Number 52
September 2005

Humanitarian Practice Network
HPN
Managed by
Humanitarian Policy Group

Network Paper

Interpreting and using mortality data in humanitarian emergencies

A primer for non-epidemiologists

Commissioned and published by the Humanitarian Practice Network at ODI

Francesco Checchi and Les Roberts

In brief

• Mortality data, properly collected, interpreted and used, have much to contribute to the appropriateness and effectiveness of humanitarian action in emergencies, and to advocacy on behalf of populations in crises. Most actors involved in relief will one day be confronted by such data, but the different ways in which this information can be collected, and their potential pitfalls, are not yet common knowledge among non-epidemiologists.

• This Network Paper describes the practice and purpose of that branch of epidemiology concerned with population mortality. It sets out the key indicators used to express mortality data, different options for how to measure mortality rates and suggestions for how to assess, interpret and use mortality reports. The paper also discusses the politics of mortality figures.

• The paper's aim is to enable readers to critically interpret mortality study reports, and to understand how these are used (or misused) to formulate policy. The intended audience is therefore all humanitarian actors, policy-makers, the media and members of affected communities, who may be called upon to comment on or make use of mortality studies, regardless of their technical background.

About HPN
The Humanitarian Practice Network at the Overseas Development Institute is an independent forum where field workers, managers and policymakers in the humanitarian sector share information, analysis and experience. *The views and opinions expressed in HPN's publications do not necessarily state or reflect those of the Humanitarian Policy Group or the Overseas Development Institute.*

odi Britain's leading independent think-tank on international development and humanitarian issues

Overseas Development Institute
111 Westminster Bridge Road
London SE1 7JD
United Kingdom

Tel. +44 (0) 20 7922 0300
Fax. +44 (0) 20 7922 0399

HPN e-mail: hpn@odi.org.uk
HPN website: www.odihpn.org

Contents

Figures and tables

Chapter 1
Introduction

Humanitarian emergencies are chaotic, continuously evolving phenomena. Their magnitude and pace often overwhelm the capacity of individual relief workers or agencies to fully appreciate their true scope and evolution. Sector-oriented relief organisations tend to focus on specific perceived needs and may lack breadth of vision, just as organisations with oversight roles, such as governments or UN agencies, may sometimes lack sufficient depth. In the face of evident injustice and suffering, it is difficult to maintain objectivity. Depending on one's perspective, it may be professionally or emotionally more expedient to convince oneself and others that the situation of affected populations is either much better, or much worse, than it really is.

Years ago, one of the authors of this paper, returning from a long stay in Liberia with Médecins Sans Frontières (MSF), was asked by a Parisian taxi driver why MSF did nothing to help the Parisian working class who, he said, were all falling sick that winter; he then started listing members of his family who had been ill with the 'flu, coughs and arthritis. Suffering is indeed everywhere, but some populations experience extraordinary crises due to natural disasters, war, political repression, displacement, hunger and epidemic disease. The common denominator of these public health emergencies, and indeed their ultimate measure, is the spiritual and physical harm they inflict on individual human beings. The former type of harm is difficult to detect and quantify (although lately the importance of doing so is increasingly recognised).[1] As for physical harm, in humanitarian emergencies its most extreme form – death – takes on rudimentary mathematical connotations as an increase in mortality from levels considered 'normal' in non-crisis times. Information on mortality, and on its evolution over time, is strikingly eloquent, offering an immediately comprehensible, overarching view of the physical experience of affected populations over a given time interval. It is to a population in distress what vital signs are to a patient.

The starting premise of this paper is that the primary, most immediate goal of humanitarian relief is to prevent excess morbidity and mortality. Similarly, any excess mortality should lead to a reaction. In this respect, mortality is the prime indicator by which to assess the impact of a crisis, the magnitude of needs and the adequacy of the humanitarian response.

What is epidemiology?

Epidemiology can be defined as the study of the distribution of human diseases, and of factors influencing their frequency. The Parisian taxi driver, however caring a family man, is everything modern epidemiology tries not to be: he demonstrates no sense of relative measures; he does not distinguish between preventable and non-preventable morbidity, or between fatal and non-fatal diseases; he makes a claim of great distress based on no temporal comparison; and, worse still, he draws his sample from an unlucky set of people who happen to surround him, and then applies his findings to the general populace. The subdiscipline of epidemiology (and demography) that focuses on population mortality attempts to replace such subjective impressions and semi-quantitative guesses with objective, precise estimates of the human toll of an emergency and, if possible, to describe at least its proximate causes.

For those who practice this science, recent years offer much reason for encouragement. In response to several dramatic setbacks, the largely unregulated humanitarian sector has strengthened its capacity for quality and accountability, with initiatives such as the Sphere Project and the Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP). At the heart of accountability is hard evidence of the appropriateness and effectiveness of relief interventions. In many cases, only well-conducted epidemiological evaluations are capable of furnishing this evidence: thus, the primacy of data over hearsay seems to be taking a decisive hold. In particular, there is a broad consensus that mortality must be the ultimate measure of how an emergency is evolving.

But there is equally great cause for concern. Unlike chaos physics, mortality epidemiology is a remarkably accessible field (as we hope to show in this paper). As such, mortality data are extremely liable to misinterpretation and manipulation. Many would argue that recent years have seen the increasing use of relief as a tool for applying international political pressure or improving the image of occupying powers among the local population. In a context where the distinction between what is political and what is impartially humanitarian is vague and often confused, scientifically objective mortality reports can easily clash with political expectations. During the recent conflicts in Kosovo, Darfur and Iraq, contradictory versions of events, including reports of civilian deaths, were offered by opposing sides, anti- and pro-war groups or their international backers. Though only partly heeded, proper epidemiological mortality studies conducted in all three contexts helped to disqualify at least one of these versions of events, and to rectify another. Even in Niger, a relatively peaceful country, the nutritional crisis of 2005 was apparently becoming a political tool.[2] While the extent of the emergency was not yet fully clear as of August 2005, Niger's president defended his view that there were no major problems by pointing out that people on the street 'looked well-fed'.

Our Parisian taxi driver might have found this conclusion flawless, but it bears no resemblance to evidence-based needs assessment. Our belief is that mortality data, properly collected, interpreted and used, have much to contribute to the appropriateness and effectiveness of humanitarian action in emergencies, and to advocacy on behalf of populations in crises. Most actors involved in relief will one day be confronted by such data, but the different ways in which this information can be collected, and their potential pitfalls, are not yet common knowledge among non-epidemiologists. This paper addresses some of these issues, in the conviction that greater awareness of the science behind mortality figures will help to maximise their operational relevance and use, whilst reducing the likelihood of misinterpretation and manipulation.

This paper is not intended to be a course on how to measure mortality in emergencies, nor does it provide detailed derivations of the statistical theory underpinning this measurement. Nor does it present a systematic review of past studies of mortality in emergencies. Where appropriate, we refer the interested reader to further, more in-depth sources of information.[3] Rather, the paper is written from the standpoint of the end-user of mortality data. We present key indicators used to express these data, different options for how to measure mortality rates and suggestions for how to assess, interpret and use mortality reports. There is also a discussion of the politics of mortality figures. The paper cannot be exhaustive: its focus is on enabling readers to critically interpret mortality study reports, and to understand how these are used (or misused) to formulate policy. The intended audience is therefore all humanitarian actors, policy-makers, the media and members of affected communities, who may be called upon to comment on or make use of mortality studies, regardless of their technical background. It should be noted here that epidemiology is an evolving science: new methods not covered in this paper may well be introduced in coming years.

Key concepts and terms

Mortality indicators, and specifically mortality rates, are expressed and calculated in different ways. This section is meant to provide readers with a common lexicon. We discuss units of measurement, and mention other common mortality indicators. These terms and concepts will then be used throughout the paper. A more extensive list of terms is contained in the Glossary (see pp. 35–36).

Mortality rates

In August 2005, a BBC report stated that, according to doctors, 15 children every week were dying from malnutrition in the Maradi region of Niger.[4] While the death of any one child from hunger is a horrible event, one can hardly be satisfied with this single figure. How many children live in the Maradi region? How many die every week usually, and does 15 constitute a significant increase? Apart from children dying from malnutrition, are others dying from conditions associated with it, like diarrhoea? Are they counted in this figure? How were these deaths ascertained? This example illustrates that an absolute number of deaths, taken in isolation, is very difficult to interpret. It becomes much more meaningful when it is related to a clearly specified population over a precise period of time, and then compared to expected mortality patterns in that population when no crisis is occurring.

To do this, we usually present mortality in terms of a rate. The term 'rate' is (mis)used in many contexts; strictly speaking, it should only be applied to express the *frequency* with which events occur as time goes on (i.e. rates should always refer to a unit of time). A **mortality rate** (MR) therefore expresses the number of events (deaths) that occur in a population of known size that is at risk for the death event during a specific period of time (usually called the **recall period**). We speak of populations *at risk* because mortality is measured in a clearly defined population (such as 'all displaced persons currently living in a camp within Gulu District, northern Uganda'). Being present within that population during the recall period (say 'January to July 2005') defines one's exposure to the risk of death, as far as the survey is concerned.

There are three ways of expressing MRs: the first (deaths per person-time) is formal and more rigorous. The second (deaths per persons per time) is a more intuitive and, in most cases, sufficiently accurate simplification, and we discuss this first. The third is deaths that occur within a population by a certain age: indicators using this concept are not often cited in emergencies (although they are by no means irrelevant), and are not discussed here.*

The simplified expression of mortality rates. The simplified expression of mortality rates (deaths per persons per unit time) is described in Box 1. This paper uses this method of expressing MRs wherever possible. It assumes that each individual, with the exception of those who die and those who are born, spends the entire recall period within the population of interest, i.e. is present from the start to the end. Given this assumption, what should be considered the population at risk during the period? Neither the population present at the beginning nor that present at the end are satisfactory solutions, since not all who were there at the start made it to the end (because of deaths during the period), and not all present at the end were there throughout (because of births). As a compromise, we estimate a **mid-period** (or **mid-point**) **population** by (i) taking the population at the end; (ii) adding one half of the deaths during the period; and (iii) subtracting one half of the births, assuming information was collected on these (if not, we simply take the population at the end and add one half of the deaths).

* Infant and child mortality rates refer respectively to deaths under one year of age and deaths under five years of age, out of 1,000 live births during a specified year. The maternal mortality ratio is the number of women dying from pregnancy-related causes out of 100,000 live births during a specified year. Methods to measure these indicators are very different from those used to measure mortality rates in emergencies. Maternal mortality ratios can, however, reach very high levels (exceeding 1,000/100,000) in crisis-stricken populations, in great part due to lack of emergency obstetric care.

Box 1

Simplified mortality rate expression

$$\text{Mortality rate} = \frac{\text{total deaths during period}}{\text{mid-period population at risk x duration of period}}$$

Example:

duration of period: 120 days
population at end: 18,300

births during period: 360
deaths during this period: 445

mid-period population at risk: 18,300 + 0.5(445) – 0.5(360) = 18,343
unit of expression: per 10,000 people per day

MR = [445/(18,343 x 120 days)] x 10,000 = 2.02 deaths per 10,000 per day

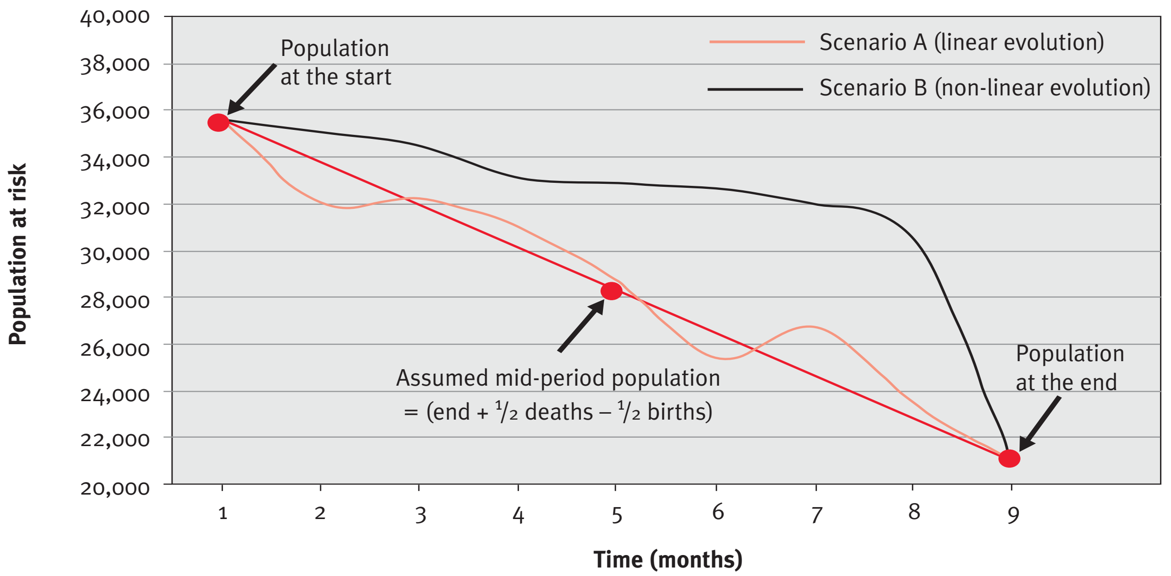
Interpretation
In this population, on average about 2 people out of 10,000 died every day during the 120-day period analysed.

The simplified expression of mortality rates seems adequate in most instances, but one can think of circumstances where this would not be so (see Figure 1). In the simplified MR expression, the way a mid-period population is estimated implies an assumption that exactly as many people leave/join the population before the middle of the period as do after it. This assumption holds in Scenario A in Figure 1, where a population declines more-or-less linearly through time. It is, however, quite inaccurate in Scenario B, where a population declines slowly at first, only to undergo a steep drop at the very end of the period (possibly due to high mortality or the sudden departure of many family members). Scenarios A and B have the same starting and ending population sizes, but more person-time is spent at risk in Scenario B, since individuals are, on average, present for longer in the population than in Scenario A. Using the simplified MR expression in Scenario B would result in a serious over-estimation of mortality, since the population denominator would be artificially small.

The formal (person-time) expression of mortality rates. In reality, not all individuals are likely to be present in a given population (i.e. be exposed to the risk of dying within that

Figure 1

An illustration of the potential problems with the estimation of population at risk in the simplified mortality rate expression (deaths per persons per unit time)



population) throughout the entire recall period: some of those present at the beginning will migrate away, die or disappear; others will join the population at some point after the start of the period, due to birth, immigration/further displacement influx or family reunification; yet others might join the population after the start of the period and leave it before the end (for example babies who die soon after birth). Furthermore, certain events, like departures or deaths, may be concentrated in a specific portion of the period, as we saw above. In short, not all individuals spend an equal amount of time (or **person-time**) in the population. The formal expression of MRs takes this into account, and consists of deaths per total person-time at risk spent by the population during the period of interest (for instance three deaths per 10,000 person-days). When total person-time is computed, each individual counts in the denominator of the MR expression only for the portion of the time period of interest that he/she actually spent in the population studied (Box 2).

The two expressions (simplified and formal) are practically interchangeable. It should be stressed that person-time MR and its simplified version express exactly the same quantity, and in most cases they are statistically indistinguishable from each other. The expression 'deaths per person-days' is equivalent to 'deaths per persons per day'. Calculating person-time is, however, not just an academic exercise: this approach is, in fact, indispensable when the recall period refers more to a situation rather than a precise length of time – such as, for example, when we wish to distinguish the MR during a refugee population's stay in a host country camp from the MR during their flight to the camp. Probably not all refugees left and arrived at the same time. The recall period is thus different for each individual, and working with person-time is the only feasible approach in this case (this is discussed in more depth in Chapter 3).

