclose what its sampling frame is but this would appear to be a list of dwellings in Iraq since it is clear that the measurement units for the mortality question are intended to be households. Within each such unit a single respondent was chosen using the next birthday method. Thus, it is assumed that each potential respondent within a household would provide identical information about mortality so that the selection of one over another does not matter. It is also assumed that respondents provide information on deaths only for persons residing within their households although we argue in section 6 that this assumption is not well founded.

ORB (2008a, New Casualty Tabs) tabulates responses to questions and includes some information on the weighting scheme which should adjust for differing selection probabilities. From Table 1 of ORB (2008a, New Casualty Tabs) we infer that answers are weighted by governorate and by

whether respondents are classified as urban or rural dwellers. For example, 474 Baghdad interviews are scaled up to 681 while 190 Salahadin interviews are scaled down to 161. 590 rural interviews are scaled up to 756 and 1,824 urban interviews are scaled down to 1,658. ORB does not explain how it arrives at its weighting or how it classifies urban and rural areas.

ORB estimates the number of deaths as follows. It considers only the 2,163 individuals who responded to the mortality question. Among these, ORB finds that 20.2% report deaths with these respondents reporting 1.26 deaths on average, where these figures incorporate the weighting scheme. Taking these figures to be representative of the household unit, ORB derives a mortality estimate by applying them to an estimated number of households nationwide. ORB states that the 1997 census in Iraq “indicated a total of 4,050,597 households”, and therefore calculates  $0.202 \times 1.26 \times 4,050,597 = 1,030,958$ , with the tiny difference from ORB’s reported number presumably being due to rounding. ORB’s “estimated range” of 946,000-1,120,000 comes from subtracting and adding 1.7% from and to its 20.2% figure where the 1.7% is approximately the margin of error on a simple random sample of size 2,160 for a yes-no question to which approximately 20% of the responses are “yes”.

ORB’s parsimony with information about its methodology is an indicator of low survey quality and weakens confidence in its estimate. However, nondisclosure of information cannot, per se, explain estimation error. We now turn, therefore, to potential sources of error.

### 3 Sampling error

ORB does not provide enough information about its sampling methodology to allow a proper assessment of sampling error but, for a variety of reasons, it is clear that its range of plus or minus 8% is inappropriately narrow.

First, the fact that ORB (2008a) used 112 unique sampling points suggests that the poll was a cluster survey with 112 clusters. Any proper error calculation must account for the survey being a cluster survey, this being especially important since households located near to each other may have quite similar violence experiences. Yet ORB does not use cluster-survey sampling error methods and, instead, calculates its error margins as if it had drawn a simple random

<sup>3</sup> This refusal was made, and this reason given, in a meeting Michael Spagat had with two ORB representatives at the ORB office in London on February 28, 2008.

<sup>4</sup> This development raises some concerns about the quality of the sampling. ORB (2007b) states that they drew a “nationally representative” sample but then ORB (2007c) notes that this sample had “limited coverage in rural districts” so ORB, therefore, performed a “rural booster”. ORB (2008a, press release) then declared this sample to also be “nationally representative.” Precise details about the sample design would be welcome.

<sup>5</sup> ORB posts all its releases online. To contain the size of our bibliography we cite only the main pages for each poll and provide information when appropriate guiding readers to further documents accessible from these main pages.

sample. The impact of ignoring clustering effects in ORB's error calculation is likely to be large. Two other much larger cluster surveys of mortality in Iraq had much wider confidence intervals. The Iraq Living Conditions Survey (ILCS), with 2,200 clusters and 21,668 households, reported a confidence interval of -25% to +21% (ILCS, 2005a). The Iraq Family Health Survey (IFHS), with 979 clusters and 9,345 households, reported a confidence interval of -31% to +48%. In comparison, ORB's claimed plus or minus 8% does not seem credible.

Second, ORB accounts for sampling error only in its estimate of the percent of households experiencing deaths but not in the average number of deaths reported per household. Third, the ORB (2008a) sample is weighted by governorate and also has rural-urban weightings. However, the error calculation does not account for the weighting scheme.

#### 4 Internal Validity Check

ORB conducted three polls within a year of each other in Iraq which asked questions on mortality. These three polls are very hard to reconcile with each other and, as a group, undermine the ORB mortality estimate derived from the second of these polls.

1. There was first an ORB poll (ORB, 2007a) conducted in February of 2007 (hereafter ORB1). In ORB1 26% of respondents reported "the murder of a *member of my family/relative*" [emphasis added] in the three previous years (table 3 of ORB1).
2. Six months later, in August of 2007, there was the ORB poll that is the main object of this paper, and which was used by ORB to create its mortality estimate (hereafter ORB2). In ORB2 20% of respondents reported deaths "as a result" of conflict/violence of a *household* member where household membership was defined by living under the same roof (table 1 of ORB2).
3. Lastly, there was a third ORB poll (ORB, 2008b) conducted in February/March of 2008 (hereafter ORB3). In ORB3 24% of respondents reported "the murder of a member of my family/relative" since the invasion (table 20 of ORB3).<sup>6</sup>

ORB itself, and virtually all commentary on ORB2, has interpreted its estimate to mean direct violent deaths, i.e., the same type of deaths measured for extended families in ORB1 and ORB3. However, it is further claimed that ORB2 measured these deaths only for formal household members rather than for any extended-family members as in ORB1 and ORB3. The claim that ORB2 was limited to formal household members is the main basis for ORBs extrapolation of its figures to a national mortality estimate.

The number of respondents with an extended family member who has been killed has to be far higher than the number of respondents with a household member killed as long as all three polls are indeed measuring the same kinds

of deaths and respondents are following the "family/relative" or "household" instructions of each poll correctly. Yet ORB gives three percentages in separate polls conducted within one year of each other that are very close to one another: one for respondents reporting deaths of household members, sandwiched in between two others reporting deaths of extended family members.<sup>7</sup>

To quantify the discrepancy between the three polls we use the network scale-up equation given in Moody (2006), which estimates the proportion of a population connected to someone with direct experience of a particular event type as a function of the total population size, the number of people with direct experience of the event type and the average number of connections per person.<sup>8</sup> We first assume that the population of Iraq is 29 million, that average extended family size in Iraq is only 26 and that 250,000 people were killed in Iraq during the ORB2 coverage period, i.e., that the ORB2 estimate is too high by a factor of 4. These figures imply that about 20.2% of the population would have had a family member or relative killed. Thus, we can generate percentages much like the ORB1 and ORB3 ones based on very small extended family sizes for a Middle Eastern country (White and Houseman, 2002) and far fewer deaths than estimated by ORB2. If we raise the extended family size to 49, still a low number, then we can lower the number of deaths to 150,000 and still 22.4% of the population would have experienced the killing of an extended family member. If we use the ORB2 estimate of 1 million people killed and an extended family size of 49 then more than 82.1% of the population would have experienced the killing of an extended family member, far more than the percentages measured in ORB1 or ORB3. Thus, the ORB2 estimate of 1 million people killed is difficult to reconcile with the ORB1 and ORB3 polls.

