

HiCN Households in Conflict Network

The Institute of Development Studies - at the University of Sussex - Falmer - Brighton - BN1 9RE

www.hicn.org

Mortality after the 2003 invasion of Iraq: Were valid and ethical field methods used in this survey?

Madelyn Hsiao-Rei Hicks*

mjhhicks@aol.com

HiCN Research Design Note 3

1 December 2006

Summary: The Lancet has published the methodology and findings of a 2006 survey by Gilbert Burnham and colleagues of mortality after the 2003 invasion of Iraq. The authors report that 40 households were interviewed a day in this survey, and illustrate the ease of this completion rate by comparison to their 2004 study in which teams interviewed 30 households in three hours (completing on average one interview every 6 minutes). This paper describes in detail the problems presented by this reported rapid interviewing rate: inadequacy of the timeframe, likely compromise to data validity, increased risk to interviewees, and the improbability of maintaining ethical standards for academic epidemiological research. Conflict-related mortality surveys should be based on valid field methods that systematically maintain an ethical relationship with the population being represented. It is suggested that Burnham and colleagues need to provide a fully detailed methodological description of their study coupled with access to their raw data to establish that these standards were met for their survey in Iraq.

Copyright © Madelyn Hsiao-Rei Hicks 2006

* Madelyn Hsiao-Rei Hicks, MD, MRCPsych *Affiliation:* Honorary Lecturer, Sections of Community (PRiSM) and Cultural Psychiatry, Department of Health Services Research, Institute of Psychiatry, King's College London, University of London, UK.

In October 2006 the Lancet published an epidemiological survey of Iraq that reported an exceptionally high estimate of conflict-related mortality. In this paper, Gilbert Burnham and colleagues (2006a) conclude that about 600,000 (95% CI 426,369 to 793,663) violent deaths occurred in Iraq by July 2006, all attributed to the invasion of Iraq or post-invasion conflict. The Iraq mortality estimate of these researchers has drawn a great deal of attention, in part because of the political impact of its release just weeks before the US elections. In fact, discussion of this study's scientific merits is often hampered by a confusion of two separate matters: the matter of political opinion about the invasion and the matter of examining the validity of the study's methods. Whatever one's political views may be, confidence in a study's findings must rest on a clearly described and scientifically sound methodology. For epidemiological studies, readers and reviewers usually focus on assessments of statistical power and on whether the appropriate statistical methods were applied. For example, the methodology of this study has drawn criticism from some conflict mortality researchers who suggest that its selection of households within clusters is systematically biased (Gourley et al., 2006). But statistics are only meaningful if they are based on high-quality, valid data gathered in the field. The present paper focuses on the adequacy of this study's field methods in terms of the recruitment and implementation of individual household interviews.

The composition of the interview teams is described in the Methods of the Lancet paper (Burnham et al., 2006a): "The two survey teams each consisted of two female and two male interviewers, with the field manager serving as supervisor." In their Results section the authors later note, "One team could typically complete a cluster of 40

households in 1 day”. Forty household interviews per day would appear to be a startlingly high completion rate for most readers who have carried out house-to-house epidemiological surveys themselves. Examination of the authors’ companion paper for additional clarification apparently confirms their report that about 40 interviews were routinely done by one intact team per day: “Once the “start house” or location is selected, then the survey team moves to the next nearest (or sometimes the second or third nearest) house until the specified number of houses are selected (often from 10-50) to be interviewed in that cluster” (Burnham et al., 2006b, p. A2).

Based on the methodology published by the authors (Burnham et al., 2006a, 2006b), the difficulty of achieving 40 household interviews per day can be initially illustrated with the following scenario. As neither paper describes how long individual interviews usually lasted or how much of the day teams spent on doing their 40 interviews, one can generously imagine that a team manages to complete 10 hours of continuous, back-to-back interviews daily. This theoretical 10 hours of daily interviewing is charitable as it would be practically unsustainable with the 130°F (55°C) summer heat and checkpoint travel delays to and from clusters that the authors describe (Burnham et al., 2006b). However, this maximum imaginable 10 hours interviewing as divided among 40 households would allow a mean of 15 minutes per household interview. This estimated 15 minutes per interview is then whittled down because it cannot include everything involved in door-to-door interviewing. For example, teams have to move from household to household. Even moving between adjacent households in a cluster, as was done in this study, takes time. Sometimes no one answers the door while residents peek through the