Box 2
Calculating the total person-time

This example of a total person-time calculation is based on a population of ten individuals and a recall period of 50 days:

- A, B and C are present from the beginning to the end of the period (thus, they count for 50 days at risk each).
- D is present at the beginning but dies on day 38 (38 days).
- E is present at the beginning but dies on day 11 (11 days).
- F is present at the beginning but leaves on day 42 (42 days).
- G joins the population on day 23 and remains until the end (27 days).
- H is born in the population on day 33 and remains until the end (17 days).
- I is born in the population on day 33 and dies on day 37 (4 days).
- J joins the population on day 9 and leaves on day 28 (19 days).

The population at the beginning is six (A, B, C, D, E, F). The population at the end is five (A, B, C, G, H). The total person-time at risk is (50x3)+38+11+42+27+17+4+19 = 308 person-days (and the average time spent at risk is 308/10 = 30.8 days).

The above are considered **crude mortality rates** (CMR).* They include all age groups and all causes of death.

Different units, same mathematics. Readers may find MRs expressed in different units, such as deaths per 1,000 people per month (or per 1,000 person-months) and deaths per 1,000 people per year (or per 1,000 person-years). In acute emergencies, however, it makes more sense to use a short time interval, and the expression **deaths per 10,000 per day** is most commonly employed. As a reference, 1 death per 10,000 per day = 3 deaths per 1,000 per month = 36 deaths per 1,000 per year.

Mortality rates in sub-groups or sub-periods. Mortality rates can refer to sub-groups within the population (Table 1), for example recently displaced refugees versus long-time residents, or certain age groups. **Under-five mortality rate**, or U5MR, is the most commonly presented age-specific MR, since it is generally indicative of the general state of health of a population. In these cases, both the numerator and the denominator change (for example the number of under-five deaths/population under five years). MR in specific sub-periods can also be calculated, for individual months, during an armed incursion or in an epidemic, for example. The implications of these sub-analyses are discussed in Chapter 4.

Indicators for mortality due to specific causes. Cause-specific MRs express mortality from a particular cause, such as measles or violence: here only the numerator changes, assuming that the entire population is at risk of the disease/event in question (this would not hold true for diseases that are sex-specific, like cervical cancer). **Proportionate mortality** expresses the portion of all deaths due to a specific cause; note that this is not a rate, but simply a proportion. This is sometimes represented by a pie chart and is, of course, related to cause-specific MR; if, for example, the all-cause crude mortality rate is 2.3 per 10,000 per day and 54% of deaths are due to epidemic

* This denomination is to distinguish them from age-adjusted rates. Because older people are intrinsically more at risk of dying, a population that contains many elderly people (say, Norwegians) may appear to have higher mortality than one whose members are on average younger (say, Algerians), despite enjoying better health care. A mathematical procedure known as age adjustment is necessary to meaningfully compare mortality in populations with different age structures. This adjustment is rarely needed in humanitarian emergencies, since any comparisons are usually made within the same population (for example, before and after displacement), i.e. based on a roughly similar age structure. Comparisons among different rural African populations are also relatively unaffected by this problem, since age structures are fairly homogenous throughout the continent.

Table 1: Common indicators of population mortality in emergencies

Indicator	Simplified formula	Common applications
Crude mortality rate (CMR)	Deaths/(population at risk x period of time)	Always presented
Age-specific mortality rate	Deaths in age group/(population in age group at risk x period of time for those within the age range)	Under-five mortality rate (U5MR)
Group-specific mortality rate	Deaths in sub-group/(sub-group population at risk x period of time)	MR among males/females; among unaccompanied children; among displaced persons vs. residents; in a special ethnic group
Period-specific mortality rate	Deaths during sub-period/(population at risk during sub-period x duration of sub-period)	Monthly MR, MR during epidemic period, MR before/after displacement
Cause-specific mortality rate	Deaths due to given cause/(population at risk x period of time)	MR due to violence; MR due to disease causing epidemic
Proportionate mortality	Deaths due to given cause/total deaths (note: this is *not* a rate)	Proportion of deaths due to violence; proportion due to disease causing epidemic
Case-fatality ratio (or rate) or CFR	Deaths due to given cause (disease)/ total cases of given disease	CFR of cholera, measles, severe malaria; important during epidemics
Excess mortality rate (total number of excess deaths)	Observed MR – expected non-crisis MR (x population at risk x period of time)	See Chapter 4

cholera, the cholera-specific mortality rate must be 2.3 per 10,000 per day x 0.54, i.e. 1.2 per 10,000 per day). The **case-fatality ratio** (or **rate**) expresses the proportion of cases of a particular disease that result in death, and is thus expressed as a percentage (epidemiologists argue about whether this is a rate, a ratio or a proportion).

Absolute and excess mortality

Any MR can, of course, easily be converted to the **total number of deaths during the period** it refers to. For example, in a population of 40,000 experiencing a CMR of 1.5 per 10,000 per day over a period of 60 days, 360 total deaths (40,000 people x [1.5 deaths/10,000 people] x 60 days) would be expected to occur by the end of the period. Sometimes, the cumulative **percentage of the population who died** during a given period is provided as an alternative indicator. In the above example, this would be 0.9% (360/40,000). MR estimates are often obtained through sample surveys: in Chapter 4 we discuss the appropriateness of calculating absolute death tolls based on these.

MRs describe the frequency with which deaths are occurring in a given population over a given time. If these are higher than the expected (baseline) MR in non-crisis conditions in that population, we can say that the difference between observed crisis and expected non-crisis MRs represents **excess mortality**, i.e. the mortality attributable to the crisis, above and beyond deaths that would have occurred in normal conditions. Exactly as above, we can apply an excess MR to the population and period it refers to, and thus obtain an **absolute number of excess deaths**.

Chapter 2
Applications of mortality data

Determining the nature and scale of the crisis

Crude and under-5 mortality rates are key indicators to evaluate the magnitude of a crisis, and a doubling of non-crisis (baseline) mortality is taken to define an emergency situation. However, different views exist on whether absolute or context-specific thresholds should be used. Different typologies of crisis result in different mortality patterns. In humanitarian emergencies, how high the mortality rate is (scale) and what people die from (nature) are key starting points for planning the size and programmatic focus of the humanitarian response. Unfortunately, unlike the assessment of food security, mortality epidemiology can only detect the occurrence of a crisis after health conditions have deteriorated.

CMR or U5MR are key indicators to define an emergency. Baseline, non-crisis CMRs in most of Sub-Saharan Africa are in the range 0.3–0.6 per 10,000 per day, with a probable current average of 0.44.[5] Based on this, in 1990 Toole and Waldman suggested an approximate doubling of CMR (to 1 per 10,000 per day) as a useful threshold for formally declaring an emergency, at least from a health standpoint.[6] This simple threshold has since been adopted widely and incorporated into various humanitarian guidelines. U5MR is usually approximately twice the CMR; hence, an U5MR of 2 per 10,000 per day or more can also be considered indicative of an emergency. The classification has since been refined by UNHCR, to distinguish between situations that are serious, and situations that are out of control.[7]

By contrast, the Sphere standards, recognising that baseline mortality is context-specific, specify that emergency thresholds should reflect a doubling of *local* pre-crisis CMR or U5MR. This approach raises a fundamental question of humanitarian ethics. Adopting different baselines (for example 0.25 per 10,000 per day in Eastern European countries and 1.1 per 10,000 per day in Darfur) is clearly useful to distinguish mild alterations in mortality from true crises that require an urgent intervention.[8] On the other hand, their strict application would mean that threshold mortality in Darfur must be five to six times higher than in Europe before emergency relief is organised, further exacerbating the already serious aid differential between African and other populations in crisis. Humanitarianism places the same value on human life irrespective of context, and seeks to diminish absolute, not relative, suffering: assuming the above baselines, a CMR of 0.5 per 10,000 per day in an Eastern European country or 1.4 per 10,000 per day in Darfur would result in identical excess mortality (+0.3 per 10,000 per day), but Darfur's CMR would not be classified as constituting an emergency. A further complication of the Sphere approach is that, in many cases, it is extremely difficult to define the start of a crisis. The question of whether humanitarian emergencies should be defined and quantified based on relative benchmarks probably merits further discussion (see also Chapters 4 and 6).

Many different classifications of emergencies have been attempted, and this paper does not propose new ones. Roughly speaking, in terms of patterns of mortality three types of crisis can be delineated:

Table 2: Mortality thresholds commonly used to define emergency situations[9]

Agencies	Assumed baseline	Emergency thresholds
Centers for Disease Control, Médecins Sans Frontières Epicentre, Academia	Fixed at: CMR: 0.5 per 10,000 per day U5MR: 1 per 10,000 per day	Emergency if: CMR: ≥ 1 per 10,000 per day or U5MR: ≥ 2 per 10,000 per day
UNHCR	Fixed at: CMR: 0.5 per 10,000 per day U5MR: 1 per 10,000 per day	CMR > 1 per 10,000 per day: 'very serious' CMR > 2 per 10,000 per day: 'out of control' CMR > 5 per 10,000 per day: 'major catastrophe' (double for U5MR thresholds)
Sphere Project Note: if baseline is not known, Sphere goal is CMR <1 per 10,000 per day	Context-specific CMR (U5MR): Sub-Saharan Africa: 0.44 (1.14) Latin America: 0.16 (0.19) South Asia: 0.25 (0.59) Eastern Europe, Former Soviet Union: 0.30 (0.20)	Emergency if CMR (U5MR): Sub-Saharan Africa: 0.9 (2.3) Latin America: 0.3 (0.4) South Asia: 0.5 (1.2) Eastern Europe, former Soviet Union: 0.6 (0.4)

- **Sudden natural disasters**, in which most mortality occurs as a result of the mechanical force of the elements or resulting injuries, and is therefore concentrated in a period of hours or days; further peaks in mortality can, however, occur weeks after the disaster as a result of hygiene- or flooding-related epidemics (for example, cholera after floods in Mozambique in 2000).
- **Acute emergencies** due to large-scale armed conflict and/or rapid displacement; where these result in relocation of the population to camps, CMR is known to fall progressively as a result of better protection and the arrival of humanitarian aid, although neglect of vaccination and disease control efforts can lead to devastating epidemics of diarrhoeal diseases or measles.
- **Slowly evolving, chronic or intermittent emergencies** in which mortality may increase slowly over the course of months and years from near-normal levels, as a result of the progressive breakdown of health infrastructures, loss of livelihoods, isolation from international aid and nutritional problems, or in which CMR can display regular peaks as a result of poor harvests, displacement waves, low-level conflict or epidemics affecting a chronically vulnerable population.

Sudden natural disasters are of such brief duration that calculating weekly or daily CMR may not make much sense. Rather, one could present mortality figures as the proportion of people dying out of all those exposed to the disaster. Examples of CMR evolution in an acute emergency (Goma, Zaire, 1994) and in a slowly evolving crisis (Angola, 2002) are presented in Figure 2.[10] Note the difference in scale and time units.

Chronic or intermittent crises are almost ubiquitous in countries affected by prolonged conflict (for example, southern Sudan, the Kivu regions of the Democratic Republic of Congo, northern Uganda, Burundi and Somalia), and sometimes last well beyond the conflict's end, as is the case in Ethiopia's southern regions today. In these settings, it can be difficult to establish a proper baseline, since the population has probably endured 'abnormal' mortality rates for years. Acute emergencies are often superimposed on more chronic, region-wide crises.

Although acute emergencies and natural disasters can be associated with dramatic peaks in mortality (see Table 3), they do not necessarily cause more deaths than other neglected, low-level emergencies which rarely make international headlines. During acute emergencies and natural disasters, dramatic peaks in mortality are observed, and this is often what makes international headlines. By contrast, other neglected, slowly-evolving crises do not usually display such peaks. Their duration and geographic extent may nonetheless make them just as deadly. For example, during the dramatic but brief flight of 400,000 Kurdish refugees from northern Iraq in March–May 1991, CMR peaked at 10.4 per 10,000 per day, and an estimated 6,200 people perished in excess of baseline non-crisis mortality.[11] By contrast, starting from a baseline CMR of 0.44 per 10,000 per day, an average CMR of 1.54 per 10,000 per day (i.e. excess mortality 1.1 per 10,000 per day) during the first seven months of 2005 in a population of 1.3 million northern Ugandan IDPs resulted in some 28,000 excess deaths.[12] Thus, the impact of an elevated CMR depends not only on its magnitude, but also on its duration, and on the size of the population experiencing it.

Figure 2

Illustration of the typical evolution of mortality in two types of crisis

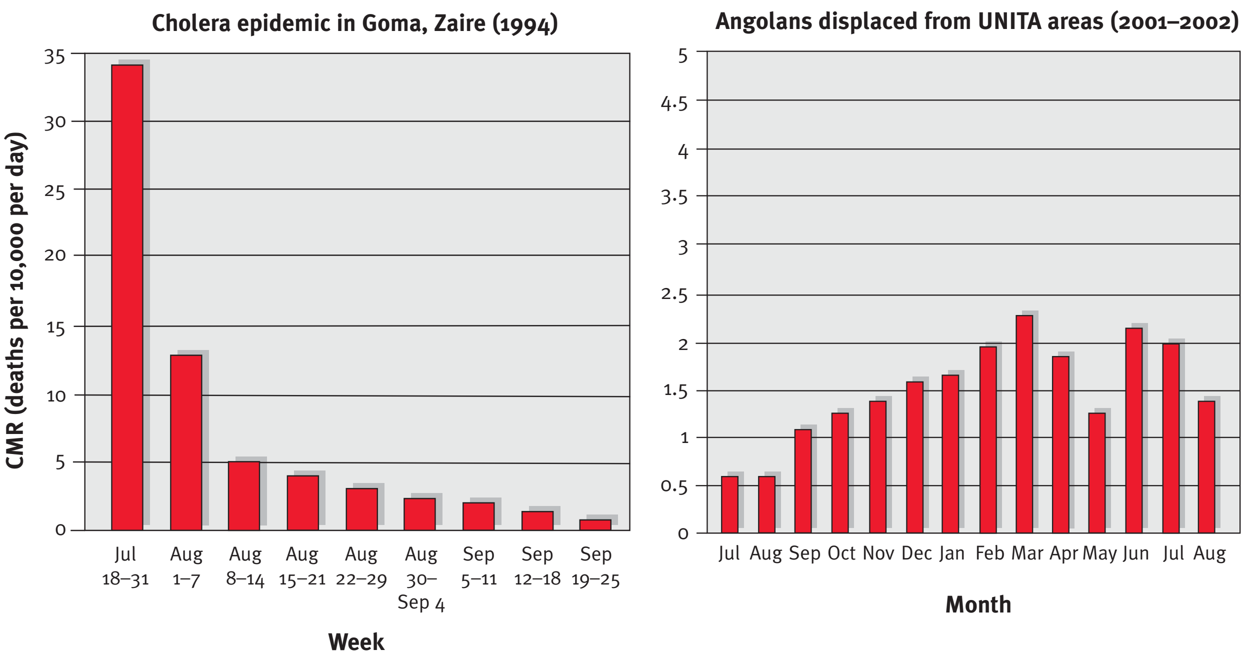


Table 3: Crude mortality rates in selected recent acute emergencies[13]

Context (year)	CMR (deaths per 10,000 per day)
Famine-affected communities in Baidoa, Somalia (1992)	16.8
Malnutrition and diarrhoeal disease epidemics among Rwandan Hutu refugees in Goma area, Zaire (1994)	34.1 to 54.5
Population under armed siege in Tubmanburg, Liberia (1996)	14.3
Famine and conflict-affected populations in Bahr el Ghazal, southern Sudan (1998)	9.2 to 26.1
Famine in Gode, Ethiopia (2000)	3.2
Famine and repeated displacement, Angolan IDPs in UNITA areas (2002)	2.3 to 3.6
Armed attacks against civilians in West Darfur, Sudan (2003–2004)	5.9 to 9.5

Monitoring the effectiveness of humanitarian relief

Whether or not mortality is established at the outset of a crisis, follow-up mortality studies are crucial to monitor trends and evaluate relief interventions. Individual programmes should probably not be judged on the basis of mortality data. CMR and U5MR, however, are key impact indicators for the entire relief operation. In addition, cause-of-death data can provide invaluable insights as to what services need to be bolstered, and are a key quality-control measure for health-related programmes.