Taken together, the three ORB polls present quite a puzzle; why is it that ORB finds a pattern of percentages of 26%-20%-24% when the middle number should be much lower than the first and third numbers? This pattern makes little sense if ORB2 is truly limited to formal household members while ORB1 and ORB3 include extended family members.

We argue that the ORB data themselves provide the answer to this puzzle.

<sup>6</sup> The number of interviews attempted for the three polls was 5,019, 2,414 and 4,000 for ORB1, ORB2 and ORB3 respectively. Thus, the ORB2 poll, the basis for ORB's national estimate of deaths due to the war, is the smallest of the three.

<sup>7</sup> The confidence intervals for ORB2's 20% and ORB3's 24% overlap even if we use the dubiously narrow margins of error claimed by ORB in its methodological bullet points. It is clear that on any proper calculation, all three confidence intervals overlap substantially with one another.

<sup>8</sup> The formula, is  $p_r = 1 - (1 - \frac{e}{t})^c$  where  $p_r$  is the proportion connected to those with direct experience,  $e$  is the number of events,  $t$  is the total population and  $c$  is the average number of connections per person. This formula can be derived analytically assuming independence in the assignment of events to people and Moody (2006) reports on simulations showing that the independence assumption can be relaxed.

Table 1 compares results of the three polls, conducted between March of 2007 and March of 2008, reported by governorate.<sup>9</sup> The figures clearly show one pattern in what we refer to as the geographic ‘Center’ and another very different pattern in the ‘South’. Note that these regions are geographically distinct from each other but internally contiguous (see map<sup>10</sup>).

First, in the South the percent of respondents reporting deaths of household members in ORB2 is consistently and substantially lower than the percent of respondents reporting deaths of family members/relatives in ORB1 [(7, 7, 4, 15, 2, 8) versus (27, 52, 23, 43, 46, 29)]. This large drop, by a factor of five on average (to 7 from 35), is reasonable and expected since the ORB1 question allows the inclusion of any extended family members, and extended family networks are much larger than households. A network scale-up calculation (Moody, 2006) also generates similar factors using plausible parameter values.<sup>11</sup> This pattern appears not only at the regional level, but also in every single governorate in the region, i.e., the pattern is robust.

Second, in the Center, these four contiguous governorates (see map), which account for more than 80% of the deaths in ORB2, show a pattern that is completely different and incompatible with the pattern found in the South. The percent of respondents reporting deaths of household members in ORB2 rises higher than the percent that had reported deaths of extended-family members in the same region in ORB1 [(43, 40, 18, 27) versus (31, 35, 18, 14)]. It is not just that the percentages fail to display the substantial and predictable drops that are seen throughout the entire South; they actually move in the opposite direction. Again, this is true not just at the regional level but also in every governorate in the region. The household percentages in these governorates in ORB2 are never below the extended-family percentages in ORB1.

Third, the ORB3 poll adds in a further problem in the South. The ORB3 extended-family numbers in this region are uniformly far below the ORB1 extended-family numbers despite the fact that another rather violent year had elapsed between the two polls [(2, 12, 1, 8, 9, 8) versus (27, 53, 23, 43, 46, 29)]. Moreover, the ORB3 extended-family numbers are strikingly similar to the ORB2 household numbers in the South [(2, 12, 1, 8, 9, 8) versus (7, 7, 4, 15, 2, 8)] despite the fact that the former should have been much higher than the latter.

The empirical patterns in the data displayed in table 1 are not credible. It appears that the boundaries of the declared units of analysis, the household in ORB2 and the extended family in ORB3, were not properly or uniformly enforced in either poll. Given this, the ORB mortality data cannot be considered valid, whatever the precise explanation for these anomalies.

One explanation that might resolve the strange pattern across the three polls in the four central governorates is that the field workers in this region did not change to the new household question for the ORB2 poll, but instead asked a family member/relative question like those asked in ORB1 and ORB3. The introduction of a different mortality ques-

tion into the ORB2 questionnaire than was used in ORB1 could have generated confusion or miscommunication in the field between different field teams working in different regions. Again, note that these regions are geographically contiguous. Thus, it seems logical that one field team or teams might cover the contiguous ‘Center’ region (Baghdad up to Kurdistan), while a separate team or teams might handle the contiguous ‘South’ (everything below Baghdad).

For the South, a similar explanation could resolve the pattern evident in the ORB3 poll. Namely, the field workers in this region correctly asked the household question in the ORB2 poll, but then incorrectly asked a household question again in ORB3. This would explain why the ORB2 and ORB3 numbers are so similar in the South, and why the ORB3 numbers are so much lower than the ORB1 numbers in the South. More broadly, this would explain why ORB3’s overall figure of 24% is lower than the 26% figure from ORB1, despite the fact that it would have been expected to be higher given the additional year of violence covered by March 2008.<sup>12</sup>

A second possible explanation for the strange pattern in the Center is that some people involved in the ORB2 field work purposefully manipulated the survey in this region with the goal of producing a high estimate. This is always a risk with politically charged research questions, and the degree to which this question would be both politically and emotionally charged in the case of Iraq in mid-2007 may have few parallels. Some known facts are consistent with this manipulation scenario. The release of the ORB2 poll was timed to coincide with the September, 2007 Congressional testimony of General David Petraeus on the impact of the U.S. “surge” policy, suggesting political motivation. Moreover, Munqeth Dagher, the director of IIACSS, the firm which implemented the field work for these polls has stated that he became a pollster in 2003 as a non-violent means of opposing the US invasion of Iraq (Research Talk, 2006, Munro, 2007) so for him political motivation seems fairly explicit. However, political biases on the part of researchers do not necessarily translate into biased data. The strange pattern across the three polls in the South displayed in table 1, where ORB3 is inappropriately low, would yield no similar political advantage and

<sup>9</sup> For completeness at the bottom of table 1 we give figures for the five governorates that were not polled all three times. The incomplete patterns in these governorates are consistent with the patterns for the governorates in the respective regions for which data are complete.

<sup>10</sup> We thank Gabriela Guerrero Serdán for providing us with this picture.

<sup>11</sup> For example, if 280,000 households experienced the murder of a member, the average extended family size is 45 and the population is 29 million then about 7% of households would have experienced deaths whereas about 35% of the population would have experienced the death of an extended family member.

<sup>12</sup> Looking only at the overall figures one might imagine that sampling error might explain why the ORB3 figure is below the ORB1 figure. However, the numbers in the South in table 1 virtually eliminate the possibility that sampling error is the right explanation for this anomaly.