curtains or check first with someone else at home. Once the door is opened, the team must be introduced and they must explain why they are there. If invited in, further introductions are made to household members, chairs are offered and refreshment is served. It is important to understand that prospective interviewees do this not just to be polite, but for the opportunity to assess who these researchers really might be, their underlying motives, how they might choose to engage with these strangers, and what they would have to gain or lose in this engagement. The time required for these tasks should not be significantly shortened by the authors' use of "word of mouth" in the community about their survey since epidemiological researchers are expected to explain their study and to obtain informed consent directly from every interviewee on an individual basis, as well as securing privacy for household interviews so that interviewees' refusal or agreement to participate, and their answers, are kept confidential.

It is after this typical routine that an interview can start. Once an interview starts, even a short questionnaire often takes more than 15 minutes, but this is particularly likely when questions focus on violence-related losses. Especially with traumatic and important losses like a violent death in the family, many people answer not with one-word 'facts' in a questionnaire format, but with stories. These stories can be important to interviewees as a way of processing the loss, or as testimonials of the value of the dead individual or of the grief and suffering of the family. How the individual died, a child, a father, is often so important to a family that the interviewee will answer with his or her narrative of that death, its circumstances and its impact on the family. This narrative can add to the

richness of the data and may answer a number of questionnaire items spontaneously, but it requires time.

It was based on this scenario of a maximum imaginable 10 hours of interviewing daily, thus permitting a mean 15 minutes per interview, that Madelyn Hicks, a public health researcher, was quoted in Science as saying she “simply cannot believe” the authors’ claim that 40 households were interviewed per day (Bohannon, 2006). Subsequent responses by the authors to explain their high interview completion rate only deepen the impression that something is amiss with the implementation of their study in the field: In an interview with BBC News, Les Roberts responded to the questioning of their 40 interviews per day with this explanation: “In Iraq in 2004, the surveys took about twice as long and it usually took a two-person team about three hours to interview a 30-house cluster. I remember one rural cluster that took about six hours and we got back after dark. Nonetheless, Dr. Hicks' concerns are not valid as many days one team interviewed two clusters in 2004” (BBC News, 2006).

Considering this clarification of the study’s interview methodology yields some disturbing conclusions. Roberts says here that in their 2004 Iraq mortality survey (Roberts et al., 2004), a two-person team would spend three hours (180 minutes) to interview 30 households in a cluster. 180 minutes divided among 30 households gives a mean *6 minutes* per household interview. Roberts then goes on to make the incredible statement that “...many days one team interviewed two clusters in 2004”, which would apparently be about 60 interviews in one day plus time needed to travel between the two

clusters. This additional information that one household interview was done about every 6 minutes is not reassuring. It only increases the degree of improbability, particularly if readers are to believe that the interviews were consistently done in an ethical and valid way as asserted by the authors' written methodology. Moreover, if in saying that this was "twice as long" Roberts is making a comparison to the 2006 interviews, then he is implying that about 3 *minutes* time was allocated per household interview in the 2006 study.

The authors of this study have offered various explanations to defend their field methods subsequent to their paper's publication. First, Burnham stated that 40 interviews could be done daily by being "well organized" (Bohannon, 2006). Later, Burnham and Roberts added that each team was actually split into two-person teams to do 20 interviews each per cluster (Boseley, 2006; Burnham & Roberts, 2006). Both these explanations are inadequate in view of Roberts's disclosure of the extremely short time spent by fieldworkers on obtaining interviews for the Iraq surveys. The short interview times are also inadequately explained by the authors' use of proximity sampling within clusters, whereby they randomly sampled one household in a cluster and then proceeded to interview the 39 households nearest to that one. The convenience of walking between adjacent households, and the team's reliance on "word of mouth among households" (Burnham et al., 2006a) to smooth the way, would not save enough time for an interview to be completed in a valid and ethical manner every 6 minutes. As described earlier, even a theoretical average of 15 minutes per interview would have been too short. With the new information that the authors of this study allocated about 3 to 6 minutes per

household interview, it is clear that a full and detailed demonstration needs to be provided of if, and how exactly, this was done.

