UCL
Université catholique de Louvain

Working paper

Estimating mortality in civil conflicts: Lessons from Iraq

Debarati Guha-Sapir
Olivier Degomme

June 2007

Complex Emergency Database

WHO Collaborating Centre for Research on the Epidemiology of Disasters
CRED

Estimating mortality in civil conflicts: lessons from Iraq

Triangulating different types of mortality data in Iraq

Debarati Guha-Sapir[1] ,Olivier Degomme[1]

June 2007

Abstract

The civil and political conditions in Iraq have steadily degenerated since the military invasion of March 19th, 2003. Civilians are increasingly bearing the brunt of relentless violence. Extremely poor security conditions are disrupting the lives and livelihoods of millions and the end is not in sight. Recent studies on the human costs of war have focused on measuring deaths due to violence. The debate around the numbers of excess deaths has opened a Pandora's Box of methodological issues related to measuring mortality in conflicts, underlining the political sensitivities that accompany such exercises.

[1] Centre for Research on the Epidemiology of Disasters (CRED)

1. Introduction

The civil and political conditions in Iraq have steadily degenerated since the military invasion of March 19th, 2003. Civilians are increasingly bearing the brunt of relentless violence. Extremely poor security conditions are disrupting the lives and livelihoods of millions and the end is not in sight.

Recent studies on the human costs of war have focused on measuring deaths due to violence [1-3]. The debate around the numbers of excess deaths has opened a Pandora's Box of methodological issues related to measuring mortality in conflicts, underlining the political sensitivities that accompany such exercises. The credibility of mortality estimates from conflict situations depends on three factors:

- Political objectivity of the originators
- Soundness of methodological foundations
- Transparency of declared limitations

In most studies of mass mortality, estimates are derived from multiple sources and are expressed in ranges [2, 4-6]. A single precise number gives a sense of accuracy that is deceptive in survey type sources.

Recent estimates of war-related deaths in Iraq vary widely and have been the subject of much discussion (Table 1). But even before the invasion and the ensuing chaos, data from Iraq presented challenges. For instance, UNICEF's Multiple Indicator Cluster Survey (MICS) 2000 and 2006 analyses [7, 8] shows a substantial reduction in child mortality rate (U5MR) between two contiguous pre-invasion 5-year periods, which requires further examination. Similarly, estimates from the post-invasion period range from 43,771 to 15 times that number.

Table 1: Mortality estimates from surveys in Iraq 1994-2006

	period covered	sample size (hh)	U5MR (95%CI)	CMR (95%CI)	war-related deaths (95%CI)	area covered	source
1.	1994-1999	24,000	131 [127;135]	-	NA	Iraq excl. Kurdish gov	Ali & Shah (2000)
2.	1994-1999	16,000	72 [68;76]	-	NA	Kurdish gov	Ali & Shah (2000)
3.	1999-2003	21,668	40	-	NA	Iraq	Iraq Living Conditions Survey (2004)
4.	2001-2006	18,144	41	-	-	Iraq	Multiple Indicator Cluster Survey (2007)
5.	2002-2003	988	-	5 [3.7;6.3]	NA	Iraq	Roberts et al (2004)
6.	2002-2003	1,849	-	5.5 [4.3;7.1]	NA	Iraq *	Burnham et al (2006)
7.	2002-2004	21,668	-		23,743 [18,187; 29,299]	Iraq	Iraq Living Conditions Survey (2004)
8.	2003-2004	988	-	7.9 [5.6 ;10.2]	98,000 [8,000; 194,000] (including violent and non-violent deaths)	Iraq †	Roberts et al (2004)
9.	2003-2006 (June 30)	1,849	-	7.2 [5.2 ;9.5]	601,027 [426,369; 793,663]	Iraq *	Burnham et al (2006)
10.	2003-2006 (June 30)	-	-	-	43,771 [41,441; 46,101]	Iraq	Iraq Body Count (2006)

* excl. Dahuk, Muthanna

† excl. Anbar

We re-estimate mortality in Iraq using data from multiple sources based on their methods and coverage. Burnham [9] and IBC [10] provided raw data and were the most complete data sources on war-related mortality. We also used the report from the Iraq Livelihoods Condition Survey [11]. The strengths and weaknesses of each are described below. We triangulate the findings of these three data sources offsetting strengths against weaknesses. We also draw methodological lessons for future mortality studies in conflicts.