Table 1: Three ORB Polls by Governorate

Region	% with death of family member/relative March, 2007 (ORB1)	% with death of household member August, 2007 (ORB2)	% with death of household member/relative March, 2008 (ORB3)
<b>Center</b>	<b>26</b>	<b>36</b>	<b>41</b>
Baghdad	31	43	45
Diyala	35	40	42
Salahudin	18	18	22
Ninewa	14	27	39
<b>South</b>	<b>35</b>	<b>7</b>	<b>7</b>
Babyl	27	7	2
Najaf	53	7	12
Muthana	23	4	1
Dhi Qar	43	15	8
Maysan	46	2	9
Basrah	29	8	8
<b>North</b>	<b>8</b>	<b>7</b>	<b>6</b>
Kirkuk/Tameem	(Kirkuk) 20	(Kirkuk) 17	(Tameem) 9
Sulaymania	2	2	6
Dohuk	0	3	1
<b>Center Ramadi/Al-Anbar (Ramadi)</b>	<b>29</b>	<b>NA</b>	<b>49</b>
<b>South Karbala</b>	<b>30</b>	<b>NA</b>	<b>13</b>
Wassit	NA	12	8
Qaddisiya	NA	11	16
<b>North Irbil</b>	<b>NA</b>	<b>NA</b>	<b>7</b>

Sources: All figures for governorates in this table come directly from three tables posted on the ORB web site. They are ORB (2007a, TablesFINALMarch07Irr) for ORB1, ORB (2008a, New Casualty Tabs) for ORB2 and ORB (2008b, FINALTABLESMarch08) for ORB3. For regional percentages we take averages across the governorates of each region using the population weightings for the governorates that are implicit in ORB's posted tables.

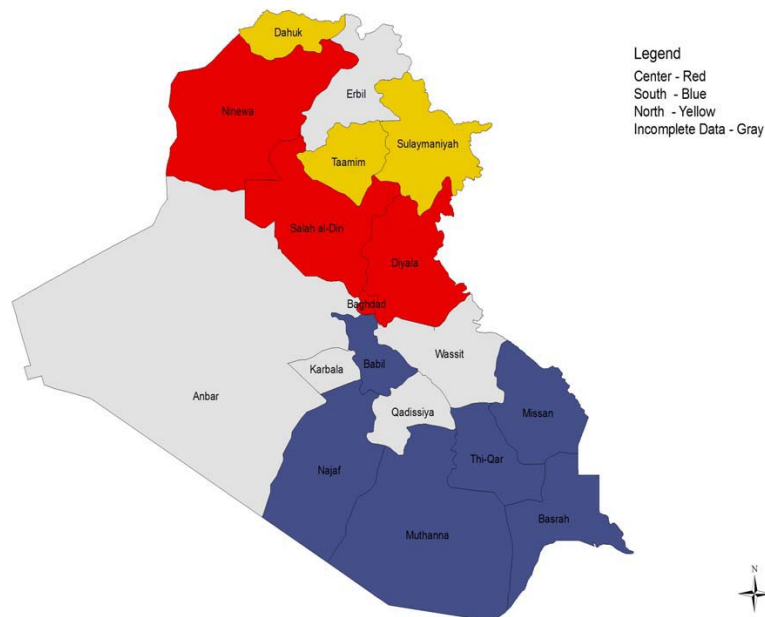


Figure 1. Governorates Linked by Reporting Pattern in Table 1 are also linked Geographically

seems more consistent with the mistaken field work scenario.

Whatever the precise explanation for these inconsistencies, the data from the three polls suggest that there were significant problems in the implementation of the mortality questions in the Center and the South during the ORB2 and ORB3 polls respectively. Specifically, the responses to ORB2's household question in the Center are much too high compared to the responses to the family/relative questions of ORB1 and ORB3, and conflict sharply with the more reasonable pattern seen in the South between ORB1 and ORB2. In addition, ORB3's family/relative figures in the South are much too low compared to ORB2's household figures.

Given the very high numbers of deaths provided to the ORB2 poll by the four central governorates, the implications of having inflated data in this region are large. The consistent and sensible pattern seen across the South between ORB1 and ORB2 suggests that the percentages in the Center should have dropped by a roughly similar factor between these two polls. Making this adjustment, while otherwise following the same estimation approach as ORB, the estimate would come to roughly 300,000, rather than 1,000,000 as reported.<sup>13</sup>

## 5 Measurement Error

In this section we argue that ORB's questionnaire is inadequate to fully exclude the reporting of non-violent deaths, focusing on the systematic error resulting from this problem although this will probably cause some variable error as well.

The original ORB (2007b) press release presents the survey as measuring the number of Iraqis who have been "murdered" (1.2 million). A few months later, a follow-up release revised the estimate down to 1 million (ORB, 2008a), but in this case the survey is described as measuring "Iraqi citizens [who] have died as a result of the conflict". The first formulation could only include direct, violent killings whereas the second is vaguer and could potentially include a wide range of non-violent deaths that respondents view as indirectly linked to the war. This inconsistency between press releases already hints at fuzziness on the boundary line between violent and non-violent deaths in the poll.

The ORB2 mortality question posted on the ORB website asks respondents about deaths "resulting from the conflict". Rather than defining such deaths clearly or checking these designations, ORB delegates to respondents the authority to decide whether or not a death is "a result of the conflict" and therefore appropriate to include:

"How many members of your household, if any, have died as a result of the conflict in Iraq since 2003 (ie as a result of violence rather than a natural death such as old age). Please note that I mean those who were actually living under your roof?"<sup>14</sup> Question 1 of the ORB2 poll.

While the parenthetical follow-up in the question does attempt to link deaths "as a result of the conflict" with violence, and may have helped to reduce the reporting of non-violent deaths, it is not clear how each respondent would interpret this part of the question. The parenthetical could

easily be interpreted as giving non-exclusive examples of a kind of death they want reported and a kind of death they do not want reported, leaving plenty of gray area in between. If ORB intended to measure violent deaths exclusively then it would have been better to just ask respondents specifically and directly about violent deaths, rather than basing its question around the often broader concept of conflict-related deaths and then attempting to correct this ambiguity by adding on another, still ambiguous, parenthetical remark.

The wording of ORB's question 1 could encourage some respondents to report deaths they perceive to have been indirectly caused by the conflict rather than direct violent deaths as normally understood in conflict research. Examples of non-violent yet conflict-related deaths are varied. Such deaths may include those directly caused by diseases but carried through impure, conflict-affected water supplies, or deaths suffered in hospitals experiencing conflict-related shortages of medical personnel, equipment or electricity. Ongoing violence in an area can also prevent access to treatment for conditions such as complicated births or heart attacks. Such deaths are easily interpreted as proper to include in an answer to ORB2's question, although doing so would convert the ORB estimate into an estimate of "excess deaths", or deaths that would not have occurred without the conflict, according to the subjective judgment of each respondent, rather than an estimate of direct violent deaths only.<sup>15</sup> A loose interpretation also becomes more likely when the closest these conflict/violence-related deaths come to being defined in the ORB question is that they are placed into a dichotomy against "natural deaths such as old age". Many respondents will feel, sometimes strongly, that the kinds of deaths described above are very much not "natural deaths such as old age", but are instead unnecessary (un-natural) deaths brought on by the environment of conflict and violence, and are appropriate to

<sup>13</sup> The following is a rough calculation which illustrates that the stakes are very high, hundreds of thousands of deaths, on the anomalies of Table 1. More than 80% of ORB's estimate of more than 1 million deaths come from the Center, i.e., more than 800,000 deaths. In the South, the percent reporting household deaths in ORB2 is 1/5 of the percent reporting family member/relative deaths in ORB1. If the ORB2 percent in the Center had also dropped by a factor of 5 relative to the ORB1 percentage then the ORB2 percentage in the Center would have been 5.2 rather than 36. Since  $800,000 \times (1 - (5.2/36)) = 680,000$  we conclude that ORB's estimate would be reduced by more than 680,000 deaths if the percentage difference in the Center for ORB2 compared to ORB1 had been similar to the more credible difference measured in the South.