To illustrate the improbability of completing valid interviews within the described short timeframe, all the elements described in the authors' published methodology, and those needed to satisfy the requirements of ethical practice in academic epidemiological research, are listed here:

- Travel from household to household within the cluster.
- Knocking on the door and waiting for this to be answered. Gaining entry and participating in social pleasantries.
- The authors write that they needed to provide “lengthy explanations of the purposes of the survey” in order to overcome initial “suspicion” (Burnham et al., 2006b, p. A3).
- *Obtaining informed consent:* After explaining a study, interviewers are supposed to determine that the interviewee understands its purpose. They are supposed to allow time for the interviewee to deliberate on whether to participate in the study, without feeling coerced. This process is referred to as ‘obtaining informed consent’; an ethical requirement for any research done with human subjects by an academic institution. ‘Informed consent’ differs from a simple consent to be asked some questions: It places a responsibility on interviewers to ensure that interviewees understand the study’s purpose, its potential risks and benefits, and the interviewee’s freedom not to take part in the survey. In some settings it is appropriate to do this verbally without the

additional requirement of the interviewee's signature or anonymous mark on a written 'informed consent form'. But in either case, a study should have written guidelines for interviewers so that they consistently explain the survey appropriately and the 'informed consent' process must be followed. Obtaining informed consent can consume more than 6 minutes before a survey interview even begins. Burnham et al. (2006a) write that they obtained "oral consent" before their interviews.

- Before describing the study's aims as part of the informed consent process, and before the interviewee is asked whether he or she agrees to participate, and then the survey questions, interviewers are expected to secure and to control an interview environment that ensures that information will be private. Any reasons to deviate from this standard, and the potential effects of such deviation on the survey results, should be anticipated by the study's researchers and a clear protocol should be established for interviewers to follow.
- *"Word of mouth" advertising and the potential compromises to the informed consent process:* In their Methods (Burnham et al., 2006a) the authors write, "By confining the survey to a cluster of houses close to one another it was felt the benign purpose of the survey would spread quickly by word of mouth among households, thus lessening risk to interviewers". One problem from relying on community-wide communication of a study's purpose is that interviewers, especially interviewers under time pressure, might not complete the informed consent process because it is tempting to assume that

interviewees 'already know' about the study. No doubt they know something if a neighbor, acquaintance, relative or stranger in the community has spoken to them about the survey before the researchers arrive; but what do they know? Information is frequently distorted, which is one reason that epidemiological studies require the researchers, not the community, to explain the study to the interviewee. There is no such thing in epidemiological research as 'community informed consent'. A second problem is that not only can the study's aims be distorted during dissemination across the community, but the very risks and benefits of participation can be changed for interviewees. By providing community members with information about the study and its targeted subjects (adjacent households are easily identified), researchers provide everyone with access to this information with opportunities to invest participation in the project with new or modified risks and benefits. Third, information about a study rarely travels equally across community members. Instead, it tends to travel, like all information, in channels directed by power, allegiance, privilege or convenience. Fourth, in epidemiological studies of potentially sensitive subjects, including violence, it can be very important that confidentiality from the community is maintained not just for the content of an interviewee's answers, but also for their very participation in the study and the study's aims. Lack of such privacy can place interviewees in the position of being coerced. Just as they are being told by other community members what the researchers will be doing in this study, they might also be told whether they should participate and what they should

say. With the awareness that researchers are publicizing the purpose of their study in the community and the households that they will be surveying, interviewees might not feel secure that their answers would not also be publicized in the community. Fifth, if coercion has occurred, researchers will rarely be able to detect it, and so will be unaware how and to what degree their data might be biased or unethically collected.

- In their Methods' 'Participants and Procedures', the authors write that interviewers asked questions to determine the following information for *every* household (Burnham et al., 2006a):

1. A listing of every current household member by sex. This is a substantial list, as the authors report that the mean number of residents per household was 7.
2. A listing of every person who lived in the household on January 1, 2002.
3. A listing of births since that date.
4. A listing of deaths since that date.
5. Who moved into the household over the recall period.
6. Who moved out of the household over the recall period.
7. Based on this information, interviewers "confirmed that the reported inflow and exit of residents explained the differences in composition between the start and end of the recall period". In interviews with 1849 households, 1474 births were reported and 629 deaths were reported.