Well-run relief programmes should be needs-driven, should strive to achieve certain standards and should produce indicators for monitoring and evaluation.[14] In addition to reporting on indicators of process (default rates from supplementary feeding programmes; percent of births attended by a clinician) and output (coverage of food distributions; litres of water provided per person per day), more sophisticated, quality-assured programmes should also be able to generate outcome data, such as malnutrition prevalence or the incidence of diarrhoeal diseases. In an emergency, any concrete impact of the humanitarian intervention must be reflected in decreased mortality (or alternatively the prevention of some excess mortality). Indeed, the primary goal of any comprehensive humanitarian programme should be to reduce mortality rates to at least the pre-crisis level. Mortality data can and should be used to evaluate the entire relief effort, viewed as an integrated system.

On the other hand, CMR and U5MR alone do not usually help to clearly identify specific gaps in the humanitarian response. Impact in terms of lives saved can be difficult to determine for individual and sector-specific programmes, unless the occurrence of a specific health event, for example a deadly measles outbreak, clearly points to specific weak areas in the humanitarian response, in this case vaccination, as in Gode, Ethiopia, in 2000.[15] For most causes of death, a reduction in mortality is the product of multi-sectoral work: for example, a decreased MR from acute respiratory infections can be a result of improved shelter, greater access to quality outpatient and inpatient care, better nutrition and higher income.

Similarly, based on causes of death it is difficult to establish which programmes are responsible for the least impact on mortality: reductions in MR due to individual programmes are relatively small and can be very difficult to demonstrate with any statistical precision. This means that it would be hard to set MR reduction targets for any individual programme, since in many cases several agencies are contributing to the same sector: one agency may be in charge of primary health care, for example, while another may operate in-patient facilities. The success of one programme usually depends on the success of related ones. In general, the more focused an intervention is on one deadly disease, such as vaccination to stop meningitis or measles outbreaks, the more easily success can be shown by mortality data.

Ideally, establishing a starting point MR (i.e. during the first weeks of the crisis) will help in planning an adequate response and judging progress. For example, given a starting point CMR of 10 per 10,000 per day, decreasing this to 0.5 per 10,000 per day versus 2 per 10,000 per day are obviously not equivalent tasks. If no such starting measure is available, mortality should still be monitored: the target must not change as a function of the problem's magnitude – rather, it must always be to bring mortality down to pre-crisis levels as soon as possible. Let circumstances, such as inaccessibility or insurmountable logistics problems, then temper our condemnation of any failure to meet that target.

In short, the measurement of mortality is an essential component of any effective public health intervention during health emergencies, from advocacy and planning to programme monitoring. Conducting a relief programme without any evidence of the extent and causes of mortality, or how these evolve over time, may be inefficient, not cost-effective and, ultimately, ethically questionable. No responsible physician would forget to check whether his or her patient was alive at the end of a procedure. The parallel with the health status of an entire population does not seem too daring. Failure to collect data on mortality rates and causes of death can be attributed to several factors, including the sub-division of health sector responsibilities among several agencies, a lack of skills and training among health staff, a reliance on incomplete surveillance mechanisms and a wish by belligerents or donor nations not to make the data available for public scrutiny.

Advocating for action

Mortality studies have occasionally played a prominent role in attracting aid and international political interest to a crisis. They can also serve to document the direct and indirect impact of war and population displacement.

The immediacy and alarming nature of mortality figures can, on occasion, have a profound impact (for better or worse) on the international response to a crisis. In Baidoa, Somalia, a 1992 survey measured a CMR of 16.8 per 10,000 per day, and estimated that 75% of under-fives had died during a seven-month period.[16] These results were an important contributing factor in the US administration's decision to send an intervention force.[17] In the DRC, repeated country-wide surveys have put excess mortality since the start of the conflict in 1998 at 3.8 million.[18] Publication of the first of these surveys in 2000, widely reported in the international press, was associated with a doubling of humanitarian aid to DRC.[19]

On the other hand, a survey in Iraq put at 100,000 the death toll from violence since the end of the 2003 invasion, and suggested convincingly that the tactics being used by Coalition forces had resulted in heavy civilian casualties.[20] Yet almost a year later no credible evidence existed that the occupying forces were doing any better at protecting the lives of civilians, as required under the Geneva Conventions. In West and South Darfur, repeated surveys in 2004 highlighted widespread killing of civilians, yet as of August 2005 these actions continued.[21] In northern Uganda, possibly the world's most neglected crisis in 2005, repeated reports of high CMR have not encouraged international attention to that conflict, and funding requests for relief for a displaced population of up to two million have gone largely unmet.[22] A survey conducted in July 2005, almost 20 years after the start of conflict, estimated that 1,000 excess deaths per week were occurring in the three most affected districts, belying impressions that the war was abating by showing a far higher number of violent deaths and abductions than reported in the media.[23]

At the very least, epidemiological documentation of past mortality directly or indirectly due to violence and war can serve as a historical record for future generations, in the hope that accumulating overwhelming evidence on the public health impact of all armed conflict, however technologically sophisticated, will stimulate the peaceful resolution of conflicts – as well as improving humanitarian practice and fostering greater respect for humanitarian law.[24]

This chapter has described the different goals of mortality studies – crisis assessment, advocacy and impact monitoring. These are not necessarily irreconcilable among themselves. They may, however, imply substantially different choices of methodology and analysis period, as discussed in the next chapter.

Chapter 3
Overview of methods to measure mortality

What is the objective?

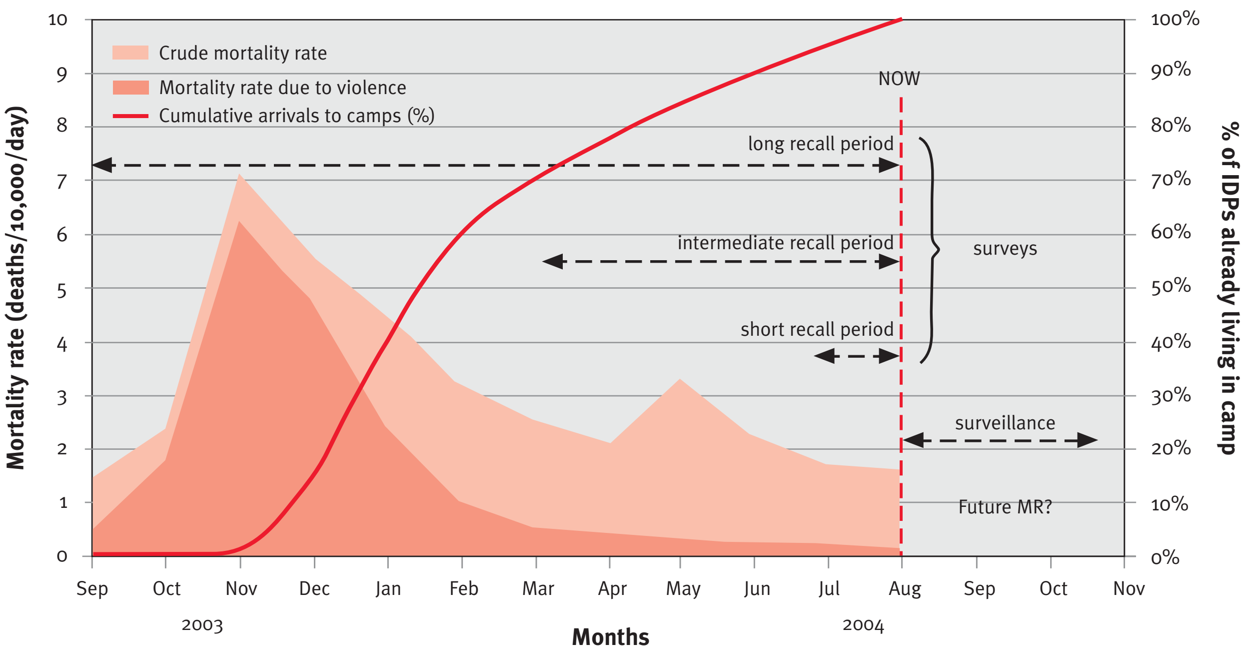
Depending on what mortality information is needed or missing, data collection will be either retrospective or prospective, and will analyse a long or a short period. Carefully defining from the outset what information on mortality is required often determines whether the findings will be of any use. This is illustrated in Figure 3. If information on past mortality is missing (it usually is), **retrospective surveys** are necessary. By surveys we mean discrete data-collection exercises carried out at a specified time, usually on a representative sample of the population and using a questionnaire to systematically collect certain quantitative variables. To gather mortality data on an ongoing basis, from the present into the future, **prospective surveillance** is the most obvious solution. Surveillance implies ongoing, systematic recording, analysis and interpretation of health data. If data must be gathered retrospectively, but the objective is to measure very recent mortality so as to inform the immediate response, a short recall period of around one or two months is appropriate.

Data in Figure 3 are approximate, and should not be quoted. They aim to illustrate broadly the evolution of mortality observed in several displaced populations living in camps in 2004. CMR peaks in November and December 2003, overwhelmingly as a result of violent attacks on villages. Survivors increasingly move to camps (shown by the red dotted line), where security is somewhat improved. Malnutrition and very poor camp conditions maintain CMR well above the emergency threshold in 2004, although the proportionate mortality attributable to violence decreases.

No data on past mortality are available in August 2004. A retrospective survey with a long recall period would be needed to document the evolution of mortality and violence since the onset of the crisis. An intermediate recall period would capture some of the violence, but not enough to demonstrate its true extent; the average CMR from this survey would also reflect a peak due to diarrhoea and measles epidemics reported around May in some camps. A short recall period would give a good idea of very recent trends, excluding 'noise' due to these earlier epidemics; however, the very short time period investigated would result in a wider confidence interval around the estimate.

The long recall period option would represent person-time spent in villages as well as in the camps (thus, the questionnaire would have to note dates of deaths and arrivals to camps). The intermediate and short recall period options would essentially represent conditions in the camps. Beyond August 2004, establishment of prospective surveillance would help to monitor CMR trends in real time. If the objective is to record the mortality experience of a population over as much of the crisis as possible (with particular reference to excess mortality due to violence and war), then a long recall period must usually be investigated. Because households

Figure 3
Choices for retrospective and prospective estimation of mortality in Darfur, Sudan, 2004



can go through significant changes over many months of crisis (displacement, births and family reunifications, disappearances and departures), a relatively complex questionnaire will be required, and the survey might have to distinguish mortality according to whether it occurred before or after important changes in households' status, such as relocation to a camp. Thinking in terms of person-time will be crucial to estimate this.

The average MR from surveys with long recall periods will reflect a mixture of conditions, and may mask an undetected or unquantifiable phenomenon, such as a measles epidemic, making it difficult to draw conclusions either about very recent mortality or about the overall impact of the crisis. Careful interviewing about when deaths occurred can partly remedy this; however, breaking down the findings into several periods (for example months) will result in wider confidence intervals for each period (see Chapter 4), since less person-time is sampled than for the entire recall period.

The degree to which causes of death are investigated is also affected by the survey objective. A survey designed to document past acts of war will try to classify the type of violent death (physical blow, bullet wound, bomb explosion). A survey oriented more towards public health intervention will attempt to rank leading causes of death (measles, diarrhoea) so as to highlight immediate public health priorities. Obviously, much depends on what information is already available: for example, well-collected health centre information on proportionate mortality could already delineate trends in causes of death; mortality surveys would add to this by quantifying death tolls in the community.

Two key measurement issues

Here we discuss different types of bias (non-sampling error) that can affect the validity of mortality data, and explain how sample size and other factors affect the precision (sampling error) of a mortality estimate (its confidence interval).

Epidemiology confronts two fundamental problems: **bias**, or **systematic/non-sampling error**; and **imprecision**, or **non-systematic/sampling error**. The measurement of mortality in emergencies is heavily affected by both, mainly since (i) objective information on deaths (such as centrally stored death certificates, or clearly marked graves) is usually not available, with consequent reliance on oral reporting by the family of the deceased; and (ii) populations are often far too large to be surveyed exhaustively, making some form of sampling necessary.

About bias (non-sampling error)

Preventing bias means ensuring that the data generated truly reflect the situation in the population studied, and are not distorted by flaws in the way the information is collected. These flaws can usually not be corrected afterwards, and are to some extent always present. No serious mortality report fails to acknowledge them, or to discuss their likely effect on the findings' validity. Indeed, potential for bias abounds in mortality epidemiology, and it is worthwhile mentioning here a few common sources of this (note that the denominations below are not all canonical epidemiological terms):

- **Household selection bias.** If only a sample of households is drawn from the population, is this sample truly representative of other households? Instances in which this is likely not to be the case include:
  - when the sample is not random, i.e. when not every household in the population has an equal chance of being included in the mortality study (perhaps only the most accessible communities are sampled due to security concerns)
  - when certain types of household are systematically left out after being selected (perhaps no one is home to provide information, and the surveyors make no attempt to trace or revisit absent household members; it may well be that unattended households have on average a different mortality experience).
- **Household size reporting bias.** Mortality surveys can easily be misinterpreted by the population as registration exercises which will lead to some relief goods being distributed to them. This can lead to inflation of true household size and, because the denominator of the mortality expression becomes artificially large, to an under-estimation of the mortality rate.
- **Event recall bias.** This bias applies to retrospective mortality studies only. It is postulated that, if the recall period that respondents are asked about is exceptionally long (more than one year), under-reporting of deaths, or erroneous reporting of their date, can occur due to forgetfulness; alternatively, certain kinds of death, such as violent ones, are remembered more vividly. These perceived constraints lead surveys to limit the duration of the recall period, usually to no more than one and a half years. The true risk of recall bias is, however, hard to gauge, probably varies widely between settings, and may be over-emphasised. It seems reasonable to presume that, if families in industrialised countries can easily provide details about past deaths, people affected by humanitarian emergencies in developing countries should be just as capable of recalling these events accurately, especially when assisted by a calendar of locally significant events.
- **Event reporting bias.** Aside from recall, there may be several reasons why respondents (or interviewers – see Chapter 5) will tend to over- or under-report deaths. If the household has strong links with an armed group, it may under-report violent deaths of family members belonging to that group. In some cultures, neonatal deaths may be perceived differently from other deaths. There can also be over-reporting if respondents perceive that this will increase their chances of obtaining aid, or multiple reporting of the same death (two neighbouring households may be part of the same extended family, and could both report an individual's death as taking place within their household).
- **Survival bias.** If the survey is retrospective, the only households which can be interviewed are those for

whom at least one member survives to the time of the survey. Some households, however, may simply have disappeared, either because all members died, or because the lone survivors, such as orphans, joined a different household. Because these extinguished households cannot be interviewed, mortality is underestimated. The extent of survival bias depends on four factors: size of households (small family nuclei are more likely to disappear), duration of the recall period, mortality rate (the higher the rate, the greater the survival bias), and clustering of the risk of death (for example if a missile hit a particular neighbourhood).

About imprecision (sampling error)

A mortality result may be satisfyingly unbiased, but disappointingly imprecise if it is derived from an inadequate sample. Imprecision refers to the degree of uncertainty (approximation) around the study estimate.

Sample-based surveys should all report a best (or **point**) estimate of mortality: 3.2 per 10,000 per day, for example. They should also report the degree of uncertainty associated with this estimate, given in the form of a **confidence interval** (CI, consisting of a lower and upper value, for example 1.8 to 5.6 per 10,000 per day). Box 3 illustrates how to interpret CIs mathematically.