<sup>14</sup> We have copied ORB's question exactly, including the errors such as placing a full stop rather than a question mark at the end of the first sentence, and a question mark rather than a full stop at the end of the second sentence.

<sup>15</sup> To the extent that ORB's data do capture non-violent deaths believed by respondents to be conflict related, and therefore measures "excess deaths" rather than just violent deaths, this would create an additional estimation problem in that the ORB poll does not provide any way to measure *negative* excess deaths. Including positive excess deaths but providing no counterbalancing measurement of negative excess deaths in regions where death rates may have declined would be a source of upward bias.

include. One related example evident in the ORB2 data is that 6% of respondents reporting deaths give “accident” as the cause of death, but deaths from accidents are not normally classified as direct violent deaths in conflict mortality surveys.

Moreover, some respondents may perceive certain non-violent deaths as results of conflict/violence when they are not. There is often no clear way to judge whether a particular non-violent death would have been avoided if a conflict had never started in the first place. Consequently, all sorts of deaths might be seen by respondents as conflict-related even when many are not really conflict-related.

Finally, ORB assumes that all respondents followed its restriction on allowable types of deaths to report: “(ie as a result of violence rather than a natural death such as old age)”. Yet, ORB2’s parenthetical clarification on allowable types of injuries is very similar to this one, “(ie as a result of violence rather than accidental injury)”, and 10% of respondents reporting injuries nevertheless gave “accident” as the cause of injury. If many respondents did not actually exclude accidental injuries when instructed to do so, then it is almost certainly wrong for ORB to assume that all of these same respondents did exclude natural deaths such as old age when instructed to do so with virtually identical wording.

Therefore, it seems inevitable that the ORB2 poll would have captured a number of non-violent deaths, including some that are likely to have been conflict-related and some that are less likely to have been conflict-related. Thus, the ORB2 mortality estimate would overstate the true number of direct violent deaths in Iraq because it can not really claim to measure exclusively violent deaths.

### Non-coverage and non-response errors

The ORB2 poll has a number of non-coverage errors and non-response errors. Again we focus on systematic error although each factor will probably add some variable error to the survey. We try to quantify upward and downward biases to the extent possible.

First, the attempt to define household boundaries in this poll, essentially with an afterthought attached to the same question, is inadequate to fully exclude the reporting of deaths of extended family members. How different respondents might interpret “actually living under your roof” is not known. The question does not pin down when, how long, or in what context either the respondent or the deceased had to live under this roof to be included in that household. With the high level of displacement in Iraq since the invasion, with individuals or whole families often migrating and moving in with other families or relatives for various lengths of time, or moving in and out of various dwellings over time, the composition of many Iraqi households will have been in a state of flux across the reporting period. A respondent could interpret this part of the question in a variety of ways depending on his or her experiences. For example, if some respondents lived in more than one dwelling at different times over the course of the long recall period used in this question, they

could consider all of the other inhabitants of each dwelling to qualify for inclusion, whereby their answers could reflect deaths across two or more households. Moreover, the above analysis still assumes, as ORB does, that each respondent really did attempt to rigorously restrict reports of deaths to only people who lived under the same roof that the respondent did. Yet, as pointed out in section 4, 10% of respondents reporting injuries ignored the instruction not to report accidental injuries. This suggests that ORB is also wrong to assume that 100% of respondents followed the instruction on household boundaries.

Even if the exact wordings of ORB2’s questions were clear on household boundaries, it is still possible that respondents nevertheless reported deaths of extended-family members. Few respondents to a household poll will understand the statistical imperative to set clear limits on household boundaries, or place such concerns above any other concerns of their own. To the contrary, many respondents might feel that, since interviewers are evidently trying to discover violent or conflict-related deaths, it is right and proper for respondents to report deaths of, e.g., cousins or uncles, even though they were not formally household members. It is likely that respondents would have a natural tendency toward what might be called “household inflation” so that they can bear witness to deaths of relatives who were near and dear to them. Taking household rosters, with demographic information on each household, is one defense mechanism against household inflation but ORB did not implement this check.<sup>16</sup>

Second, the treatment of non-response to the ORB2 mortality question is unsound and its assumptions unacknowledged. In making its estimate, ORB ignores the 251 (weighted) responses out of 2,414 attempted interviews in which the respondent either did not answer the main mortality question or responded “don’t know”. ORB then extrapolated the remaining 2,163 responses as if they were the complete sample. The effect of this treatment of non-response is to introduce an unstated assumption that these non-response households had the same proportion of conflict deaths as the average of the other households which did respond. Effectively, it is assumed that there were also conflict deaths in 20% of these households that did not actually report any deaths. This baseless assumption has a large impact on the ORB2 estimate, possibly generating upward bias of as large as 10% (i.e., about two percentage points).

Third, ORB (2007b) notes that, for security reasons, it was unable to sample in the governorates of Al Anbar and Karbala, implying that this is a source of downward bias. Al Anbar in particular has been one of the most violent governorates in Iraq so it is plausible that these security problems did cause some downward bias in the ORB2 estimate. On the other hand, for reasons related to Kurdish sovereignty, ORB2 was also unable to sample in the peaceful governorate of Irbil. ORB includes the populations of these three un-

<sup>16</sup> On the importance of household rosters see (SMART, 2006, p. 75) and (London School of Hygiene and Tropical Medicine (undated) p. 109).



sampled governorates in its national population estimate and extrapolates a national estimate out of data collected from only 15 out of Iraq's 18 governorates.<sup>17</sup> This estimation procedure effectively assumes that these three un-sampled governorates have the same death rate as the average death rate across the 15 sampled governorates. According to the Iraq Body Count database (IBC), these three governorates contain roughly 8.6% of the violent deaths of civilians during the ORB2 coverage period while, according to the IFHS, they contain about 12.3% of the population. Thus, it is not at all clear that these coverage problems would have been a source of downward bias as suggested by ORB, and in fact could have been another source of upward bias.