- In every household that reported a death, a series of additional questions were asked to ascertain the study's mortality data. With 629 deaths reported by 1849 households, interviewers needed to collect data on about 1 death for every three households. The authors report that the following questions were asked for every death reported (Burnham et al., 2006a):
 1. Had the decedent lived in the household continuously for 3 months before their death?
 2. "Additional probing" was done to establish the cause and circumstances of the death, taking into account family sensitivities. As a co-author of this study has correctly emphasized in other writing the importance of "the ethics of sensitive data collection" (Checchi & Roberts, 2005), this study's interview teams should be expected to understand the need to allow interviewees adequate time for describing this sensitive information as they saw fit, rather than cutting off descriptions due to time constraints.
 3. Interviewers asked to see a copy of the death certificate.
 4. The death certificate would need to be located and brought to the interviewers.
 5. The interviewers then recorded the presence of a death certificate.
 6. When there were differences between the household account of the cause of death and the cause recorded on the certificate, further discussions were carried out to establish the primary cause of death.

7. The authors report that interviewers asked to see death certificates for 545 of reported deaths, and that certificates were provided for 501 of these deaths.

How is it possible for an interview team to do all this in such a short time, reportedly as a matter of routine? The time factor is clearly a problem. Furthermore, a particular ethical issue is raised when the issue of short interview duration is coupled with the fact that the study's interviewers routinely fostered "word of mouth" to spread knowledge of their household interviews across the neighborhood: How could the teams control their interview environment in order to keep interviewees' responses private and confidential within such a short time, especially when the whole neighborhood was aware of who was being interviewed and why? In their description of the study, the authors repeatedly emphasize the need to protect the interview teams from danger (Burnham et al., 2006a, 2006b), which is indeed essential in a survey. But by carrying out rushed interviews publicized by word of mouth on a potentially highly sensitive topic, (the study's violent deaths include criminal deaths, some of which will inevitably have involved perpetrators or collaborators in the local community who were aware of interviewees' participation), this study easily may have endangered the interviewees. This is analogous to interviewers announcing to a community that they are doing a domestic violence survey, going door-to-door in sight of extended family and neighbors, and asking women to account for injuries suffered in the previous year while allowing husbands to sit in the room. Lack of confidentiality from the perpetrators of the very violence being studied increases interviewees' risk. Because interviewees are highly attuned to their own level

of risk, this can bias survey results as they seek to moderate their risk through how they answer survey questions. In the study by Burnham and colleagues, interviewees might have faced a variety of risks if their answers on the cause or perpetrator of death were not completely private: violent retribution from perpetrators in the community; general household or personal risk from being associated with certain political or sectarian allegiances; or criminal liability and shame to the household if another household member were responsible for the death. In general, when events of injury or death are already known to others and an interviewee is asked to report the perpetrator, the interviewee may attribute responsibility to a third party in order to neutralize potential risks from identifying the real perpetrator, thereby biasing the study's results.

Given the information that the authors have provided so far, there is an evident mismatch between the time required to include all necessary elements for ethical and valid epidemiological interviews and the markedly short time that was spent on the study's interviews. This suggests that either fewer interviews were actually completed than were reported or that all the interviews were completed but not in accordance with the published methodology and required ethical standards. This problem stands whether the mean time allocated per interview were 3 minutes, 6 minutes or 15 minutes.

In view of the disjuncture between the time spent interviewing and the number of tasks that would need to be covered, there is now strong reason to ask that the authors of this study provide the additional information needed to demonstrate their study's validity, for example, by providing the raw, anonymized data necessary to show how long

individual interviews lasted. Start and end times for dated interviews are usually recorded in epidemiological studies in order to monitor the feasibility of the method and the reliability and productivity of the interviewers. Data on which interviewers carried out each interview are easily coded to protect interviewer anonymity, and this allows studies to monitor for anomalies in interviewers' work or data and thus to ensure the quality of the data. In addition, the authors could clarify the matter by providing detailed, specific descriptions of how interviews were carried out and the time and logistics involved; something that the co-author, Professor Lafta, would be eminently qualified to do due to having supervised the interview teams in the field. Finally, it would be extremely helpful if others who have done household surveys in Iraq since the invasion could describe their experiences with carrying out interview methodology in the field. For example, Fafo-AIS, a Norwegian NGO, was closely involved in carrying out the UNDP Iraq Living Conditions Survey, a survey that like the authors' Iraq study was also epidemiological, cluster-sampled, and gathered conflict mortality data in Iraq's very difficult post-invasion field conditions (UN Development Programme, 2005). Fafo prepared the survey methodology and field manuals, conducted training and directly monitored the survey data that were regularly transmitted from the field. Input from these researchers on their experience of the time requirements and feasibility of surveying mortality in Iraq would be illuminating.