2. Summary analysis of the Burnham survey dataset

The Burnham survey estimated deaths caused by the US led invasion using a national, cross-sectional mortality survey. The survey itself, undertaken between May – July 10th, 2006 covered 47 clusters and a recall period from January 1st, 2002 to June 30th, 2006. The sample estimates were extrapolated to the 2004 population of all governorates (UNDP Census). Thus each survey sample death represented about 2000 population-based deaths. The authors presented a point estimate of 654,965 deaths due to the 2003 invasion, an average of 594 civilian war casualties everyday over 3 years and 4 months. Of these, nearly 92% were attributable to war related violence based on a pre-invasion baseline of 2 violence related deaths. The study has received wide publicity in both academic as well as political circles, underlining as it did, the devastating effects of the war on civilian populations [12-14].

We checked the data for accuracy and validity of the numerator, denominator and sampling weights of governorates. The following errors and methodological weaknesses were identified.

Validity of numerator: Eleven deaths were removed from the numerator because the households in which they occurred were excluded from the denominator due to missing household size. In addition, since the recall period ended on June 30th [9, 15], 24 car bomb deaths that occurred in cluster 33 in July 2006 were eliminated.

Sampling distortions: Three governorates with the highest violence-related mortality rates were over sampled and 3 governorates with the lowest violence-related mortality rates were under sampled (Table 2). In order to correct this, we applied weights to each governorate according to its population to re-estimate nationwide mortality rates.

Table 2: Population and sample distribution by governorate

	% killed due to violence after the invasion	Rank	Sample	Population	Ratio Sample/ Population
			OVERSAMPLED		
Anbar	5.9%	3	7.5%	5.1%	1.47
Thi-Qar	2.4%	13	7.1%	5.7%	1.25
Ninewa	6.5%	2	11.1%	9.8%	1.13
Baghdad	4.2%	6	27.1%	25.1%	1.08
Diyala	9.1%	1	5.7%	5.3%	1.08
Erbil	2.7%	9	5.8%	5.4%	1.07
Babylon	2.7%	10	6.0%	5.6%	1.07
			PROPORTIONALLY SAMPLED		
Najaf	2.6%	11	3.7%	3.7%	1.00
Salah Al-Din	5.7%	4	4.3%	4.3%	1.00
			UNDERSAMPLED		
Tameem	1.6%	14	2.0%	3.3%	0.61
Kerbala	3.2%	8	1.8%	2.9%	0.62
Qadissiya	4.0%	7	2.2%	3.5%	0.63
Wassit	1.3%	15	2.5%	3.7%	0.68
Sulaymaniyah	0.9%	16	4.5%	6.6%	0.68
Missan	2.5%	12	2.1%	3.0%	0.70
Basrah	5.6%	5	6.6%	6.9%	0.96
Total			100.0%	100%	1.00

Top 3 high violence (arrows to Anbar, Ninewa, Diyala)

Top 3 low violence (arrows to Tameem, Wassit, Sulaymaniyah)

Design Effects: We calculated the design effects (DEFF) for the overall and cause-specific violence-related mortality rates[2]. The DEFF for all violence-related deaths was 5.4 and for some specific causes, 9.6. This indicates that the violent death phenomenon was highly clustered over time or space, making it difficult to reliably generalise to the entire country.

Validity of denominator: Recalculating household size from each record, we obtained 66 households with less than 1 member per household and 33 of those had a negative (<0) household size in 2002. More importantly, the authors did not take population migration into account for the denominator, the implications of which can be sizeable. A household member who enters early in the recall period (e.g. Feb. 2003) is at risk of death for 53 months, whereas a person entering in Feb. 2006 is at risk for 5 months. In general, the mid-period population in a scenario where most out-migration occurrs before in-migration will be lower than when in-migration precedes out-migration,