Fourth, ORB (2007c) also states that they were unable to sample in "more volatile areas", which suggests that even in governorates where the field teams did operate they failed to conduct interviews at some selected "sampling points" due to security problems. If so, these failures could be a source of downward bias, although they would also point to further problems with the way this poll has been presented to the public. ORB should have at least give a report on these problems.

Fifth, survivor bias, i.e., the general problem in conflict surveys that one cannot interview members of households which were wiped out completely, would be a source of downward bias in ORB2. However, this is thought to be a fairly small source of bias in most conflict surveys (Gakidou and King, 2006).

We can find no indication that ORB did anything to account for the non-coverage and non-response errors discussed in this section.

In sections 5 and 6 we focused on systematic errors but it is clear that the issues we discuss will contribute to variable errors as well, although these would be difficult to quantify. Moreover, in section 3 we argued that, properly calculated, the sampling error in the ORB2 survey should be much larger than claimed by ORB. Together these considerations suggest that a reasonable confidence interval for an estimate based on ORB's data could be very wide indeed.

## 7 External Validity Checks

In this section we place the ORB2 estimate within the context of other evidence on violent mortality during the Iraq conflict. This analysis further reinforces the conclusion that ORB has overestimated violent deaths in Iraq by a wide margin.

Table 2 compares ORB2 with four other surveys: the ILCS, the IFHS, Roberts et al. (2004), hereafter designated as "L1", and Burnham et al. (2006), to be designated "L2".<sup>18</sup> The table also builds in the data of IBC (continuously updated) which tallies violent civilian deaths based on media reports and data from hospitals, morgues and NGOs. Since IBC uses a uniform methodology and includes the coverage periods of all the surveys, it provides a useful means for comparing surveys with different coverage periods.

The ORB national mortality estimate is much higher than those of the ILCS, IFHS and L1. While the latter 3

surveys exceed IBC figures for their respective coverage periods by factors between 1.7 and 3.1, ORB and L2 both exceed IBC figures for their respective time periods by a factor of 12.2. Although the national estimates of ORB and L2 appear to be well in line with each other, they diverge substantially on the geographical distribution of violent deaths, with ORB placing a much higher fraction of its deaths in Baghdad than L2 does. The IFHS and ILCS are much larger and higher quality surveys than ORB and L2. For example, the IFHS and ILCS took full household rosters for each of their respondents. This procedure is considered a crucial step in household conflict mortality surveys to establish clear household boundaries and to reduce inaccurate or fabricated reports of deaths (footnote 16). L2 and the ORB poll did not take household rosters, which is perhaps instructive, given that these two surveys produced far higher estimates than those which did follow this procedure. Another quality indicator is that the Arabic and Kurdish questionnaires for the ILCS and IFHS are freely available online whereas neither ORB nor the L2 researchers have disclosed the field versions of their questionnaires. ILCS (2005b) and IFHS (2008) and Iraq Family Health Survey Study Group (2008) both give detailed accounts of how they designed and field tested their questionnaires, developed their sample frames, drew their samples, trained their staff, conducted and supervised their fieldwork and coded and quality-checked their data. ORB provides virtually no information on any of these issues.<sup>19</sup>

To close this section we present three further anomalies generated by the ORB2 data.<sup>20</sup> First, the ORB2 estimate implies roughly 220,000 deaths in car bombings, about 130,000 of which are in Baghdad.<sup>21</sup> In contrast, the IBC database has roughly 11,500 car-bomb deaths during the ORB coverage period, roughly 5,500 of which are in Baghdad. Thus, the ORB2 estimate implies that the international media have noticed only about 5% of car-bomb deaths both nationwide and

<sup>17</sup> It is inappropriate in the first place for ORB to simply extrapolate its data gathered in just 15 governorates to all 18 governorates. Accepted practice would have been to publish an estimate covering only the governorates actually sampled.

<sup>18</sup> The shorthand "L1" is in reference to this being the first war mortality survey of Iraq published in the *Lancet* whereas "L2" was the second such survey.

<sup>19</sup> Here we give just two other indicators of poor quality control in the ORB2 poll. ORB states that its interviews were "face-to-face" yet it did not establish the genders of 4% of its respondents. The first table of results released by ORB implied that the population of Baghdad is 60% Christian but then, without explanation, ORB replaced this table with another one that reduced this percentage to 3% (McLean, 2007).

<sup>20</sup> Blogger Will McLean has posted (McLean, 2007/2008) a number of good critiques of this kind on the ORB poll.

<sup>21</sup> ORB (2008a, New Casualty Tabs, Table 2) gives a breakdown of causes of death. Nationally, 22.3% are attributed to car bombs which would work out to 220,000 deaths if this percentage is applied to 1,033,000. In Baghdad 21.5% are attributed to car bombs which would work out to 130,000 deaths if this percentage is applied to 600,000 (notes to Table 2 above). Alternate calculations could produce slightly different figures but these would not affect our analysis.

Table 2: ORB and Surveys of Violent Deaths in Iraq since March 2003

	ILCS	L1	IFHS*	L2	ORB
Coverage Period Ends	May 1, 2004	September 20, 2004	June 30, 2006	July 10, 2006	August 31, 2007
Survey Estimates of Violent Deaths	26,000	56,700	98,000 or 151,000	601,000	1,033,000
Ratio to IBC	1.7	3.0	2.0 or 3.1	12.2	12.2
Violent Deaths in Baghdad	8,063	18,900 - 27,000	52,920 or 81,540	150,000	600,000
Ratio to IBC in Baghdad	1.0	1.9 - 2.7	1.9 or 2.9	5.2	12.5
Number of clusters	2,200	32	971	47	112
Field Questionnaire Available	Yes	No	Yes	No	No
Household Roster Taken	Yes	Yes	Yes	No	No

Sources: Pedersen (2007) for the ILCS; EPIC (2004) discloses that 57,600 of the L1 estimate are violent and Deltoid (2004) that between 7 and 10 of the 21 violent deaths in the L1 sample are in Baghdad; Iraq Family Health Survey Study Group (2008) for the IFHS; Spagat (2008) for L2. The calculation for the Baghdad ORB figures is 0.61 deaths per household in Baghdad times 988,346 households in Baghdad (4,050,597 households in Iraq of which about 24.4% are in Baghdad according to IFHS (2008).

\*The IFHS argued that there is a general tendency for the underreporting of deaths in household surveys. On these grounds, the IFHS adjusted its estimate upwards by more than 50% compared to a conventional estimate as produced by all the other surveys in this table. It is, therefore, best to remove this adjustment when comparing across surveys so in the IFHS column we always provide two estimates; the first is a conventional estimate and the second is an adjusted one as given in the IFHS.

inside Baghdad.<sup>22</sup> Yet, the IBC database shows that, on average, a lethal car-bombing generates six independent media reports. The notion that 95% of all car-bomb deaths go unreported is exceptionally implausible. Indeed, car-bombings are highly visible and newsworthy events that generate loud explosions that are heard over wide areas followed by noisy and obvious emergency responses. Many Iraqis carry cell phones that can be used to report car bombings. The perpetrators of car-bombings have a strong interest in making sure that these events are noticed by the international media, an interest shared by the Iraqi and US governments for different reasons. Moreover, the violent-death figures of the Iraqi Ministry of Health and the Baghdad Morgue use different counting systems and are very similar to the IBC figures both inside and outside of Baghdad, so ORB's figures would imply that hospitals and morgues have also failed to detect the vast majority of car bomb deaths.<sup>23</sup> In fact, if the ORB figures are right then it would appear that the number of car-bomb deaths alone is already higher than the total number of violent deaths of all causes recorded officially, even in Baghdad.