The broader the CI, the greater the range of plausible results – that is, the less precise the estimate. This breadth is influenced by the following factors:

- **The sample size.** Intuition suggests that, if mortality in a population of 100,000 is being estimated, sampling only ten households will yield very imprecise results, even if these ten households are drawn perfectly randomly from the population and provide very reliable information. The smaller the sample size, the broader the CI.
- **The length of the recall period.** Investigating a period of 100 days will yield more precision than a ten-day period: in the formal expression of mortality rate, the denominator consists of person-time. In general, the shorter the period, the broader the CI.
- **The extent of mortality itself.** At higher MRs, with sample and population size remaining the same, CIs will be broader in absolute terms, but narrower in relative terms, i.e. as percent variation compared to the point estimate.
- **The sampling design**, i.e. how the sample is drawn. Cluster surveys pay for simplicity with reduced precision because of a design (cluster) effect (this is discussed below). Cluster designs thus result in broader CIs.

These effects are illustrated in Table 4 (p. 15) for hypothetical values of the above factors.

Retrospective surveys

Because populations are often too large to be studied exhaustively, surveys almost always try to estimate mortality based on a sample. Different options exist for draw-

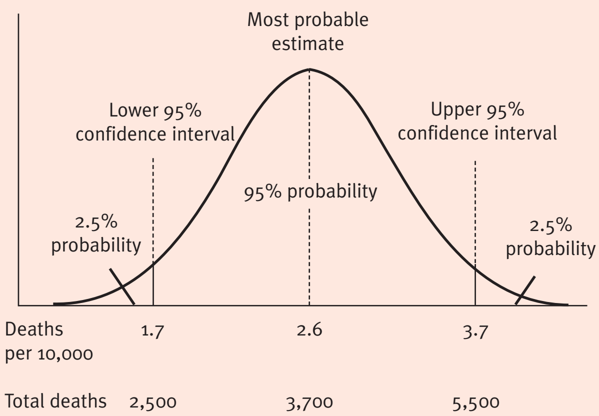
Box 3

Confidence intervals based on an example from Darfur, 2004[25]

A 2004 retrospective cluster study of mortality among 74,900 IDPs living in a camp in Murnei, West Darfur, reported a violence-specific MR of 2.6 per 10,000 per day with a 95% confidence interval (CI) of 1.7 to 3.7 per 10,000 per day. When converted into a total number of violent deaths by applying this MR to the entire population of IDPs over a recall period of 193 days, 3,700 violent deaths were estimated (95%CI 2,500 to 5,500). The correct way of describing this finding would be: 'the most probable estimate of the true mortality rate due to violence in this population, over the period investigated, is 2.6 deaths per 10,000 per day (or 3,700 total violent deaths); furthermore, we are 95% confident that the true mortality due to violence lies somewhere between 1.7 per 10,000 per day (2,500 deaths) and 3.7 per 10,000 per day (5,500 deaths). There is a small (2.5%) chance that the true mortality is actually below 1.7, and a 2.5% chance that it is above 3.7'.

The 95% level of confidence is merely a widely accepted convention in statistics: it reflects the general consensus that an error of up to 5% is acceptable when reporting results. We could, of course, settle for a lower level of confidence, and, based on the same dataset and retaining the same best estimate, report that 'we are 90% confident that the true mortality rate lies somewhere between 1.8 and 3.5'. By doing so, we have narrowed our CI – but now face a 10% risk that the true mortality is not within our CI.

CIs quantify the extent of imprecision in the estimate. Statistical theory underpinning the CI calculation demonstrates that, as we move away (in an increasing or decreasing fashion) from the survey estimate (or point estimate), it becomes less and less probable that the true population value lies at such values. This probability curve is bell-shaped around the point estimate. The true mortality is always more likely to be close to the point estimate (the peak of the probability curve), that is, around higher probability values.



ing a representative (that is, unbiased) sample, but cluster sampling is the most used because of feasibility issues. It does, however, have important limitations. Apart from sampling designs, surveys rely on well-constructed questionnaires: the current and past household census methods are two approaches to gathering mortality data for individual households.

Sampling methods

Assuming that the population cannot be surveyed exhaustively, and once the required sample size is established, the challenge becomes how to select households to be sampled so that they are representative of the population being studied. There are three established methods for doing this:

- **Simple random sampling.** This method requires a listing (**sampling frame**) of all households in the community, and some unique identification for each, such as a number. Households are chosen by drawing random numbers. In practice such a listing is almost never available or very incomplete: therefore, this sampling method is rarely feasible.
- **Systematic random sampling.** This method does not rely on any listing, but does require that households be grouped in some identifiable way, such as in villages or camps, and that the total number of households be counted. Only the first household is drawn entirely at random; subsequently, every *n*th household in the sequence is selected based on the required **sampling step** (equal to the total number of households divided by the required sample size). In addition to geographically distinct settlements, this method can occasionally be applied where households are taking refuge in public buildings, and can thus rapidly be counted, as in a 2003 survey in northern Uganda.[26]
- **Cluster sampling.** By far the most frequently used method, cluster sampling is a simple way to draw a representative sample even where (as in most circumstances) there is neither a listing nor a known total number of households (that is, no individual household sampling frame), and households are arranged in chaotic patterns. Clusters are groups of neighbouring households (usually 30 or more) out of which only the first is chosen at random, and the remainder by proximity, usually by picking the household closest to the one just surveyed (see Figure 4). Cluster sampling simply requires knowledge of administrative or geographic divisions within the population (districts; villages; sectors of a camp), and of their population size (in fact, only a measure of the relative size of each sub-division is truly necessary). The more populous the sub-division, the more clusters it is allocated. The total number of clusters varies depending on the desired sample size, but should never be lower than 25, below which statistical theories underpinning the estimation start to break down (this is intuitive: basing one's estimate on a certain number of clusters, say 30 or 40, does ensure that a reasonable range of situations is sampled, whereas if only three or four clusters were picked, it would be unlikely that these represent the full diversity of mortality experiences within the population).[27]

Among these methods, cluster sampling is the most appropriate in the field. However, it suffers from two major drawbacks:

- A mortality estimate is provided for the entire population, but geographic sub-divisions within this population cannot be studied (that is, no estimates can be generated for, say, individual districts or camp sectors; this will be discussed in the next section) unless a distinct sample of 30 or more clusters is selected for all the sub-divisions.
- Precision is lower than with a random or systematic survey of similar sample size. Because of their proximity, households belonging to each cluster are likely to resemble each other much more than if the households were selected entirely at random in the population. There is, therefore, a loss in sampling variability (or rather, excessive homogeneity within clusters). This phenomenon is known as the **design effect** (or cluster effect), and results in wider CIs around the estimate (as shown in Table 4). Anticipating this problem, cluster surveys usually assume a design effect of 2, and adjust their sample size accordingly (a design effect of 2 results in a doubling of sample size). Mortality causes that make households within clusters more homogeneous (and, conversely, increase heterogeneity among clusters) are responsible for high design effects: in surveys conducted in West Darfur, design effects up to 11.3 were noted, largely attributable to the fact that violence, the dominant cause of death, had affected households in some clusters far more than others.[28]

In practice, 30 clusters (and occasionally more) of at least 30 households (so 900 households or more) are sampled. Increasing the number of clusters is statistically preferable to increasing the number of households per cluster (less design effect, and so improved precision). On the other hand, one advantage of cluster sampling is that it enables rapid household selection while limiting the movement of survey teams to a few random points (the starting points of each cluster), which can be very important when time, logistics and/or security concerns are paramount: in this respect, varying the number of households per cluster will be more advantageous. In summary, a 50 cluster x 30 households/cluster survey will be more precise, whereas a 30 x 50 survey may be more feasible.

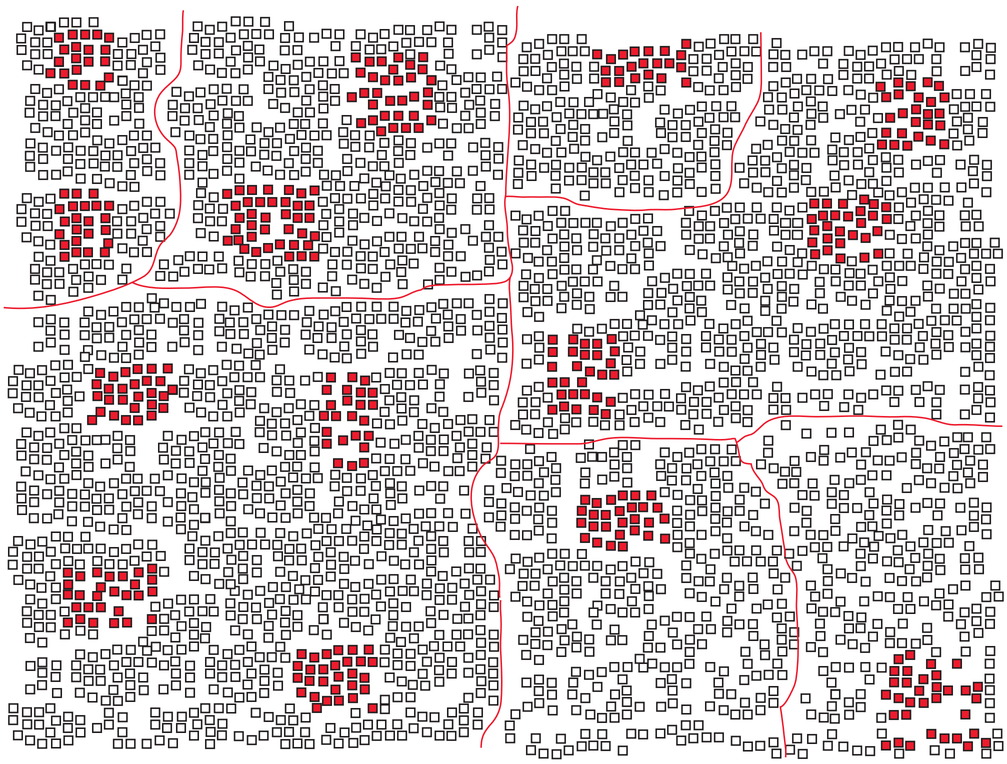
It should be noted that the above cluster method implies **population-proportional sampling**, that is, areas with greater population are allocated more clusters. **Spatial sampling**, whereby clusters are distributed proportionally to surface area, can partially be applied to mortality surveys (for example, a Global Positioning System can be used to select the first household in a cluster). The spatial approach potentially suffers from a rural bias, in that people or settlements with more space are more likely to be sampled.

Table 4: 95% confidence intervals (imprecision) of crude or under-5 mortality rate estimates according to different scenarios

	Sample size: 30 clusters x 30 households (4,500 people, 900 under-5s)			
	Recall period: 30 days		Recall period: 180 days	
	Design effect: 1.0 (i.e. none)	Design effect: 2.0	Design effect: 1.0 (i.e. none)	Design effect: 2.0
CMR: 2.0	1.4–3.0	1.1–3.5	1.7–2.3	1.6–2.5
CMR: 4.0	3.1–5.3	2.7–5.9	3.6–4.5	3.4–4.7
U5MR: 2.0	0.7–4.6	0.4–6.6	1.4–2.9	1.2–3.3
U5MR: 4.0	2.2–7.6	1.6–9.8	3.1–5.2	2.8–5.8
	Sample size: 60 clusters x 30 households (9,000 people, 1,800 under-5s)			
	Recall period: 30 days		Recall period: 180 days	
	Design effect: 1.0 (i.e. none)	Design effect: 2.0	Design effect: 1.0 (i.e. none)	Design effect: 2.0
CMR: 2.0	1.5–2.6	1.4–3.0	1.8–2.2	1.7–2.3
CMR: 4.0	3.3–4.9	3.1–5.3	3.8–4.5	3.7–4.6
U5MR: 2.0	1.1–3.8	0.8–4.9	1.5–2.6	1.4–2.8
U5MR: 4.0	2.6–6.3	2.2–7.6	3.3–4.8	3.1–5.2
Assumptions: Five people/household, under-5 children = 20% of population. Ranges shown in each cell are lower and upper bounds of the 95%CI resulting from alternative values of CMR or U5MR (4.0 versus 2.0), design effect (2.0 versus none), recall period (180 versus 30 days) and sample size (30 clusters x 30 households or 60 clusters x 30 households).				

Figure 4

Cluster sampling design in a camp



Each square corresponds to a household, and red lines denote borders of sectors. More clusters (shown as groups of red households) are likely to be drawn in large, populous sectors than in smaller ones.

In practice, the vast majority of mortality surveys are nested within classic 30 x 30 cluster nutritional assessments. The mortality questionnaire usually comes second to the anthropometric measurement of under-five children, and may be complemented by other questions on measles vaccination, food consumption, access to services and possession of essential non-food items. While coupling nutritional and mortality assessments may sometimes be necessary (or even an opportunity not to be missed when access to a site is very precarious), the limitations of this approach are significant. First, it discourages any rigorous examination of actual sample size requirements (because nutritional surveys usually include 900 children, mortality is also measured on 900 households). Second, it complicates the use of the past household census method (see below), which requires more time per interview. Third, it probably results in lower-quality interviews. Fourth, survey teams may not gather mortality information from households in which no children eligible for the nutritional assessment are found; because of this mistake, a 2005 district-wide nutrition survey in northern Uganda reported apparently normal CMR and U5MR, contradicting a simultaneous survey in the same district which focused on mortality and showed very high mortality rates.[29] Fifth, crucial mortality findings may be lost in a myriad of other data on malnutrition, vaccination and access to services. For these reasons, mortality surveys should, whenever possible, be carried out independently of other assessments. Often this may not be feasible, as agencies may not wish to forego the opportunity to gather other precious information. While this is understandable, it should be clear to all involved in the planning of a survey that there is no good shortcut to measuring mortality properly: it requires proper interviewer training, significant space in a questionnaire, exhaustive questioning of the household and careful data analysis.

Data collection

Apart from an appropriate sampling design, the key elements of a well-conducted mortality survey are:

- A carefully designed questionnaire. This should be standardised, to contain very specific questions that limit the subjectivity of individual interviewers, translated into the local language, back-translated so as to verify the reliability of the translation and pre-tested on a small group of households before starting the survey in earnest (this pre-testing provides a chance to adjust the questionnaire). Having fewer questions in the survey tends to result in higher-quality information.
- A calendar of locally recognisable events that will help respondents to accurately recall the month or approximate period when certain events, such as death or displacement, actually occurred; the beginning of the recall period should always be set at a very easily identifiable date, such as Christmas, New Year, the end of Ramadan, a major climactic event (in Angola, a solar eclipse was used) or political occurrence (in Angola, the death of rebel leader Jonas Savimbi).
- Clear criteria for classifying the causes (see below) and circumstances of death: information on these may include location (in the community of origin, in flight, in the camp; alternatively, at home versus in a health facility, which provides a measure of health service coverage), and date.
- Unambiguous definitions to aid in coding other possible questionnaire responses, such as what differentiates spontaneous departure from the household from disappearance (for example, based on whether the household has any news of the absent person).

There are two established questionnaire designs for collecting information on household mortality during the recall period of interest.[30] These are described below.

- **Past household census.** The first step of the mortality interview is to list all household members (by age and gender) who were present at the beginning of the recall period, and establish which of these original members are still present in the household on the survey day. Any additional members present now but not at the beginning are identified, and the approximate date on which they joined the household is noted. The fate of original members who are no longer present is then established: they may have died, disappeared or simply moved away. The date and/or location of these events is also noted, along with any other supplementary information, such as the cause and circumstances of death. Finally, the interviewer tries to learn about individuals who came into the household after the start of the period but left before the survey date, for instance children who were born and then died. The end result of the past household census method should be an accurate reconstruction of each household's demographic evolution from the beginning to the end of the recall period, including deaths and other leaving/joining events, and when/where these occurred. Mortality is defined as deaths per person-time at risk during the recall period, namely using the formal expression of MRs (see Chapter 1).
- **Current household census.** In this simplified method, surveyors first establish the composition of the household on the day of the survey. They then ask about any births or deaths occurring since the beginning of the recall period. They may also ask about other leaving or joining events, such as disappearances. In so doing they obtain sufficient information on the numerator of the mortality expression (deaths), but may miss some of the changes in the size and composition of households, that is the denominator (population at risk). This method is less applicable in settings where households are frequently breaking apart. Mortality is usually calculated in these surveys using the simplified MR expression (see Chapter 1).