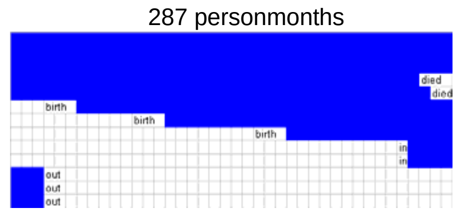
---

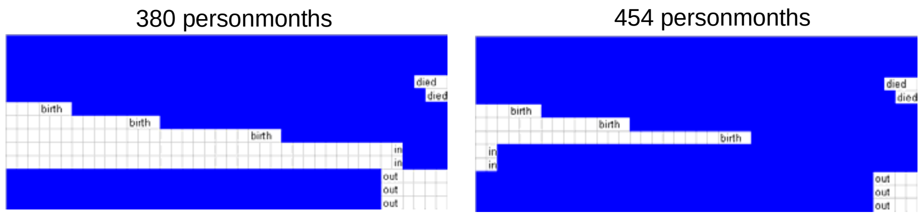
[2] The design effect, which measures variance, gained or lost, by using a complex, rather than a simple random sample, depends on cluster size and cluster homogeneity. A DEFF of 2 (commonly used as an acceptable standard) requires a sample twice the size of a simple random sample for the comparable level of precision.

changing the denominator and the rates. A graphical representation of the potential effects is presented in Figure 1.

Figure 1: Schematic representation of the effects of in- and out-migration on the denominator

HH size (2006)	8
Births	3
Deaths	2
In	2
Out	3





In Table 3, we present mortality estimates derived from the partially corrected Burnham dataset using two extreme migration scenarios. These scenarios produce estimates that are between 15 – 30% lower than the Burnham estimates with similarly wide confidence intervals. As long as the same basic dataset is used, the few adjustments that are statistically possible, do not substantially change the estimates although they point to a serious need for greater thoroughness in data processing and verification, and accuracy in data presentation.

Table 3: Corrected mortality estimates based on Burnham et al.

	Burnham et al.	Guha-Sapir and Degomme	
		scenario 1	scenario 2
Pre-invasion			
- CMR	5.5	5.6	5.0
Post-invasion			
- CMR	13.2	11.0	11.4
- excess deaths	654, 965	464,203	552,257
95%CI	[392,976; 942,636]	[211,062; 713,352]	[303,210; 796,555]

However, four important weaknesses did not allow for correction.

First, design effects were higher than the customarily acceptable levels. The heterogeneity of violent deaths in Iraq will generate high DEFFs but larger sample sizes would have increased the estimate precision. The sampling design should have been adapted in light of this, e.g. by stratifying the clusters in governorates according to high and low violence areas.

Second, displacement information was not taken into consideration and is likely to have significant consequences on the estimation of rates. Population movements from areas heavily affected by violence towards more peaceful regions have increased considerably since 2003, changing the population distribution by governorate and therefore distorting the sampling weights.

Third, Burnham et al. used recall periods which ended the day of the interview and therefore varied from day to day, spanning two months. In a period of high escalation of violence in certain areas like Baghdad or Diyala, the day in which the cluster is surveyed can have a disproportionate impact on the result.

Fourth, the nation-wide estimates are unstable due to very small numbers. An error of one death (e.g. on the pre-invasion baseline of 2 deaths from which excess violent mortality is calculated), can change the results significantly.

Nevertheless, there are several strengths in the Burnham dataset, not the least of which is its unique contribution to the estimations of war-related mortality in Iraq. The survey collected specific data on war-related deaths using a cluster design, widely acknowledged to be an appropriate method for these purposes. It systematically covered all except two provinces and has obtained information on causes of civilian death – a rare achievement in war conditions. To date, it remains the most systematic dataset on invasion-related mortality, to be available to the public.

3. Summary Analysis of IBC data set

The Iraq Body Count (IBC), an UK based voluntary project, started monitoring non-military casualties in 2003 using a census type enumeration of deaths. Criteria for entry in the IBC database requires reporting by at least two independent media sources that have met specific standards such as: at least daily website updates, English language reporting and public access. IBC also requires the source to separately archive all reports on their site.