Second, the ORB2 findings on injuries seem to be well out of line with existing evidence for two main reasons. First, if injuries are estimated the same way that ORB estimated deaths, the injury data suggest that there should have been roughly 1 million injuries between the beginning of the war and the end of August, 2007, a number that seems to be too high by a wide margin.<sup>24</sup> Civic Worldwide (2003) identified 4,000 injuries during the first 50 days of the war. IBC (2003) found "at least 20,000" civilian injuries as of August, 2003. IBC (2005) found "at least 42,500" civilian injuries between the beginning of the war and March 19, 2005. The Iraqi Ministry of Health (MoH) recorded 53,634 injuries due to "terrorist incidents and military clashes" outside of Kurdistan between April 5, 2004 and June 1, 2006 [Roug and Smith, 2006].<sup>25</sup> The UN Assistance Mission for Iraq (UNAMI) reported 36,685 injuries in 2006 (UNAMI [2007]). The MoH reported 38,609 injuries in 2007 (NINA [2008]) It is difficult to extract a single clear figure for the whole ORB2 coverage

period from these overlapping data, but it is clear that the number of injuries consistent with these sources during that period would be much closer to 200,000 than to 1 million. Thus, the ORB2 data would suggest that 800,000 or more injured Iraqis were either never treated for these injuries or have never been recorded as receiving any treatment.<sup>26</sup>

The other problem with ORB's injury findings is that the ratio of injuries to killings, as measured by ORB is implausibly low, roughly just 1 to 1, and inconsistent with other sources on this question as well. Coupland and Meddings (1999) report ratios of wounded to killed for 9 conflicts ranging from 1.9 to 13.0 with a median of 4.0. Civic Worldwide (2003) recorded a ratio of 2 injuries per death during the first 50 days of the war. IBC (2003) and IBC (2005) found ratios of 2.5 to 1 and 1.7 to 1 respectively. The MoH recorded a ratio of 2.8 to 1 between April 5, 2004 and June 1, 2006. UNAMI (2007), with a ratio of 1.1 to 1, is the only source with a ratio nearly as low as ORB's but it applies only to 2006 when a large portion of killings were by execution, leaving few injured behind. For 2007 the MoH reported a ratio of 2.3

<sup>22</sup> IBC continuously monitors dozens of international media sources.

<sup>23</sup> Roug and Smith (2006) obtained data from the Iraqi Ministry of Health and Baghdad Morgue and concluded that "at least 50,000 Iraqis had died violently since the 2003 U.S. led invasion". Their figures closely match the IBC ones both at the national and governorate levels. Gamel (2009) obtained annual national data on death certificates issued for violent deaths from 2005 through February of 2009 that are also close to IBC figures.

<sup>24</sup> According to ORB (2008a, New Casualty Tabs, Table 3) there were 0.25 conflict injuries per household, which is precisely equal to the number of deaths per household shown in ORB (2008a, New Casualty Tabs, Table 1).

<sup>25</sup> Roug and Smith (2006) published data on violent deaths but not injuries. However, the authors of the article collected MoH data on both deaths and injuries which they shared with us: personal communication with Doug Smith.

<sup>26</sup> ORB2 data suggest roughly 575,000 injuries in Baghdad, a figure that exceeds the MoH Baghdad figure of 28,343 injuries for the shorter period April 5, 2004 to June 1, 2006 by a factor of about 19.

injuries per killing (NINA [2008]). Even ORB's ratios in car bombings, aerial bombardments and other blasts are respectively just 1.3, 1.2 and 1.4. Such ratios are lower than those measured in explosions in a variety of different contexts.<sup>27</sup> Thus it would appear that if ORB has not overestimated the number of deaths in Iraq then it has underestimated the number of injuries. Yet, as argued above, ORB's injury numbers already appear to be implausibly high to begin with, so it is difficult to square this circle.

Third, the ORB2 numbers imply that approximately 600,000 Baghdad residents have been killed (notes to Table 2 above). If so, then there should be a noticeable gender imbalance in Baghdad. Several independent sources show that more than 80% of violent deaths during the ORB timeframe were adult males. The roughly 600,000 people killed in Baghdad implied by ORB's estimate would include roughly 500,000 adult males and 55,000 adult females.<sup>28</sup> Roughly 60.6% of the population of Iraq is adult (ILCS, 2005a) and the estimated population of Baghdad in 2006 was 6,962,650 (IFHS, 2008), implying an adult population in Baghdad of about 4,200,000. The ORB figures, together with these population figures, suggest that nearly a quarter of the adult male population of Baghdad had been killed by August of 2007.<sup>29</sup> If ORB's mortality figures for Baghdad are reasonably accurate the remaining adult females should outnumber the remaining adult males by a ratio of about 1.3 to 1.<sup>30</sup> Yet ORB2 itself found a ratio of adult females to adult males in Baghdad of just slightly under 1 and ORB1 and ORB3 found, respectively, ratios of females to males in Baghdad of roughly 1.04 and 0.92 to 1.<sup>31</sup> Note that ORB uses the "next birthday" method of selecting respondents within households, a method that is not biased toward selecting either males or females.<sup>32</sup> Thus, the gender balance within Baghdad measured in ORB's own polls, with possibly even more males than females, appears to be inconsistent with ORB's mortality estimate.<sup>33</sup>

## Conclusion

Without a doubt the war in Iraq has inflicted immense human costs on the Iraqi people. Yet the conflict mortality estimate of 1 million deaths published by ORB does not withstand scrutiny. It is inconsistent with the survey data of the ILCS and the IFHS which are much larger and higher-quality surveys. The ORB2 estimate also generates a number of implications which are not plausible: about 200,000 car-bomb deaths unnoticed by all international media, hundreds of thousands of unnoticed injured people who have apparently either not sought treatment or not been recorded as receiving any, a surprisingly low ratio of injuries to killings, and an implied gender imbalance in Baghdad that seems inconsistent with ORB's own polling data.

ORB gives little information about its methodology, for example, not disclosing its sample design or translations of its questions, while information that has been released suggests some significant weaknesses. ORB's presentation of margins of error and "ranges" for its estimates suggests that ORB has not calculated proper 95% confidence intervals. ORB2 did not record household rosters, asked an ambigu-

ously worded mortality question and did not account for non-response in its estimates or "ranges". The claim of ORB (2007b) to have drawn a "nationally representative sample" followed immediately by the ORB (2007c) admission of the need for substantial further rural sampling also casts doubt on ORB's sampling procedures.