One way that the suffering of Iraq's populace from war can be acknowledged in the field of medical epidemiology is by producing accurate mortality findings. This can only be done if epidemiological techniques are applied properly at all levels of the

epidemiological survey. In view of the significant questions that remain unanswered about the feasibility of their study's methods as practiced at the level of field interviews, it is necessary that Burnham and his co-authors provide detailed, data-based evidence that all reported interviews were indeed carried out, and how this was done in a valid manner. In addition, they need to explain and to demonstrate to what degree their published methodology was adhered to or departed from across interviews, and to demonstrate convincingly that interviews were done in accordance with the standards of ethical research. If the authors choose not to provide this further analysis of their data, they should provide their raw data so that these aspects can be examined by others. Even in surveys on the sensitive and potentially risky subject of community violence, adequately anonymized data are expected to be sufficient for subsequent reanalysis and to be available for review. In the case of studies accepted for publication by the Lancet, all studies are expected to be able to provide their raw data.

An appropriate acknowledgement of the losses of Iraq's populace to armed conflict cannot be provided by the simple production of high mortality numbers. Accurate and valid quantification of conflict-related mortality does have its role and can wield substantial moral and scientific weight. But it must have a solid basis on good data and on research that systematically maintains an ethical relationship with the population being represented. Adequate standards of ethical and valid research conduct need to be maintained in the field for a study to be worth its effort, and certainly if the resulting findings are being presented as an academic study. If these standards are compromised

due to actual or perceived urgency, then the resulting limitations of the study need to be fully elucidated and recognized.

References

BBC News. October 30, 2006. Have Your Say: Lancet author answers your questions. (http://news.bbc.co.uk/1/hi/talking_point/6099020.stm) (Accessed October 31, 2006.)

Bohannon J. 20 October 2006. 'Iraqi death estimates called too high; methods faulted.' *Science* 314: 396-397.

Boseley Sarah. October 24, 2006. 'UK scientists attack Lancet study over death toll.' *The Guardian*: 17.

Burnham G, Lafta R, Doocy S, Roberts L. 2006a. 'Mortality after the 2003 invasion of Iraq: a cross-sectional cluster sample survey'. *The Lancet* 368:1421-1428.

Burnham G, Doocy S, Dzeng E, Lafta R, Roberts L. 2006b. The human cost of the war in Iraq: a mortality study, 2002-2006. John Hopkins Bloomberg School of Public Health and Al Mustansiriya University School of Medicine.

http://web.mit.edu/CIS/pdf/Human_Cost_of_War.pdf (accessed Oct 23, 2006).

Burnham G, & Roberts L. November 24, 2006. Letters: A debate over Iraqi death estimates. *Science* 314 (5803): 1241.
<http://www.sciencemag.org/cgi/content/full/314/5803/1241b> (accessed Nov 24, 2006).

CDC. 2006. Probability Proportional to Size Cluster Sampling. Step 3: Select elements within the cluster. <http://www.cdc.gov/descd/MiniModules/pps/page07.htm> (accessed Nov 24, 2006).

Checchi F, Roberts L. 2005. Interpreting and using mortality data in humanitarian emergencies: a primer for non-epidemiologists. Humanitarian Practice Network, Network Paper. Overseas Development Institute: London.

Gourley, Sean; Jukka-Pikka Onnela, Neil Johnson, Gesine Reinert & Michael Spagat. 2006. Conflict Mortality Surveys. Department of Economics, Royal Holloway College, University of London.
<http://www.rhul.ac.uk/economics/Research/conflict-analysis/iraq-mortality/Home.html>.
(accessed November 23, 2006).

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.

UN Development Programme. 2005. Iraq living conditions survey 2004.
<http://www.iq.undp.org/ILCS/overview.htm> (accessed Oct 23, 2006).