For each incident, the following information is entered: date, time and location of the incident, the target according to military sources, weapons used, civilian deaths (minimum and maximum reported) and sources. Entries are adjusted when further details become available at a later date. For the post-invasion period covered by the Burnham survey (March 2003-June 2006), the number of casualties reported by IBC was between 41,423 and 46,081 or an average of 34 to 38 deaths per day.

The IBC gave us access to their dataset with information on the methodology and limitations. Their data can be used for different types of analyses. First, aggregated monthly figures can be obtained for the entire country as well as specific governorates. This allows for time trend analyses and the comparison of these trends from different areas.

Second, the number of reports per incident can be obtained by governorate. This information can be used as proxy for the coverage of deaths by the media in a governorate. An analysis of this data shows that in Baghdad, about 75% of the incidents were reported by more than 3 independent sources, with an average of more than 5 sources per event. However, the same analysis shows that other governorates have lower media coverage. Diyala and Ninewa have only half of the events reported by more than 3 sources, Anbar less than 60%. This suggests that in governorates other than Baghdad, fewer events are reported by the media and thus not included in the Iraq Body Count database. In addition, the non-inclusion of incidents reported by one source may contribute to the underestimation of casualties.

While IBC is undoubtedly missing some deaths in Baghdad, it is unlikely that they would miss an average of over 100 violent deaths a day, given the level of media coverage in the city. We therefore conclude that their Baghdad mortality estimate is close to complete, further corroborated by the ILCS estimates (see below).

4. Other data sources

Besides the two data sets analysed above, there are three notable mortality data sources available to the public.

The first of these is the Iraq Living Conditions Survey conducted by UNDP and Iraq's Central Organization for Statistics and Information Technology (COSIT) with technical assistance of the Fafo - Institute for Applied International Studies (Norway). The survey had a nationwide sample drawn by a standard stratified two-stage design of about 22,000 households, more than 10 times the size of Burnham's sample. It gathered data on war-related deaths dating from the invasion until May 2004. To ensure the technical quality and integrity of data collection, the local supervisors and interviewers were given an intensive 3 week training by Fafo in Amman. In addition, external teams monitored the interviewers and when necessary they were sent back to complete or correct the questionnaires. Due to a careful sample design, the confidence interval for the point estimate 23,743, was substantially narrower compared to the Burnham survey. Data was continuously sent to Fafo headquarters where it was checked for validity and consistency. In addition, to ensure accuracy, survey supervisors re-interviewed 10% of the households for interview errors and omissions.

Overall, supervision was externally guaranteed by 2 Fafo employees visiting different governorates and one to oversee data entry in Baghdad. In view of the close external quality control of the fieldwork, the methodological detail and the levels of precision, we consider the results to be one of the most accurate available for the year following invasion.

The Brookings Institution's Iraq Index [16] depends on the IBC as one source for their data and therefore was not included in this analysis. The UNICEF's MICS dataset collects data only on child mortality.

5. Triangulated CRED estimates.

Two independent datasets that were complete both in terms of geographic coverage and time were IBC and Burnham datasets. The third independent dataset was the Iraq Living Condition Survey which covered deaths until May 2004.

The Burnham estimates of deaths in the post invasion period are much higher than any other estimate. Even the lower limit of its 95%CI is higher than the highest estimate from any other source (Table 1). Further, weaknesses cited earlier as well as several inconsistencies in their published work undermine the reliability of their estimates. We therefore triangulated the Burnham estimate with IBC and ILCS data to propose revised death tolls.

For period 1 (March 2003 – April 2004) data from IBC, ILCS and the Burnham study were available. Given the methodological advantages, we assume the death estimates from ILCS to be the closest approximation to reality, accounting for a total of 23,743 casualties. For Baghdad, the IBC death toll was close to the ILCS casualty count, with a difference of 9%, indicating close to complete coverage. On the other hand, the Burnham figure for the comparable period was 13,244 – 64% higher than ILCS. For the rest of the country however, the coverage by IBC is likely to be lower as indicated in our analysis (cfr section 3) and by the difference with the ILCS (6,130 vs. 16,196 respectively) (Table 4).