The current household census method makes for shorter interview times, provides for a relatively simple analysis and is probably valid when the recall period is short (for periods during which no major demographic changes, aside from mortality, are expected to have occurred). It assumes that inward migration to the sample households roughly equals departures from those households. The past household census method is far more rigorous, as it allows surveyors to more clearly identify missing household members (by virtue of the cross-checking logic of the questionnaire), and estimate very precisely the person-time denominator. It is indispensable when the recall period is long and the population has a complex history of displacement and separation. For example, in 1996 a group of Rwandan Hutu refugees living in Zaire were chased out of their camps and pursued for hundreds of miles during their westward flight by Zairean pro-government forces, who mounted regular attacks. Later that year, remnants of this group turned up as far afield as the Republic of Congo. A survey based on past household census was able (and indeed necessary) to reconstruct their appalling experience (see Figure 5).

A third method for obtaining mortality data, known as **'children ever born'** or **'previous birth history'**, is used to estimate infant and child mortality in the context of national health surveys, such as UNICEF's Multiple Indicator Cluster Surveys. It has been suggested for emergencies.[32] Mothers are asked about all the children they have ever had, and the fate of each child born within the recall period of interest is noted, yielding infant or under-five death rates. Unfortunately, this method does not capture orphans, who probably have a far higher risk of mortality, and is not adapted for settings where the survey comes late with respect to the crisis onset, and significant under-five mortality has already occurred. The method has been shown to grossly under-estimate under-five mortality.[33]

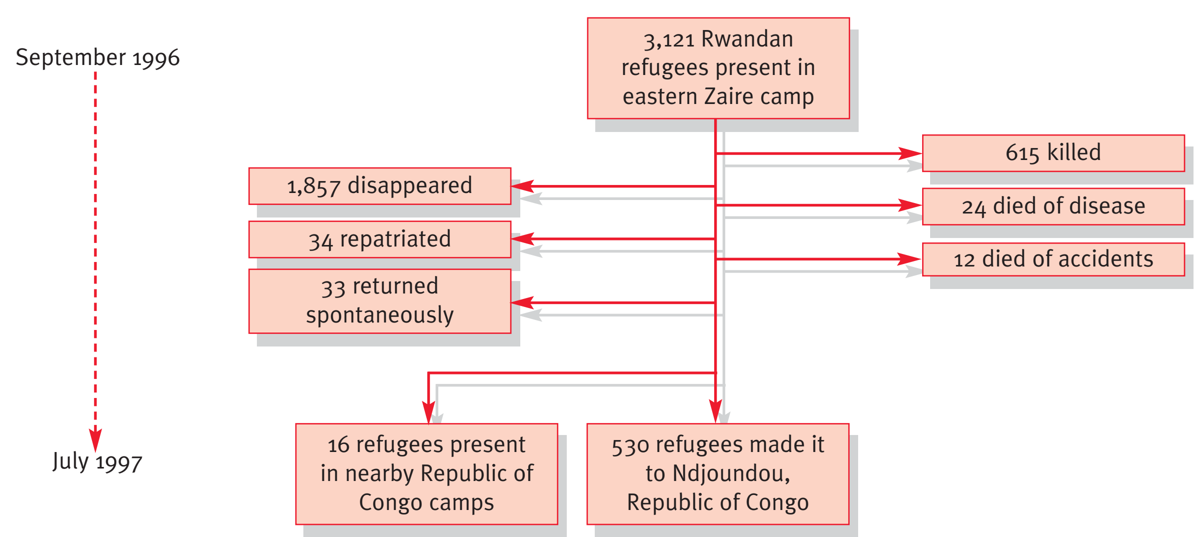
Whatever the interviewing method used, establishing **causes of death** (with the exception of violence) can be challenging, and no clear guidelines on this exist for field surveys. An in-depth interviewing procedure of next of kin, known as 'verbal autopsy', is used in more sophisticated research projects, for example where researchers are attempting to establish the impact of certain interventions on a particular disease, such as the effect of impregnated bed nets on malaria.[34] The verbal autopsy technique, however, requires a lot of training and some medical knowledge; it is also very protracted, and thus not feasible in humanitarian emergencies. In the absence of alternatives, some authors recommend only classifying causes of death for which unambiguous, local terms are known.[35] Usually, most populations will recognise conditions such as measles, neonatal tetanus or diarrhoea, and will have local words to describe them. The two manifestations of severe malnutrition, marasmus (wasting) and kwashiorkor (oedema), may also be recognisable (in Angola, respondents referred to this as *anemia*). This is not the case for other common diseases such as acute respiratory infection or malaria, although anthropological research has identified terms that correspond well with a cerebral malaria diagnosis, such as *degedege* in Tanzania and *soumaya* in Mali.

The ethics of sensitive data collection

Surveys are increasingly invoked to document past abuses of human rights and humanitarian law.[36] Retrospective mortality surveys are not thought to require approval from a research ethics review board. They do, however, require at least verbal consent from respondents, who must be adequately informed of the intended use to which the survey will be put. Respondents may experience deep stress whilst recounting their experiences: they must therefore be made aware that they can refuse to answer any ques-

Figure 5

Evolution of a sample population of Rwandan refugees, September 1996–July 1997, reconstructed by the past household census method[31]



tions. Finally, they must be assured of confidentiality, and their names should never be recorded for survey purposes. In general, all data on violence must be treated with the utmost care, from collection to publication. Furthermore, semi-public interviews by rapidly trained surveyors are not an acceptable vehicle for gathering information on events such as rape or torture, which may cause profound mental trauma and social stigma in their survivors. Nevertheless, rape in particular has been cited as a systematic, purposeful war crime, and it seems important to quantify its full extent more rigorously than is possible through collating sporadic reports from health facilities, which probably vastly under-estimate the true burden.[37] To do this, existing methods from industrialised countries may have to be adapted for emergency settings (a full discussion of the epidemiology of sexual violence is beyond the scope of this paper). Care must also be taken to ensure that survey teams and their sponsoring agencies do not face intolerable security risks as a result of such studies.

How many resources are necessary for a survey?

Mortality surveys are intense, full-time, but brief efforts. They should only be led by experienced people, preferably with epidemiology and statistics training, and backed up by a senior scientist for questions on design, questionnaire building and analysis. Teams of full-time, literate surveyors are needed, preferably able to converse in local languages and the language of the survey leader, along with adequate transport and logistics back-up (food, communications devices, stationery and, preferably, a computer). The number of teams often determines the duration of the survey (one team is usually made up of two people, and can complete between one and two clusters per day). On the other hand, limiting the number of teams probably ensures better trained teams, and better monitoring. Execution of the survey, including data entry, analysis and reporting, can take between one week and one month.

Prospective (real-time) surveillance

Mortality surveillance systems require teams of trained home visitors who record deaths as they occur in populations of known size. This approach should always be considered, as it enables real-time monitoring of mortality trends, facilitating a quick reaction. It does, however, require ongoing supervision.

Surveillance of mortality implies active detection of all deaths occurring in the community, not just in health structures. To this end, teams of home visitors need to be organised, each of whom is assigned to a specific sector or village, and tasked to visit each household in his/her catchment area on a regular basis (every day or once a week) so as to inquire about any deaths, as well as to regularly update the total population size of the area, thereby providing a reliable denominator for the MR calculation. In some cases, local residents can be trained to record all deaths in their community. Methods to estimate population size accurately (and their limitations) are beyond the scope of this paper, but the importance of working with reliable population figures should not be underestimated when implementing surveillance systems. Rough estimates, such as from food distribution data, may be reasonable to start with, but should as soon as possible be replaced by figures obtained either though a validated estimation and mapping exercise, or, better still, a full (and unbiased) census and registration. Partly because of this, surveillance requires a considerable number of home visitors – at least one per 1,000 people.[38]

The analysis of surveillance data is relatively straightforward and can be performed by anyone armed with a pencil. For example, if at the end of the week a home visitor reports four deaths in a sector of 1,540 people, the CMR must be 4/1,540 x 10,000/7 days = 3.7 per 10,000 per day. For surveillance to make any sense in an emergency, such analysis must be weekly. This provides an opportunity to observe trends in real-time, and mount a rapid response. It should be noted that, especially if small sub-populations such as camp sectors are under observation, some apparent weekly fluctuations may simply be a result of chance: in the above example, one death would yield a CMR less than 1 per 10,000 per day, whereas with two the emergency threshold would be crossed. Similarly, if population estimates are updated upwards or downwards, MR will appear to shift suddenly, while actual mortality may have remained unchanged.

Generally, the presence of an active, prospective surveillance system, coupled with other crucial community health activities such as nutritional screening, referral of severe cases and health information, is a sign of programmatic quality, and of a proactive rather than reactive approach to humanitarian relief. Epidemiology is often employed when epidemics strike or an unexplained deterioration in health status occurs, but this is hardly good timing: the primary use of epidemiology in emergencies should be to prevent and detect health problems. Mortality surveillance is, therefore, highly advisable from the very onset of any humanitarian intervention.

Surveys and surveillance

Surveys and surveillance are by no means mutually exclusive, but could rather be seen as complementary. Their relative strengths and weaknesses are listed in Table 5. Surveillance enables real-time monitoring of mortality trends, and thus a quicker reaction. However, it also requires some regular epidemiological supervision, and its quality may not be sustainable over many months. Furthermore, it is appropriate mostly for camp-dwelling or regimented populations. Surveys can generate very reliable data, but do not reflect trends in sufficient detail. On the other hand, they can be adapted to almost any setting, however remote and arduous, and constitute a one-time effort.

Rapid convenience surveys

In order to obtain results rapidly and with minimal effort, surveys are sometimes carried out without a proper sampling design, but rather on readily accessible populations, and relying on small sample sizes. During the very early days of an emergency, or when conducting rapid site assessments, mortality is often estimated roughly by drawing non-representative samples from the population, based on criteria of feasibility and speed of data acquisition. Examples of these methods include interviewing heads of households standing in line for a food distribution, or mothers bringing their children to a vaccination point. Alternatively, local people of importance (chiefs, leaders of women's groups) may be asked to list recent deaths in the communities under their oversight.

Convenience surveys seem attractive because of their simplicity, but in reality are riddled with potential bias; apart from very rapid site assessments, convenience surveys should not serve as decision tools. They tend to ignore the more disenfranchised and vulnerable members of the population, who may not have access to distributions or may be marginalised by the leadership of the affected community (for example, because they belong to a different ethnic group). Ultimately, they may be less cost-effective than proper studies. Occasionally, freshly dug graves may be counted, assuming local customs include the use of central cemeteries. Grave-watchers may also be hired to follow trends in burials prospectively. Problems with this method include difficulty in distinguishing new from old unmarked graves (especially after rain), incomplete census of all burial grounds in the community (some may be small and informal), and inability to distinguish residents from displaced people among the dead.

Mortality data from non-representative convenience surveys can be instrumental in sounding an alert, and the collection of such data should not necessarily be discouraged. This data must, however, be confirmed by properly designed, scientifically valid surveys. Most importantly, data from convenience surveys suggesting low mortality should under no circumstances lead to a conclusion that death rates are indeed low. Unfortunately, non-random sampling methods are common in emergencies.[39] For example, during Ethiopia's famine in 2000, 46% of nutritional surveys had convenience sampling; of the remainder, only 9% had a valid cluster design.[40]

Table 5: Comparison of relative strengths and weaknesses of surveys and surveillance for estimating mortality

Retrospective surveys	**Prospective surveillance**
Strengths	**Weaknesses**
Can obtain MR estimate without knowing population size	Needs updated, accurate population size
Can be performed in rural or camp settings	Only feasible in camps or regimented populations
Epidemiological input needed, but only for duration of survey	If cluster design is used, no sub-area analysis is possible
Requires minimal epidemiologist supervision	Requires large teams of home visitors on a long-term basis
Can analyse MR by sub-area	
Data quality can be highly controlled	
Requires a team of surveyors for a short period	
Weaknesses	**Strengths**
MR estimate comes after the fact, and often too late for meaningful intervention	Enables real-time monitoring of trends, quick response
MR is an average of past period, may not reflect trends in the past few days/weeks	Highlights weekly fluctuations in MR
Impact difficult to measure due to lack of sub-period detail (weekly MR obtainable from surveys is very imprecise)	Detects impact of specific interventions immediately
High possibility of bias, especially response bias (population may perceive that the survey is a registration- or distribution-connected activity)	May minimise response bias (population becomes used to surveillance)

Chapter 4
Interpreting and using mortality data

Validity of mortality studies

Mortality reports should present information clearly, and non-epidemiologists should be able to evaluate their quality. Significant bias may be suspected, but should not automatically result in a study being rejected; rather, it should be balanced with information needs.

Judging the quality of a mortality report

Non-specialists should read mortality reports critically, and interpret their findings in light of both sampling/non-systematic error (imprecision) and non-sampling/systematic error (bias). Box 4 gives a checklist for evaluating the probable quality of mortality data obtained either through surveys or surveillance.

No mortality report should be dismissed simply because the MR findings are 'obviously too low' or 'obviously too high' – at least not before very detailed site assessments have taken place. Certain MRs can, however, justify some degree of suspicion, and should lead to a more detailed evaluation of the work behind such estimates. In particular:

- any CMR below 0.5 per 10,000 per day (or any U5MR below 1 per 10,000 per day) in the acute phase of a crisis should be viewed as exceptionally low, at least in Sub-Saharan Africa; there may be a problem with under-reporting bias; and
- it is exceptional, at least in Sub-Saharan Africa, for CMR to be higher than U5MR, unless a known cause of mortality, such as armed attacks, disproportionately affects adults.

Working with bias

No study is perfect, and mortality studies carried out in emergency contexts are among the most difficult of epidemiological investigations. Common sources of bias have been listed in Chapter 3. Some can occasionally be spotted. For example:

- patterns in the dataset may provide some clues: a particular interview team may consistently report far higher mortality than the other teams, for instance, possibly due to a misinterpretation of the questionnaire;
- any inflation of household size by respondents can be verified against other sources, such as World Food Programme registration lists; and
- discussion with survey teams may reveal misunderstandings about the length of the recall period.

Most other systematic errors, however (especially recall and response bias), are not measurable, and cannot be corrected *post facto*. Neither can an unexpectedly high design effect.

Box 4
Proposed checklist for evaluating a mortality report

Retrospective surveys

- The choice of a recall period is justified by quantitative or anecdotal evidence about past mortality in the population, and by the stated survey objectives.
- A sample size calculation is provided that corresponds with the stated survey objectives.
- The sampling design (simple random, systematic random, cluster) is clearly explained.
- The method for defining, identifying and enrolling households is described.
- A copy of the questionnaire is provided, and authors report that it was field-tested before the start of the study.
- An attempt was made to classify causes of death, at least into the two main categories of 'violent/accidental' and 'medical'.
- The proportion of households replaced during sampling is reported.
- 95%CI are reported alongside each point estimate, and (if a cluster design was used) the design effect is provided, or the authors state that they accounted for design effect in their calculation of 95%CI.
- Potential sources of bias are listed, and their potential effect on the validity of the study is discussed.

Prospective surveillance

- The population size (denominator) used to calculate MR is clearly reported (as well as the date when it was last updated), and the method by which this was estimated (census of households; area mapping; collation of agency data from food and other registrations) is described.
- The division of the population into sectors under the responsibility of each home visitor is clearly shown (ideally a map should be provided).
- The method of mortality data collection is described (frequency of visits to households; method of ascertaining deaths, and prevention of likely biases such as multiple reporting of the same death among neighbours).
- A copy of the questionnaire/data collection tool is provided.