Most importantly, our analysis of figures across the three ORB polls finds fundamental flaws in the data underpinning ORB's estimate. The ORB2 data are not suitable for deriving any credible estimate but, given proper scrutiny, it is clear that ORB has overestimated by a wide margin.

## Acknowledgements

We are especially grateful to two anonymous referees and to Peter Lynn for substantially improving our paper. We also thank Beth Daponte, Hamit Dardagan, David Kane, Will McLean and Robert Shone for excellent comments on an early draft. Finally, we thank Mohammed Ali for an early observation that led us eventually to the section of the paper on internal validity.

<sup>27</sup> Coupland and Meddings (1999) give wounded-to-killed ratios ranging from 3.3 to 20.0 with a median of 7.0 in a list of incidents involving explosives between 1996 and 1998.

<sup>28</sup> Gender proportions are derived from IBC, IFHS and L2. IBC covers the whole ORB time-frame, while IFHS and L2 are the closest of the existing surveys, covering to mid-2006. These sources give 83%, 89% and 85% respectively for adult male deaths, and 9%, 3% and 5% for adult females. In order to give ORB the best chance of consistency on gender balance, our calculations use the lowest of the three for males (83%) and the highest of the three for females (9%). ORB itself did not collect gender information on deaths.

<sup>29</sup> Note that, even using ORB's low measured ratio of injuries to killings of about 1 to 1, if one quarter of adult males in Baghdad have been killed then one third of the remaining adult males should be injured. Using a standard rule of thumb of 3 injuries per killing in conflict we would expect that all remaining adult males in Baghdad should be injured.

<sup>30</sup> The number of adult males in Baghdad should be something like  $((4,200,000)/2)-500,000 = 1,600,000$ . The number of adult females in Baghdad should be roughly  $((4,200,000)/2)-55,000 = 2,045,000$ . The ratio of the two numbers is 1.28. This calculation sets aside migration (see footnote 36).

<sup>31</sup> The sources are ORB (2007a, TablesFINALMarch07Irrq, Table 11), ORB (2008b, New Casualty Tabs, Table 7) and ORB (2008b, FINALTABLESMarch08, Table 51).

<sup>32</sup> "When knocking on the door of the household the interviewer initially collects the birthdays of each member of the household. If a respondent does not know the date of their birthday (something that is common amongst older Iraqis) then a random birthday is generated from a table each interviewer has. The individual with the next birthday is then selected for interview." (ORB, 2008b) Senior ORB staff confirmed to Spagat that they used this method in the ORB2 poll.

<sup>33</sup> Conceivably migration might resolve this discrepancy if, among the adult population, there has been large-scale, predominantly female outflows from Baghdad, large-scale predominantly male inflows into Baghdad or some combination of the two. We are not aware of data that would enable a convincing argument along these lines to be made.