For period 2, which the ILCS did not cover, we had to calculate a correction factor by adjusting the IBC value for Baghdad by the ratio IBC/ILCS (=0.91) we obtained from our analysis from period 1. We thus increased the IBC death toll of 18,657 by approximately 10% and obtained 20,582 for Baghdad. We then calculated the Burnham/IBC ratio (= 4.2) which reflected the magnitude of Burnham's overestimation. Finally, we adjusted the

Burnham estimate downwards for the rest of Iraq using this ratio. We thus obtained an adjusted estimate of 81,418.

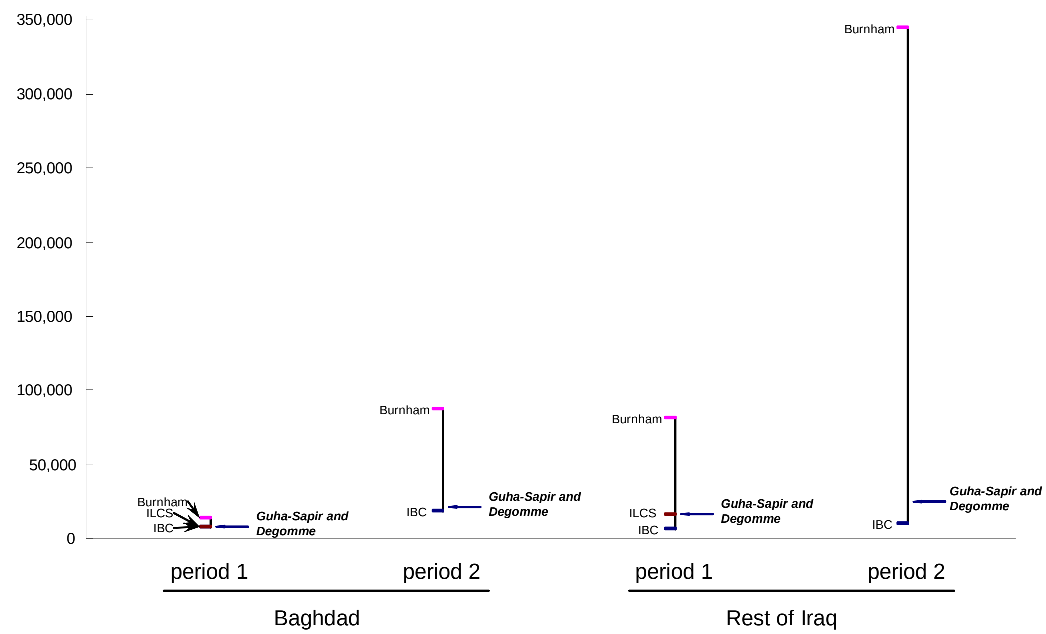
Our re-estimation of total war-related death toll for Iraq from the invasion until June 2006 is therefore around 125,000.

We would like however, to underline that in uncertain circumstances, exact numbers such as these are misleading in their precision and should only be indicative of the magnitude of mortality. Mortality could be half again as much but is unlikely to be 5 times as much.

Table 4: Triangulated (CRED) estimates of war related deaths among Iraqi civilians using survey and census type data

	Baghdad		Rest of Iraq	
Period 1 (March 2003 – April 2004)	7,547	ILCS estimate	16,196	ILCS estimate
Period 2 (May 2004 – June 2006)	20,582	IBC with correction for underreporting	81,418	Burnham with correction for overreporting

Figure 2: Death tolls in post-invasion Iraq by period and source



6. What have we learnt?

First, studying war related deaths in politically charged situations such as the one in Iraq requires extreme vigilance for errors in the data processing and in data presentation. The sensitivity of the context will inevitably expose the data to detailed scrutiny. Therefore, slip ups in these areas are not only gratuitous but provides easy openings for critics of the study findings.

Second, point estimates of numbers of deaths (e.g. 509,123) from single surveys arguably focus public attention but often end up with a boomerang effect – discrediting the cause. In both Darfur and Iraq, high numbers of war-related deaths of unconvincing precision have generated a debate that has not been helpful in pushing forward the humanitarian agenda. Meta-analyses of multiple data source estimates which propose ranges of deaths are a more credible approach.

Third, recall periods for conflict mortality surveys should not be longer than six months and the same period should be applied to all surveyed units. Due to population displacements and rapid escalation of violence, typically inherent to these situations, different recall periods introduce unnecessary bias and methodological complexities.