Bias assessment should accompany every mortality report, and not just occur when certain parties manifest doubts about the findings. When bias is suspected or the results are questioned, the attitude must be cool-headed and scientific. If the study is of great importance, a panel of experts could rapidly be assembled to review its methodology and analysis, and decide how realistic the findings are in light of all available data about the health status of the surveyed population.

When deciding whether to accept a moderate risk of bias and act upon the findings, relief managers might wish to reflect on the following:

- what is the size of population to which the mortality findings apply? The bigger the population, the bigger the potential mistake;
- how will the data be used, and with what urgency are they needed?;
- is it likely that another survey can be performed in the near future, or that surveillance can rapidly be implemented?; and
- are there other sources of data, such as mortuaries or clinics, that could confirm or refute these findings?

It should be noted that not all significantly biased data are necessarily useless:

- if an under-estimate is suspected and the MR still exceeds the emergency threshold, the survey, however biased, still indicates that urgent action is needed;
- in case of an under-estimate and reported MRs below emergency thresholds, one cannot conclude that there is no emergency; and
- if an over-estimate is suspected, repeating the survey becomes necessary, unless the reported MR is below the emergency threshold, which indicates strongly that the situation is under control.

Point estimates and confidence intervals

In any MR estimate – say 2.5 per 10,000 per day, 95%CI 1.6 to 3.9 – the point estimate (2.5) is the most likely value of mortality, and the confidence interval (1.6 to 3.9) expresses the degree of uncertainty around it. Both should be reported, but their interpretation becomes difficult in borderline situations, where mortality is close to the emergency threshold.

How precise can we be?

Assuming that the survey relied on representative sampling and controlled other potential sources of bias, the point estimate is the most likely value of mortality in the population of interest. The point estimate is therefore the value to which an emergency threshold should be compared. As discussed above, the confidence interval describes the degree of uncertainty surrounding the point estimate; it is not a reflection of bias, but rather of sampling imprecision. Both point estimates and CIs are extremely informative, and both should be reported when describing mortality in any population.

The narrower the CI, the more confident one can be about the result of the survey. But how narrow is narrow, and how broad is broad? There are, unfortunately, no cut-offs to decide this, and much depends on the actual MR. Common sense, however, would not be satisfied with, say, a finding of CMR 2.0 per 10,000 per day and a 95%CI ranging from 0.3 to 5.8. Such a finding would denote poor survey design, namely that investigators did not define from the outset a sample size that would be sufficient to detect elevated mortality with sufficient precision. Obviously, the CI from such a survey will become narrower if we accept a lower level of confidence – 90%, 80%, 70% and so on. Whether 95% confidence is actually too strict a standard in the context of emergency mortality surveys perhaps deserves further discussion among the experts, since the implications of CIs in this context are very different from those in most other areas of epidemiology. In general, the less regret one will have when acting on data (in terms of deaths averted, or costs incurred) that are in error, the less important confidence intervals become. For example, a high child mortality estimate attributed to measles would be likely to lead to an immunisation and vitamin A campaign which, even if uncalled for, might still serve those children in years to come. An equally high error in estimating mortality attributed to violence might lead to a military or diplomatic response which, if inappropriate, would induce a relatively large amount of regret.

As the case may be, lowering the level of confidence cannot replace the requirement for properly designed surveys and adequate sample sizes. At the very least, surveys should have a sample size that is sufficient to clearly identify seriously elevated mortality. Without wishing to be dogmatic, we suggest that a minimum of 900 households (30 clusters x 30 households) and a recall period of at least three months should be investigated whenever possible (or 450 households if random sampling can be done). This sample size detects a CMR of 1.5 per 10,000 per day with ±0.5 per 10,000 per day precision, assuming a typical household size of five individuals.

Sometimes, access to populations in crisis may be patchy at best, and for a variety of reasons (the population is on the move; the agency is hesitant to deploy survey teams due to security reasons), one may decide to carry out a study with a less than adequate sample size, especially if there is a strong suspicion that CMR could be high. In this case, 450 households (30 clusters x 15 households, or 225 with random sampling) may be acceptable. This sample size would detect a CMR of 2.0 per 10,000 per day with precision ±1.0.

Proposed interpretation rules

If surveys are carried out, one of the following four scenarios may be encountered. We propose below a set of responses corresponding to each:

- **Scenario 1: Both the point estimate and the 95%CI are clearly above the emergency threshold (example: CMR 2.3 per 10,000 per day, 95%CI 1.3 to 3.1).** The emergency is not under control and significant excess mortality has occurred, even if the lower bound of the CI is considered. The situation could be as bad as the upper bound of the CI. An urgent improvement and scaling up of relief operations is needed; data on causes of death reported by the survey, as well as other available assessments, may point to those interventions which are likely to be most beneficial. Repeat surveys and/or surveillance should be planned so as to monitor the evolution of mortality.

- **Scenario 2: The point estimate crosses the emergency threshold, but the 95%CI includes values below it (example: CMR 1.1 per 10,000 per day, 95%CI 0.6 to 1.6).** The relief community has little choice in the short term but to act upon the assumption of an out-of-control situation, and thus strengthen humanitarian assistance. In so doing, it accepts the risk that the relief operation may be excessive compared to needs. Repeat surveys and/or surveillance are urgently needed to confirm and monitor this finding.
- **Scenario 3: The point estimate is below the emergency threshold, but the upper 95%CI crosses it (example: CMR 0.8 per 10,000 per day, 95%CI 0.4 to 1.3).** Such a result is encouraging, but there is evidence that excess mortality persists. Indeed, mortality could be as high as the upper CI. There is insufficient evidence to justify scaling down relief operations. The relief community should compare this finding with any previous estimates available, and decide whether a downward or upper trend is evident. Most importantly, a repeat survey should be organised, this time with sufficient sample size to show a clear reduction of MR below the emergency threshold.
- **Scenario 4: Both the point estimate and the 95%CI are clearly below the emergency threshold (example: CMR 0.5/10,000, 95%CI 0.2 to 0.8).** The emergency is under control, and relief efforts, at least from the standpoint of minimising mortality, are probably adequate (note that other services, such as psychological care and schooling, may still be scarce). The upper 95%CI, however, suggests that mortality may still be elevated compared to baseline levels, and repeat surveys or prospective surveillance are necessary.

As discussed in the previous chapter, good precision in surveys requires a significant sample size, a relatively long recall period and a low design effect. When mortality is very high, achieving very good precision is not crucial. The operational response to a CMR of 20 per 10,000 per day is unlikely to differ much from the response to a CMR of 10 per 10,000 per day, and even a broad CI will succeed in conveying the gravity of such a situation. In borderline situations, however, the limitations of survey sampling become all but insurmountable: for example, assuming a recall period of six months and cluster sampling with design effect of 2.0, classifying a CMR of 1.1 per 10,000 per day as being unequivocally above the emergency threshold would require a precision of ±0.1 per 10,000 per day (that is, a lower 95%CI bound not below 1.0), namely a sample of 46,953 households. Epidemiologists sometimes struggle to convey this important limitation when arguing for adequate resources for surveys, and explaining their findings to non-specialist audiences. In short, in situations where only survey data are available, and where the CI is both above and below the emergency threshold, the interpretation is mostly conservative, and assumes worst-case scenarios; it should, however, be informed by a good dose of objectivity and common sense.

It must be noted that a well-organised, weekly prospective surveillance system which captures deaths in the population exhaustively through a capillary network of home visitors will yield absolute MRs with no CIs around them (that is, no sampling error), thereby circumventing scenarios 2 and 3 above. Quality surveillance is certainly superior to quality surveys when MR is close to emergency thresholds; furthermore, it generates weekly updates that will clearly highlight any declining or increasing trend.

Whenever mortality data are interpreted, there will probably always be a range of vested interests attempting to demonstrate that an emergency is continuing, or that it is definitely over. Entrusting the conduct and interpretation of mortality studies to experienced staff who are independent from major political powers and who can exercise scientific objectivity is of great importance if misuse of data is to be prevented.

Confidence intervals and proportionate mortality

There is perhaps one situation in which the importance of CIs is diminished, namely when a single cause of death is overwhelmingly responsible for the excess mortality observed. For example, in 2003–2004 a population of 75,000 West Darfurians experienced a CMR of 9.5 per 10,000 per day before reaching Murnei camp, and 93% of these deaths were due to violence.[41] This dramatic finding alone is probably sufficiently eloquent, and CIs here would not add much operationally relevant information.

This example assumes that the survey captured a sufficient number of deaths to yield precise estimates of proportionate mortality. The proportion of deaths due to any given cause also has a CI associated with it, if this proportion is drawn from a sample. This CI becomes broader as the total number of deaths decreases, and as the design effect becomes larger. For example, four diarrhoea-related deaths out of a total of ten reported, and 40 diarrhoea deaths out of 100 would both yield a proportionate mortality of 40%, but the respective 95%CIs (assuming a design effect of 2.0) would be 7% to 83% and 27% to 55%. Proportionate mortality data should thus be interpreted with caution when they are based only on a small number of deaths.

Interpolation, extrapolation and stratification

Mortality rates may be generalised to total numbers of deaths in the population and period studied (this is called interpolation). Extrapolation (application of the findings to different populations and periods) is more risky, but may be performed cautiously if certain conditions are met. Stratification refers to sub-group (or sub-period) analysis. Mortality estimates in sub-groups are always less precise than in the entire sample. Single sample cluster surveys do not allow for sub-area stratification. This is a major limitation of this sampling design.

Estimating death tolls within and outside the study period and population

Total death toll figures have a far greater policy and media impact than mortality rates. Whereas epidemiologists may be satisfied with the information provided in MRs, relief programme managers, NGO communications and fundraising staff, donors, politicians and journalists are far more interested in the hard numbers.

We define **interpolation** as the process by which mortality rate findings are applied to the entire population from which the sample was drawn, so as to obtain total numbers of deaths or percentages of the population that died over the recall period being investigated. We call **extrapolation** any extension of the findings to a population and/or period that was not represented by the sample. If the survey is well-conducted and relatively bias-free, results may be considered representative of the population from which the sample was drawn, during the period of time covered by the questionnaire. Interpolation is therefore statistically justified and readily performed (see Box 5). However, death-toll estimates may sometimes be very imprecise. The first response when considering such a figure should be to ask for a confidence interval, and a CI should accompany, if not supersede, any total death toll estimate.

Extrapolation has little scientific justification, and can greatly distort actual survey findings, discrediting individual research groups and the disciplines of demography, epidemiology and statistics. Nevertheless, at least one understandable reason why extrapolation is often performed can be cited: where it is insufficient and/or impossible to collect data. Extrapolation across different populations works by association: if two populations appear to be experiencing similar conditions, for example the same patterns of disease or food security, or a similar proximity to conflict, it is assumed that the mortality experience of one can be applied to the other. Extrapolation to periods before or after that investigated by the survey is more risky, and more difficult to justify.

We believe that extrapolation is not acceptable when *any* of the conditions below are met:

- it applies a mortality finding to a far wider area/population than that covered by the survey which generated that finding, and does so without any attempt to compare qualitative and quantitative data on the population surveyed with data on the populations to which the extrapolation is applied;
- it relies on highly unreliable population figures;
- it is driven essentially by political motives, or a preconceived notion of what mortality 'must be'; and
- it is a short-cut around the requirement for rigorous surveys and surveillance.

Conversely, we believe that some form of cautious extrapolation, guided by logic and a deep knowledge of the field context, and accompanied by detailed acknowledgement of its likely limitations, is acceptable and occasionally even useful when *all* of the following conditions are met:

- data from several different surveys/surveillance systems are available and, though they may not cover the entire population experiencing the crisis (or may reflect different periods in this crisis), they do offer some degree of geographical representation;
- it would clearly be impossible or very difficult to collect exhaustive data on the entire population of interest; and
- the agency performing the extrapolation has an objective view of the situation.

There are as yet no guidelines about how to perform extrapolation in mortality measurement, and every case is likely to be different. Whatever extrapolated estimate is provided, however, should be accompanied by a full explanation of how it was derived, and of the assumptions made. A range of likely lower- and upper-end values, rather than a single value, may be a more prudent and meaningful way of expressing and interpreting such estimates.

Significant examples of recent interpolations and/or cautious extrapolations performed on the basis of mortality studies can be cited from:

- the Goma area of eastern Zaire (now the Democratic Republic of Congo), where epidemiological investigations in the wake of the 1994 Rwandan refugee influx (and based in this case on both surveillance and surveys) reported that almost 10% of the population had died during the first month of the crisis, mostly due to diarrhoeal disease epidemics;[42]
- Kosovo, where a survey done after the 1999 withdrawal of Yugoslav forces estimated a total death toll of 12,000 due to violence;[43]
- the DRC, where nationwide surveys by the International Rescue Committee have put at 3.8 million the total number of excess deaths due to conflict since 1998;[44]
- Iraq, where a conservative interpolated estimate put excess deaths after the 2003 invasion at 98,000 (95%CI 8,000 to 194,000).[45] A noteworthy aspect of this survey was the decision by the authors to exclude from the analysis a cluster in the city of Falluja which accounted for a disproportionate number of violent deaths, and for which the possibility of strong bias could not entirely be discounted.
- Darfur, where a WHO survey, combined with other site-specific surveys and projections, led to an estimate in September 2004 of 4,500 to 9,000 excess deaths per month in a population of 1.8 million IDPs. A recent doc-

ument has attempted to estimate death tolls since the start of the conflict by collating available data.[46]

Stratified analysis

Stratification refers to the sampling and analysis of subgroups or sub-periods. A stratum is simply a category of individuals in the population: women, for example, or 5- to 14-year-olds, unaccompanied children, an ethnic group which may have suffered persecution disproportionately, or recently arrived refugees. It may also refer to a portion of the recall period: some number of months, or a specified period before or after a particular event. Calculating U5MR is the most common use of stratification. The main drawback when stratifying is that the denominator becomes smaller because fewer people are considered in the analysis or a more limited timeframe is used, resulting in a loss of precision (that is, wider CIs). For this reason, it is unwise to present and interpret any stratified MR without a CI.

Stratifying can be statistically legitimate when working with surveillance data, or data from surveys in which sampling was random (simple or systematic: see Chapter 3). By contrast, a limitation of cluster sampling designs is that mortality in sub-areas cannot be estimated. For example, a programme manager, having commissioned a cluster survey of several refugee camps which his or her agency is running, might wish to know which of the camps experienced the highest mortality; to do this would imply calculating CMR separately for each camp based on the clusters sampled in each, but would not, in fact, be a statistically valid procedure: as discussed in Chapter 3, 30 clusters can be considered representative of an entire population, but, say, three clusters cannot adequately represent the full range of mortality in the sub-area they were sampled from. Any sub-area analyses in a cluster survey report should therefore be disregarded.

When distinct sub-populations are likely to have experienced radically different mortality rates – urban versus rural, for example, or IDPs versus non-displaced – it is statistically advantageous to select a separate sample from each population. This stratified analysis will give a far more precise final estimate, although typically it results in a larger number of clusters and households visited, and therefore more work and expense.