## References

- American Association for Public Opinion Research. (2009). *AA-POR Finds Gilbert Burnham in Violation of Ethics Code*. [http://www.aapor.org/AAPOR\\_Finds\\_Gilbert\\_Burnham\\_in\\_Violation\\_of\\_Ethics\\_Code/1383.htm](http://www.aapor.org/AAPOR_Finds_Gilbert_Burnham_in_Violation_of_Ethics_Code/1383.htm).
- Biemer, P., & Lyberg, L. (2003). *Introduction to survey quality*. Hoboken, New Jersey: Wiley.
- Bloomberg School of Public Health. (2007). *Johns Hopkins Researchers Respond to ORB Poll on Iraq Mortality*. [http://www.jhsph.edu/refugee/publications\\_tools/iraq/burnham\\_ORBpoll.html](http://www.jhsph.edu/refugee/publications_tools/iraq/burnham_ORBpoll.html).
- Bloomberg School of Public Health. (2009). *Review Completed of 2006 Mortality Survey*. [http://www.jhsph.edu/publichealthnews/press\\_releases/2009/iraq\\_review.html](http://www.jhsph.edu/publichealthnews/press_releases/2009/iraq_review.html).
- Bohannon, J. (2006). Iraqi death estimates called too high: methods faulted. *Science*, 314, 396-397.
- Burnham, G., Lafta, R., Doocy, S., & Roberts, L. (2006). Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey. *The Lancet*, 368, 1421-1428.
- Burnham, G., & Roberts, L. (2007). *U.S. must face huge death toll of Iraqi civilians*. Baltimore Sun, October 9, <http://web.mit.edu/humancostiraq/further-reading/burnham-roberts2007.html>.
- Civic Worldwide. (2003). *Iraqi Civilian War Casualties*. <http://civilians.info/>.
- Coghlan, B., Brennan, R. J., Ngoy, P., Dofara, D., Otto, B., Clements, M., et al. (2006). Mortality in the Democratic Republic of Congo: a nationwide survey. *The Lancet*, 367, 44-51.
- Coupland, R., & Meddings, D. (1999). Mortality associated with use of weapons in armed conflicts, wartime atrocities, and civilian mass shootings: literature review. *British Medical Journal*, 319, 407-410.
- Daponte, B. (2007). Wartime estimates of Iraqi civilian casualties. *International Review of the Red Cross*, 89, 943-957.
- Dardagan, H., Sloboda, J., & Dougherty, J. (2006). *Reality checks: some responses to the latest Lancet estimates*. <http://www.iraqbodycount.org/analysis/beyond/reality-checks/>.
- Deltoid. (2004). *Breakdown of violent deaths in Lancet study*. <http://timlambert.org/2004/12/lancet11/>.
- EPIC. (2004). *An interview with EPIC advisor Richard Garfield*. [http://www.epic-usa.org/An\\_Interview\\_with\\_EPIC\\_A.html](http://www.epic-usa.org/An_Interview_with_EPIC_A.html).
- Gakidou, E., & King, G. (2006). Death by Survey: Estimating Adult Mortality Without Selection Bias from Sibling Survival Data. *Demography*, 43, 569-585.
- Gamel, K. (2009). *AP Impact: Secret Tally has 87,215 Iraqis Dead*. Associated Press, April 23.
- Guha-Sapir, D., & Degomme, O. (2007, June). *Estimating Mortality in Civil Conflicts: Lessons from Iraq: Triangulating different types of mortality data in Iraq*. CRED Working Paper.
- Iraq Body Count. (2003). *Adding indifference to injury*. <http://www.iraqbodycount.org/analysis/beyond/indifference-to-injury/>.
- Iraq Body Count. (2005). *A Dossier of Civilian Casualties in Iraq 2003-2005*. <http://www.iraqbodycount.org/analysis/reference/press-releases/12/>.
- Iraq Body Count (IBC). (n.d.). <http://www.iraqbodycount.org/>.
- Iraq Family Health Survey. (2008). *IFHS web site*. [http://www.emro.who.int/iraq/surveys\\_ifhs.htm](http://www.emro.who.int/iraq/surveys_ifhs.htm).
- Iraq Family Health Survey Study Group. (2008). Violence-related mortality in Iraq from 2002 to 2006. *New England Journal of Medicine*, 358, 484-493.
- Iraq Living Conditions Survey 2004. (2005a). *Content of IMIRA*. <http://www.fao.no/ais/middeast/iraq/imira/content.htm>.
- Iraq Living Conditions Survey 2004. (2005b). *Volume II: Analytical report*. [http://reliefweb.int/rw/RWFiles2005.nsf/FilesByRWDocUNID\\_FileName/KHII-6CC44A-undp-irq-31dec2.pdf/\\$File/undp-irq-31dec2.pdf](http://reliefweb.int/rw/RWFiles2005.nsf/FilesByRWDocUNID_FileName/KHII-6CC44A-undp-irq-31dec2.pdf/$File/undp-irq-31dec2.pdf).
- Johnson, N., Spagat, M., Gourley, S., Onnela, J., & Reinert, G. (2008). Bias in epidemiological studies of conflict mortality. *Journal of Peace Research*, 45, 653-664.
- KUOW Radio. (2008). *How many Iraqi Civilians Died because of the War*. <http://kuow.org/program.php?id=16092>.
- Laaksonen, S. (2008). Retrospective Two-Stage Cluster Sampling for Mortality in Iraq. *International Journal of Market Research*, 50, 403-417.
- London School of Hygiene and Tropical Medicine (LSHTM). (undated). *The use of epidemiological tools in conflict-affected populations: open-access educational resources for policy-makers*. <http://www.lshtm.ac.uk/hpu/conflict/epidemiology/index.htm>.
- McElwee, P. (2007). *A Million Iraqis Dead? The US Press Buries the Evidence*. Fair.org, <http://www.fair.org/index.php?page=3321>.
- McLean, W. (2007). *Baghdad now 60% Christian*. <http://willscommonplacebook.blogspot.com/2007/09/baghdad-now-60-christian.html>.
- McLean, W. (2007/2008). *Iraq Posts*. <http://willscommonplacebook.blogspot.com/search/label/Iraq>.
- Moody, J. (2006). Fighting a Hydra: A Note on the Network Embeddedness of the War on Terror. *Structure and Dynamics: eJournal of Anthropological and Related Sciences*, 1, 1-5.
- Munro, N. (2007, December 1). *Iraq's Slippery Polls*. National Journal.
- Munro, N., & Canon, C. (2008, January 4). *Data bomb*. National Journal.
- National Public Radio. (2007). *Survey puts Iraq War Dead above One Million*. <http://www.npr.org/templates/story/story.php?storyId=14501232>.
- NINA. (2008, June 21). *MOH Reveals Figures of 2007 Violence Victims*. [http://www.ninanews.com/English/News\\_Details.asp?ar95\\_VQ=EDGKJM](http://www.ninanews.com/English/News_Details.asp?ar95_VQ=EDGKJM) (login required).
- ORB. (2007a). *Despite Violence only 26% Preferred Life under Sadaam*. [http://www.opinion.co.uk/Newsroom\\_details.aspx?NewsId=103](http://www.opinion.co.uk/Newsroom_details.aspx?NewsId=103).
- ORB. (2007b). *Iraq Casualties Poll Update*. [http://www.opinion.co.uk/Newsroom\\_details.aspx?NewsId=111](http://www.opinion.co.uk/Newsroom_details.aspx?NewsId=111).
- ORB. (2007c). *More than 1,000,000 Iraqis Murdered*. <http://www.just-international.org/index.php/More-than-1000000-Iraqis-murdered-since-2003-invasion.html>.
- ORB. (2008a). *Iraqis Confident in Security by Concerned about Economy*. [http://www.opinion.co.uk/Newsroom\\_details.aspx?NewsId=126](http://www.opinion.co.uk/Newsroom_details.aspx?NewsId=126).
- ORB. (2008b). *Update on Iraqi Casualty Data*. [http://www.opinion.co.uk/Newsroom\\_details.aspx?NewsId=120](http://www.opinion.co.uk/Newsroom_details.aspx?NewsId=120).
- Pedersen, J. (2007). *Mortality and Missing Family Members in Iraq Following the 2003 Invasion*. unpublished manuscript, Fafo-AIS, Norway.
- Research Talk. (2006). *ESOMAR: Life and Death MR*. <http://www.researchtalk.co.uk/rt/index.php?s=Munqeth+Daghir>.
- Roberts, L., Lafta, R., Garfield, R., Khudhairi, J., & Burnham, G.

- (2004). Mortality before and after the 2003 invasion of Iraq: cluster sample survey. *The Lancet*, 364, 1857-1864.
- Rosenblum, M., & van der Laan, M. J. (2009). Confidence Intervals for the Population Mean Tailored to Small Sample Sizes, with Applications to Survey Sampling. *International Journal of Biostatistics*, 5, 118.
- Roug, L., & Smith, D. (2006, June 25). *War's iraqi death toll tops 50,000*. Los Angeles Times.
- SMART. (2006). *Measuring mortality, nutritional status, and food security in crisis situations: SMART METHODOLOGY*. [http://www.smartindicators.org/SMART\\_Methodology\\_08-07-2006.pdf](http://www.smartindicators.org/SMART_Methodology_08-07-2006.pdf).
- Spagat, M. (2008). *Mainstreaming and Outlier: The Quest to Corroborate the Second Lancet Survey of Mortality in Iraq*. unpublished manuscript. <http://personal.rhul.ac.uk/uhte/014/Mainstreaming.pdf>.
- Spagat, M. (2010). Ethical and Data-Integrity Problems in the Second Lancet Survey of Mortality in Iraq. *Defense and Peace Economics*, 21(1), 1-41.
- Spagat, M. (forthcoming). *Ethical and Data-Integrity Problems in the Second Lancet Survey of Mortality in Iraq*. Defense and Peace Economics. <http://personal.rhul.ac.uk/uhte/014/Standards.pdf>.
- Spiegel, P., & Salama, P. (2000). War and mortality in Kosovo, 1998-99: an epidemiological testimony. *Lancet*, 355, 2204-2209.
- Steele, J., & Goldenberg, S. (2008, March 19). *What is the Real Death Toll in Iraq?* The Guardian.
- Tapp, C., Burkle, F. M., Wilson, K., Takaro, T., Guyatt, G. H., Amad, H., et al. (2008). Iraq War Mortality Estimates: a Systematic Review. *Conflict and Health*, 2(1).
- UN Assistance Mission for Iraq (UNAMI). (2007). *Human Rights Report: 1 November - 31 December 2006*. <http://www.uniraq.org/FileLib/misc/HR%20Report%20Nov%20Dec%202006%20EN.pdf>.
- White, D., & Housman, M. (2002). The Navigability of Strong Ties: Small Worlds, Tie Strength and Network Topology. *Complexity*, 8, 72-81.