Fourth, methods other than population based surveys should be considered to establish numbers of deaths due to war. Sample surveys may not be appropriate in situations where displacement is common and inaccessibility of populations due to poor security compromise sampling. Studies estimating deaths from mass violence frequently employ a combination of testimony data, eye-witness accounts, records of executions and forensic inspections of mass graves to gather individual death records. The Office of the Prosecutor of the International Criminal Court triangulated databases of death records to reconstruct war-related deaths in Bosnia and Herzegovina (Tabeau and Bijak 2005). Heuveline [17] used the UN electoral lists to estimate the population size after the Khmer Rouge regime and the extent of "excess" mortality in the 1970s. More recently, Silva and Ball have used Multiple Systems Estimation techniques as well as capture–recapture methods for their estimations of deaths in Timor Leste.

Finally, the fundamental purpose of conducting mortality surveys - which cost money and where lives of both interviewers and interviewees are placed at risk - need to be clearer. Do these exercises actually serve operational purposes such as help redirect resources? Are the results available fast enough to save lives? Are they any use to advocate for a war to end? Do they have any real effect on political decision making? Do point estimates with large confidence intervals help human rights investigations?

7. References

1. Coghlan, B., et al., *Mortality in the Democratic Republic of Congo: a nationwide survey.* Lancet, 2006. **367**(9504): p. 44-51.

2. Guha-Sapir, D. and O. Degomme, Estimating Death Tolls in Darfur - Report prepared for the International Criminal Court 2006.

3. Roberts, L., et al., Mortality before and after the 2003 invasion of Iraq: cluster sample survey. Lancet, 2004. **364**(9448): p. 1857-1864.

4. Kiernan, B., The Demography of Genocide in Southeast Asia: The Death Tolls in Cambodia, 1975-79, and East Timor, 1975-80. Critical Asian Studies, 2003. **35**(4): p. 585-597.

5. Silva, R. and P. Ball. The Demography of Large-Scale Human Rights Atrocities: Integrating demographic and statistical analysis into post-conflict historical clarification in Timor-Leste. in meetings of the Population Association of America. 2006. Los Angeles (USA).

6. Tabeau, E. and J. Bijak, War-related deaths in the 1992-1995 armed conflicts in Bosnia and Herzegovina: A critique of previous estimates and recent results. European Journal of Population-Revue Europeenne De Demographie, 2005. **21**(2-3): p. 187-215.

7. Ali, H.M. and I.H. Shah, *Sanctions and childhood mortality in Iraq.* Lancet, 2000. **355**(9218): p. 1851-1857.

8. UNICEF, Findings from the Iraq Multiple Indicator Cluster Survey 2006. Preliminary Report. 2007.

9. Burnham, G., et al., Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey. Lancet, 2006. **368**(9545): p. 1421-1428.

10. Iraq Body Count. [cited; Available from: http://www.iraqbodycount.org/.

11. UN Development Programme. *Iraq living conditions survey 2004*. 2005 [cited 22/05/2007]; Available from: http://www.iq.undp.org/ILCS/overview.htm.

12. Bohannon, J., Epidemiology. Iraqi death estimates called too high; methods faulted. Science, 2006. **314**(5798): p. 396-7.

13. Brown, D., Study Claims Iraq's 'Excess' Death Toll Has Reached 655,000, in Washington Post. 2006.

14. Giles, J., Death toll in Iraq: survey team takes on its critics. Nature, 2007. **446**(7131): p. 6-7.

15. Johns Hopkins Bloomberg School of Public Health. *Updated Iraq Survey Affirms Earlier Mortality Estimates*. 2006 [cited 07/07/2007]; Available from: http://www.jhsph.edu/publichealthnews/press_releases/2006/burnham_iraq_2006.html.

16. The Brookings Institution. *Iraq Index*. [cited 07/07/2007]; Available from: http://www.brookings.edu/iraqindex.

17. Heuveline, P., 'Between one and three million': towards the demographic reconstruction of a decade of Cambodian history (1970-79). Popul Stud (Camb), 1998. **52**(1): p. 49-65.