Missing populations

High mortality or displacement can dramatically alter the demographic structure of a population. Surveys which collect the age and sex of each individual household member can display this phenomenon visually by plotting an **age-sex pyramid** of people present in the households on the day of the survey, and comparing it, for example, to that for a neighbouring country that is not experiencing a crisis. In conflict settings, surveys often highlight a deficit of young men of fighting age, who could have been killed, been conscripted to fight or gone into hiding. Men may also migrate to other regions of the country to look for work. This deficit is sometimes very alarming if coupled with information on

Box 5
Example of interpolation and extrapolation from a survey in Angola[47]

In August–September 2002, a retrospective cluster survey was conducted in Angola to measure mortality in a population of IDPs who had been ordered to assemble at resettlement centres following the April 2002 ceasefire. These IDPs had previously lived under the command of UNITA rebels. The survey covered a recall period of 427 days, including both pre- and post-ceasefire periods, and was representative of 149,106 UNITA IDPs. In total, 900 households (30 clusters x 30 households) were sampled. Of 390 all-cause deaths recorded, 69 (18%) were reported to be due to violence. The violence-specific MR was thus 0.27 per 10,000 per day (95%CI 0.20 to 0.36; design effect 1.57, mid-period sample = 6,056 people).

Can we interpolate a total number of deaths due to violence in the study population over the entire recall period?
First, we can estimate the percentage of persons who died due to violence in the population. Overall, 1.15% (95%CI 0.85% to 1.54%) were killed (1.15% = 0.27 per 10,000 per day x 427 days recall divided by 100, so as to change the unit from 'per 10,000' to 'percent'; the same operation is carried out for the lower (0.20) and upper (0.36) limits of the 95%CI). Applying these percentages to the entire population from which the sample was drawn, we obtain an estimate of 1,715 (95%CI 1267 to 2296)* people killed within this population from the beginning of the recall period (June 2001) to the survey date in August–September 2002 (1,715 = 1.15% x 149,106).

Can we extrapolate the findings to the entire population of UNITA IDPs living in resettlement camps?
The survey only included camps where Médecins Sans Frontières (MSF) was working. These contained 38% of the entire UNITA IDP population registered for resettlement. If we extend the findings to all UNITA IDP camps, the total estimate of people killed becomes 4,513 (95%CI 3,334 to 6,042) (4,513 = 1,715 estimated deaths in the population surveyed/0.38). How likely is this extrapolated estimate? The survey did have good geographical spread. However, MSF is likely to have selected camps where mortality was especially high. On the other hand, some camps not surveyed were not accessible at the time due to insecurity and landmines: violent mortality there may have been higher than in the survey camps. Extrapolation in this case would probably be too risky. The survey does, however, strongly suggest that violent mortality among *all* UNITA IDPs was seriously elevated, and bears adequate witness to the plight of this population.

* This is a slight under-estimation. The figure of 149,106 is the end-of-period population, whereas we should have used the mid-period population, which would have been higher (we could adjust for this based on observations in the sample).

systematic violence in the region. Its consequences are also significant, as the resulting women-headed households are often more vulnerable than the average (and should thus be identified and targeted for additional relief/protection). HIV/AIDS has resulted in a dramatic thinning of the population as age progresses: older age groups are increasingly a minority, and children under 15 are demographically dominant. Where under-five mortality is high, one can also observe a thinning of the under-five age group with respect to older age groups, and of the under-one group if infant mortality is particularly elevated. Some demographic information is, however, necessary to make the link between high child mortality and the shape of an age-sex pyramid: reduced fertility can also lead to shrinkage in the youngest age groups.

Understanding causation

Cause-of-death information (other than violence) is often not specific, and thus may have only limited operational relevance. Certain causes of death, such as malaria and respiratory diseases, can be difficult to diagnose. Moreover, mortality studies mostly do not explore the distal determinants of poor health. War and displacement increase the risk of disease and poor disease outcome – emergency relief can only address some of this risk.

Violent causes of death are easily recognisable provided little response bias occurs (that is, respondents are honest), and can have immediate implications for the protection of civilians and international advocacy. The age and sex profile of such deaths must be investigated carefully. Any killing of children and women clearly points to violations of humanitarian law. Many violent deaths among adult men may reflect casualties among combatants. It will sometimes be difficult to appraise this proportion, and such information should generally not be requested of respondents, as it may place them at risk. Background knowledge on the evolution of the crisis, evidence on the composition and location of combatant armies and anecdotal accounts of how and against whom violence was perpetrated can help in deciding whether adult male violent deaths are mostly civilian in nature.

As discussed in Chapter 3, classifying medical causes of death in mortality surveys is very difficult in the absence of verbal autopsy techniques. The validity of such data is uncertain. Causes of death are probably more easily interpretable and more relevant in children:

- Neonatal tetanus is usually recognisable because of the spasm typically associated with it; tetanus deaths, if noted frequently in the sample, suggest poor antenatal care.
- Measles is usually well-classified, but its true burden tends to be under-estimated, since the characteristic rash is not always visible and many measles deaths are due to complications such as acute respiratory infections or diarrhoea; the occurrence of measles deaths in the sample should lead to a recommendation to reinforce vaccination services.
- Malaria and ARI have great symptom overlap and are often confused in clinical practice; surveys will almost certainly not discriminate any better between the two, unless a serious malaria epidemic is occurring.[48]
- Diarrhoea can probably be classified with reasonable specificity if interviewers find a locally valid definition of what constitutes severe diarrhoea; high proportionate mortality due to diarrhoea should lead to better nutrition (see below) and water and sanitation; however, bloody diarrhoea, and especially epidemic shigellosis (which requires a specific response, including adapted antibiotics), may go unnoticed in mortality studies.
- Malnutrition itself causes few deaths, even in nutritional crises: most malnutrition-related deaths will be due to severe diarrhoea and dehydration, and others to malaria or acute respiratory infections; indeed, malnutrition is considered to be the underlying cause for almost half of preventable childhood deaths.[49] Malnutrition and mortality rates correlate well in emergencies.[50] Thus, high global acute malnutrition prevalence, greater than 10%, should always be assumed to cause excess mortality, primarily in children.

Information on the causes of infectious disease death in adults is usually of limited operational relevance, with one possible exception: a high proportionate mortality due to diarrhoea is unusual in adults, and indicates that a serious epidemic of either cholera or shigellosis may be occurring. Along with improved water and sanitation services, such a finding should lead to reinforced surveillance and a full epidemiological investigation.

Lastly, it should be noted that causes of death say very little about social, economic and anthropological obstacles on the road from illness to cure: it is very difficult simply based on mortality data to decide whether, for example, new health centres should be opened, or a community sensitisation campaign conducted to encourage patients and caregivers to seek early care at existing facilities.

Distal versus proximate determinants of mortality

There is a limit to the potential benefits of relief, which becomes more and more apparent the closer one gets to the scene of the action. Emergency relief deals with the proximate causes of mortality, but rarely affects the most distal ones. No water and sanitation intervention, however well-conducted, can fully ameliorate the ill-effects of overcrowding and poor site planning, just as no food and nutritional programme can ever replace the naturally balanced and plentiful diet of a population allowed to farm and trade freely. For example, at the time of writing about 1.5 million northern Ugandans are prevented by systematic armed terror from leaving overcrowded and unsanitary displacement sites: even a furtive trip outside the camps, to cultivate once well-tended fields, can result in death, abduction, rape or mutilation.

Beyond the direct effects of violence, displacement and/or a breakdown of health and other infrastructures triggers a myriad of other domino effects, increasing the risk of various fatal and debilitating diseases, and decreasing the probability of recovery. Examples include:

- Tuberculosis (TB) treatment presently requires at least six months of directly observed therapy and follow-up, which most relief agencies are reluctant to commence in the acute phase of emergencies; diagnosing new TB cases, as well as cases of TB treatment failure, requires laboratory facilities.
- The reproductive rate (transmissibility) of measles and meningitis is higher in overcrowded refugee/IDP camps compared to other settings.[51] This means that epidemics tend to spread faster (leaving less time for reaction). It also means that, in order to prevent epidemics through herd immunity, measles and meningitis vaccination campaigns must achieve a higher coverage in refugee camps than they must in, for example, Western Europe.
- The resurgence of sleeping sickness (human African trypanosomiasis) in Sub-Saharan Africa is largely attributed to wars in Angola, the Congos, Sudan and Uganda, where a large epidemic in the late 1980s was probably initiated by imported cases among southern Sudanese refugees.

A review of communicable disease risks in emergencies and of their determinants was recently published by Connolly et al.[52]

Acting upon the findings

Any evidence of excess mortality should lead to a reaction. The amount of evidence needed in order to act depends on (i) how much can reasonably be collected; and (ii) what the data will be used for. When data are insufficient, the criterion for action should be to minimise the risk of overlooking or under-estimating a crisis.

Emergency thresholds revisited

We have stressed the difficulties in interpreting MRs when they are very close to the emergency threshold. It is very important, however, not to lose sight of the fact that any such threshold is (and always will be) merely an arbitrary value, whose role is primarily to provide a framework for evaluating the magnitude of an emergency, and for justifying the implementation of a relief operation. Although the evidence for this is empirical at best, there is a recognition among relief workers that, in most emergencies (apart from those occurring in industrialised countries), even the most efficient and needs-driven relief operation will struggle to bring CMR far below 1 per 10,000 per day: therefore, the emergency threshold, or rather the target of CMR<1 per 10,000 per day (and U5MR<2 per 10,000 per day) is perhaps best viewed as a reflection of the degree to which humanitarian assistance can – and therefore should – minimise excess mortality in affected populations. Thresholds are immensely important alarm bells. However, mortality in excess of baseline expectations is the more fundamental indicator: whether expressed in terms of MRs or actual total excess deaths derived by interpolation or extrapolation, this is actually the most informative measure of the impact of an emergency, and that which best describes the gap which remains to be filled by humanitarian agencies. Any excess mortality should lead to an immediate reaction. The Sphere approach of considering context-specific baseline MR may make more sense as a way to estimate excess mortality than as a starting point for defining thresholds, since the calculation of excess mortality requires knowledge of what the expected non-crisis MR is in the setting in question.

How much evidence is needed?

Mortality data are usually insufficient, yet important decisions must be taken rapidly based on them. How much evidence on mortality is actually needed to justify a reaction largely depends on the programmatic focus of the study.

Magnitude of the emergency. Objectively deciding how much humanitarian and political assistance a population in crisis needs requires an implicit risk analysis. This analysis balances the provider risk of over-estimating the extent of a crisis (and thus deploying an unjustified amount of resources) with the beneficiary risk of under-estimating or failing to detect a crisis, with consequent excess mortality. Both risks are probably maximised when no data are available, and both are reduced by well-conducted mortality studies. The larger the population at risk, the greater the potential consequences of inaction.

A problem arises when mortality data do not cover the entire affected population, and filling all the knowledge gaps would either take too long or is not feasible due to government hostility or lack of access, funding and/or epidemiological know-how. Most in the humanitarian community would probably be prepared to accept a relatively high provider risk, if this meant keeping the beneficiary risk to a minimum. Furthermore, it seems logical that populations should not be punished with a higher beneficiary risk simply because data on their health status cannot be collected. How much evidence is needed is, therefore, a question of how much can reasonably be collected without putting on hold an urgent humanitarian response. This done, the programmatic interpretation should, as suggested above when discussing bias and confidence intervals, be cautiously conservative, leading to action based on worst-case scenarios, while recognising that this entails a significant provider risk.

Monitoring trends. If a relief operation is ongoing, and individual sites within the target area continue to report high MRs, it would be reckless to scale down programmes simply based on qualitative impressions. Indeed, the burden of proof during monitoring is on the side of the provider and its donors, who should only justify a reduction in assistance based on evidence of decreased mortality everywhere. Common sense suggests that sites which cannot be monitored due to insufficient resources or inaccessibility are more likely to experience elevated mortality – all the more reason to make conservative assumptions when data are missing.

Advocacy. Where populations in crisis are inaccessible, mortality studies can sometimes open a small window on their plight: for example, surveys of small groups of Burmese Karen refugees in Thailand and North Korean migrants to China have corroborated qualitative information about human rights abuses in Burma and famine in North Korea.[53] At the very least, such studies confirmed the existence of phenomena vehemently denied by the respective governments. Mortality studies have occasionally been able to estimate the full impact of a crisis, as in Kosovo and the DRC.[54] Between these two extremes lies a wealth of fragmentary information produced by site- and period-specific studies which, were it assembled in one document, would demonstrate, for any country and setting, the striking extent to which war and socio-economic vulnerability to disasters have both direct and indirect effects on human health.[55]

If mortality studies are to be used as an advocacy tool, then they should be designed to be as representative as possible of the entire population and period during which the crisis (or, more specifically, the violence) took place. Failing this, epidemiology's achievement may simply be to confirm that a certain threshold of acceptability has been crossed, for which those who initiated the crisis, or failed to prevent it, should now be held accountable. Such epidemiological findings can then be integrated into a wider appraisal of the crisis, which combines hard data and qualitative information.

Predicting mortality in time

When emergencies strike, the question on many minds is 'how many people will die if no help reaches them in time?'. There is an understandable and even healthy wish to use past experience to forecast the evolution of mortality. Past attempts to do this have, however, met with little success. For example, in 1996 allied forces invaded Hutu refugee camps in eastern Zaire (ostensibly to drive out the Rwandan *génocidaire* militia based there), and denied access to humanitarian agencies. MSF epidemiologists estimated that CMR would climb to ten per 10,000 per day, and that 13,600 people would die within three weeks, a claim that led to a call for military intervention. The claim was in fact wrong: in reality, the diversion of aid by the militia controlling the camps had created a false perception of how much assistance was necessary to maintain the refugees. In July 2003, the United States Agency for International Development (USAID) predicted that, in the absence of significant aid, CMR in Darfur would eventually climb to 20 per 10,000 per day.[56] CMR thankfully did not reach such levels. Although these particular attempts may have been somewhat rudimentary, it is conceivable that, in the future, through a combination of mathematical modelling, advanced demographic techniques and expert feedback, realistic projections of mortality scenarios in a crisis (or at least of a likely range of outcomes) could inform advocacy for more assistance and access to affected populations.

Chapter 5
The politics of mortality

Numbers have political uses in many facets of our lives, from unemployment and other economic indicators to numbers of asylum-seekers. It is no surprise that figures describing the number of people dying in emergencies will also be used for political ends. The idea of people dying, especially non-combatants dying in times of conflict, always provokes sympathy in onlookers. Prolonged violent conflict almost always results in excess mortality, either directly from the violence, or indirectly, from economic and social disruption. Those who are responsible for the conflict or who are benefiting from it will want to downplay the issue of civilian deaths. Those who oppose a conflict will want to highlight the human cost.

Unlike deaths from old age and chronic diseases, deaths resulting from the direct and indirect consequences of violence are best prevented by the cessation of the violence. Thus, public health workers pursuing the fundamental task of averting death in times of war often find their work and recommendations more politically oriented than they are in non-conflict settings. Moreover, virtually all public health researchers begin studying a pattern of disease or death because they are fundamentally opposed to those episodes of disease or death occurring. While this approach seems scientific when discussing a specific pathogen, researchers who set out to document the ill-effects of war are often accused of having a political agenda. By implication, their work is suspect.

The political problems associated with mortality data can be divided into two types: problems associated with data collection and acceptance of the findings; and problems associated with the exaggeration or minimisation of the results. While these problems can often arise with surveillance data, the discussion here focuses on mortality survey data, as it is most commonly at the centre of political controversies.

Bias and manipulation

Following the Gulf War of 1991, the destruction induced and the economic sanctions imposed resulted in concern in some quarters for the health of the Iraqi people. A research team from Harvard University hired and trained young, well-educated Jordanians to collect a sample of data from within Iraq in an attempt to measure under-five mortality. The results, published in *The Lancet*, implied that hundreds of thousands of children had died as a result of the invasion and the disruption that followed.[57] Post-publication, a review of the data showed that one interviewer recorded many, if not most, of the excess deaths. An embarrassing retraction was requested by a Harvard review committee.

Interviewer-induced biases can also work in the opposite direction. In 2001, the International Rescue Committee conducted a mortality survey in Kalima health zone in the DRC. Investigators hired two interviewers suggested by Oxfam who could speak some French. The investigators quickly became concerned about the rigour of one of the interviewers, who found only child deaths. He was therefore watched closely. On the second day, an investigator overheard the suspect interviewer being told of a violent death. When the data forms were handed to the investigator, that death was not listed. A second visit to the house confirmed that the death had indeed occurred. It was decided that the interviewer would not be confronted, but all his data sheets were excluded from the analysis. It was subsequently discovered that the suspect interviewer lived with the rebel chief of security in Kalima, and that Oxfam had been manipulated into suggesting this individual. Another member of the household later attempted to give diamonds to the investigators, presumably to get them arrested.[58] Investigators suspect that the interviewer had attempted to under-report deaths to minimise the evidence of high mortality in an area where the rebel government was both brutal and unpopular.

This sort of bias may also be induced by the interviewees. Data collected in 1999 during an immunisation survey in Katana health zone in the DRC showed that mothers who approached interviewers in public areas were twice as likely to have immunisation cards as mothers who were found at their homes. Thus, the availability or eagerness of people to be interviewed can induce a bias which shows the importance of rigorous survey-enrolment procedures. Likewise, Taylor and Becker have both highlighted the under-reporting of infant and child deaths during surveys.[59] In particular, when mothers recalled all the children born to them, they excluded most births that resulted in an infant death if the woman being interviewed was over 19 years of age. It is believed that older women who have experienced more births and who had to recall births more than a year or two years before simply did not, knowingly or unknowingly, report most infant deaths.

While these two examples may not constitute politically-motivated under-reporting, the potential exists for gross under-reporting for political purposes. For example, a survey in the DRC by the International Rescue Committee in 2002 found under-five mortality of 5.8 per 1,000 per month, much higher than the 3.4 per 1,000 per month reported by a UNICEF Multiple Indicator Cluster Survey (MICS) conducted the same year.[60] The head of UNICEF's MICS programme believed that this discrepancy arose because the MICS survey, which asked mothers about births and under-five deaths among their family in the past five years, suffered from the same under-reporting described by Taylor a decade earlier. While few public health scientists who read both reports believed the UNICEF finding, the DRC government cited only the UNICEF figure because the IRC finding would have meant that the

DRC had the world's highest under-five mortality for a country of 30 million people or more. Here, under-estimates produced by interviewees' under-reporting served a national political agenda. Conversely, it may be politically expedient to exaggerate the extent of mortality. For example, findings from a small sample of North Korean families were used to support claims that perhaps three million famine-induced deaths had occurred in North Korea in the early 1990s.[61] While the investigators did not stand by this finding, it was widely used by critics of North Korea as evidence of governmental incompetence and wrong-doing.

Awkward findings may simply be suppressed or rejected. Towards the end of the Nigerian civil war of 1966–70, a young Epidemic Intelligence Officer from the US Centers for Disease Control, Carl Western, conducted a convenience survey which estimated that half of the population in Biafra – perhaps 1.5 million people – had perished after the borders were sealed.[62] To our best knowledge, the data were never published in a peer-reviewed journal. At the end of the conflict, US diplomatic efforts were focused on reconciliation, and on preventing Nigeria from falling under Soviet influence. In terms of the percentage of the population that died, the Biafran conflict may have been the deadliest in recent history, but evidence of this was largely set aside, presumably for diplomatic reasons. IRC's surveys in DRC in 2000 and 2001 were the only evidence provided by the Kinshasa-based government when it accused the invading armies of Rwanda and Uganda of genocide. This was despite clear statements in the reports that both sides were significantly responsible for the violence, and that the vast majority of deaths were indirect and due to infectious diseases.[63]

More recently, there has been significant dispute over the number of civilian deaths in Iraq following the US-led invasion in March 2003. Since the conflict began, a web-based network called the Iraq Body Count (IBC) has been attempting to record civilian deaths from violence.[64] By its own admission, the IBC's estimates are likely to be on the low side. Its coordinators suggest that the IBC monitoring system detects at least 25% and probably more than half of all civilian fatalities (source: John Sloboda, pers. comm.). However, a population-based survey published in *The Lancet* estimated that 100,000 excess deaths had occurred in the first 18 months after the invasion.[65] While the *Lancet* figure included excess deaths from all causes, not just violence, thereby accounting for some of the difference with the IBC estimate, its estimate of violent deaths was at least four times higher. Supporters of the war and government officials cited the IBC number of 17,000 as more credible. British Prime Minister Tony Blair publicly stated that one could not 'extrapolate' from a small sample, and that the number of deaths could only be known by counting.[66] Pundits have also attempted to minimise the political impact of the *Lancet* study.[67]

Table 6 shows various estimates of the number of violent deaths occurring in occupied Iraq. The first three are surveillance-based, the second three survey-based. The last reference by the Iraqi Kaffi is reported as a door-to-door tally from the first six months of the conflict. As the table shows, the IBC figure is among the low-end estimates, but it remains the primary death toll cited by the press in the US.

This tendency to minimise the adverse effects of conflict by those who induced them is not new. The German death toll from the First World War rarely includes deaths due to famine in the years that followed, which would roughly double the figure. This famine is believed to have been the direct result of a British and French embargo on imports of seeds and fertilizer into Germany.

Minimising manipulation

Given the innately political nature of mortality data during conflict, the onus is on the investigator to anticipate and minimise the potential for misuse of the data. The key here is to predict the nature of the problems which will arise over the course of data collection and dissemination. A key approach often employed is to conduct a sensitivity analysis by reporting the results given a range of assumptions. For example, a major weakness of many conflict-based studies is that some percentage of the population is inaccessible or not at home. Investigators may explore the potential effect of this by assuming the best and the worst plausible experiences in those populations, and showing how these assumptions would affect the overall conclusions. Likewise, a detailed self-critique of methods can be useful in curbing misinterpretations. Below are specific

Table 6: Estimates of violent deaths per day in occupied Iraq

Source	Date of information	Violent deaths per day implied
Iraq Ministry of Health[68]	5/4/04 – 5/05	22
Iraq Body Count	1/3/03 – 1/2/05	32
NGO Coordination Committee of Iraq (unpublished)	2004	50
Iraq Multiple Indicator Rapid Assessment (IMIRA)[69]	1/3/02 – 30/5/04	56
Lancet research (violent deaths only)[70]	1/3/03 – 21/9/04	101
Mental health study, 2004[71]	2003 – 2004	133
Iraqi Kaffi	3/03 – 10/03	152

problems, and measures to reduce those biases in the mortality reporting process.

- **Interviewers are making up deaths.** Have investigators with interviewers as much as possible, make sure that some trusted investigators are locals or fluent in the local dialect, investigators are shown a death certificate or a grave for a sample or all decedents, analyse data to see if some interviewers have disparate results.
- **Interviewers are hiding deaths.** Have investigators with interviewers as much as possible, make sure that some trusted investigators are locals or fluent in the local dialect, revisit a subset of households with trusted or new interviewers, attempt to confirm findings with local official records (which, however, almost always underestimate mortality).
- **Those who need to act will not accept the results.** Avoid inflammatory language or recommendations, follow a peer-reviewed publication process, develop a cadre of experts who can explain or defend the data, be conservative in the reporting of the data.
- **Groups are taking data out of context and overstating the results.** Be conservative in the presentation of results, have a detailed limitations section describing how the data should not be interpreted or used, develop a cadre of experts who are familiar with both the data and the methods, who can explain or defend the data and publicly correct any misuses of the data.

Final thoughts

Measuring mortality in times of war is a complex and politically volatile operation. Leaders in the field suggest that one side is usually responsible for most civilian deaths.[72] Thus, the results of mortality surveys often have the effect of accusing a particular group or regime of murder. Adding to the volatility of the subject is the fact that no organisation is charged with documenting deaths in times of war. Those that do so are usually self-selected, driven by humanitarian imperatives, and working for a non-governmental agency with limited political clout. In most major conflicts a death toll is never established while the fighting is under way. Thus, not only is the subject of war deaths political in nature, but the media and world community at large have little opportunity to digest and respond to death-toll information.

For some, the political nature of mortality is itself an argument against acquiring data on it. But if the purpose of public health data is to avert death and suffering, wartime mortality data should have a higher priority than data showing the number of cases in a disease outbreak, or other epidemiological data commonly collected. This is because war-related deaths are all within human control, but the field of war epidemiology is so nascent that the tools for preventing these deaths, be they legal, military or media-based, are largely undeveloped. The political attention given to estimates of conflict-induced death has the potential to set precedents and influence future conflicts and the development of International Humanitarian Law. At present, when water supplies are destroyed or people flee from their homes and die from the indirect consequences of violence, there is no mechanism by which international law can try the perpetrators. Because indirect deaths usually outnumber violent deaths, this means that most of the mortality induced by combatants is ignored. Through the consistent and rigorous collection of mortality estimates, the perpetrators of war can be held responsible for their actions, and the wealthy societies that sponsor wars in places like Chechnya and Iraq may for once be brought to understand the consequences of their actions.

Chapter 6
Conclusion

Is there a right to good data?

In February 2005, a report by the *Washington Post* rightly referred to mortality estimates in Darfur as 'statistical anarchy'.[73] At the time of writing, disparate attempts, essentially based on extrapolation and educated guesswork, have been made to quantify the death toll from that conflict.[74] In this confusion of figures, we may perhaps lose sight of two critical points. First, these rough estimates are the response of civil society to a government's denial of killing, rape, pillage and other atrocities conducted by militia operating with its connivance and/or open support; they are an attempt to quantify the humanitarian disaster caused by these attacks, measure the level of violence, and thereby give some voice to the survivors. Second, 'statistical anarchy' is inevitable when governments discourage or openly obstruct the objective assessment of the health status of populations in crises, and when combatants threaten relief workers[75] or intimidate victims.[76] Preventing the collection and dissemination of objective data on the plight of a conflict-affected population, or manipulating existing figures, are stratagems through which belligerents try to forestall public scrutiny of their actions and intents. Missing and misleading data can be as deadly as shells. In future crises, we must demand more strongly unfettered access for the scientific assessment of what affected populations have undergone, and what they require for dignified survival.

Mechanisms for collecting mortality data

Even where there is reasonably good access, resources for surveys are sometimes allocated inadequately and too late; good-quality prospective surveillance systems are usually not considered a programmatic priority; and relief agencies may act without coordination and without proper reflection about what sort of data they need. Mechanisms should thus urgently be put in place to:

- establish the magnitude of an emergency from the outset;
- implement mortality surveillance systems and/or regular monitoring surveys; and
- where humanitarian assistance arrives too late, document the past health impact of any crisis.

The kind of comprehensive, geographically representative data necessary to evaluate a relief operation in its entirety can best be collected when agencies work in coordination and pool their resources, while assigning the implementation of epidemiological work to a reputed, impartial research body. Well-planned, sufficiently sampled surveys covering the entire population affected by a crisis will remove the need for risky extrapolation of scarce site-specific mortality findings. The budgets of large relief operations can easily run into tens of millions of dollars a year; even a region-wide comprehensive survey will rarely cost more than $100,000, and will usually cost less. It does not seem unreasonable to use 1% (or even 5%) of a relief budget to help assess whether the other 99% (or 95%) is having an impact. The point is not to do many surveys, but to do a few well: ideally, epidemiologists should aim for redundancy.

Who should carry out such surveys? Ideally, documenting death and suffering occurring in or induced by UN member states should fall within the functions of UN agencies themselves. In the past, the UN has largely not fulfilled this role. NGOs and academic centres have partly filled the gap, establishing a proven track record in implementing surveys and surveillance systems. Recently, however, the World Health Organisation has successfully led region-wide mortality assessments in both Darfur and northern Uganda. Direct UN involvement may not always be possible, especially if the survey objectives include documenting past violence and abuses: typically, this is due to stringent security rules or the vital need to maintain relations with belligerent parties. In such cases, however, the UN should actively support and empower an impartial, reputable agency or academic institution to collect mortality data, and insist on seeking out full documentation on the impact of the crisis.

A standardised approach

As the centrality of mortality figures is increasingly accepted, an increasing onus is placed on epidemiologists and agencies sponsoring their work to collect good-quality data. The consequences of bad science can be counted in human lives when, on the basis of incorrect findings, agencies or donors decide to scale down or abandon life-saving activities, or allocate them improperly. Standardisation of methods is probably the best guarantee against biased, imprecise and otherwise contestable results, especially if more agencies become involved in the collection of mortality data. We have attempted to lay out certain core requirements of any mortality-data collection, such as the proper training of home visitors, regular updating of population figures and an adequate sample size. The Standardised Monitoring and Assessment of Relief and Transitions (SMART) initiative is developing guidelines and tools for mortality and other emergency assessments, and has published a step-by-step protocol for nutrition and mortality assessments.[77] Such initiatives are welcome, but the challenge will lie in disseminating these standards.

Making sure that there is professional capacity to collect and interpret valid health data is as important as standardising methods. Experienced epidemiologists are in short supply: as with other health professionals, many tend to leave the humanitarian sector for more stable work. In this respect, not-for-profit institutes with a strong link to field operations and proven scientific expertise, such as Epicentre or some academic centres of excellence, are desperately needed to infuse experience and judgment into organisations wishing to collect and use mortality data.

Reporting and interpretation

We have attempted to provide basic suggestions for interpreting point estimates and confidence intervals, performing interpolation and extrapolation, and acting on the basis of imperfect data. There is, however, a need for further discussion and consensus-building among epidemiologists and relief programme managers, to ensure that consistency and scientific rigour prevail in the reporting and use of mortality data. Likewise, resources for fast expert review of study protocols and reports should be made available to less experienced research teams.

Further discussion on emergency thresholds will also be useful, since there is an obvious divergence of approach (for example, between Sphere and other institutions): standardising these thresholds internationally can only be beneficial and guarantee equal assessment of any crisis. At the same time, the discussion on thresholds must not divert attention from the primary aim of humanitarian programmes, namely reducing *any* excess mortality. In summary, thresholds are still needed so as to set targets that relief operations must be geared to reach. Data on excess mortality, however, are equally needed so as to highlight the limits of relief and the need for political action to address the root causes of mortality, as well as giving a measure of the gap still to be bridged.

A call for action

The worst fate of a mortality report is oblivion. We do not know how many of the mortality studies performed in the past have actually led to significant improvements in the health status of surveyed communities. An analysis of these studies' operational impact would be very helpful in clarifying this. What is clear, however, is that in many crises excess mortality remains unacceptably high, and neither humanitarian assistance nor measures to protect civilians can be considered sufficient. In the case of Darfur, the DRC, Uganda and many other emergencies, the evidence of mortality has failed to elicit an appropriate international response.

Nevertheless, epidemiology in general, and mortality studies in particular, are increasingly appreciated as key tools to guide humanitarian action and foster respect for humanitarian law. The UK Department for International Development has instigated an initiative to refine benchmarks for humanitarian action and impact assessment. Concurrently, there is a drive for the creation of a Humanitarian Severity Index, in which mortality data, standardised and understood by all concerned actors, will be fundamental.

The fact that relief efforts may be unfailingly evaluated by a professional financial accountant but only seldom by a project evaluation specialist can only decrease the effectiveness of aid agencies, and lessen their credibility. Similarly, the incomplete quantification of government-sponsored violence in places like Chechnya and Darfur is a major blemish on the record of the international community. We can easily do better, and recording the rates and circumstances of deaths during all crises seems like a minimal start. We hope therefore that future years will see more resources allocated in a timely fashion to operational research, and that mortality findings in emergencies will help to hold combatants, host governments, relief agencies, donors, international governments and the media accountable for their failures to respect, protect and assist affected populations. Equally, these studies may document successes in the management of crises. Mortality-data collection should continue to be conducted in a spirit of impartial and needs-driven humanitarianism, with the primary purpose of improving assistance to populations in need. The responsibility for properly interpreting and using these data lies, not just with scientists, but with the entire relief community.

Annex 1
Glossary

Some of the terms below have a more general meaning in epidemiology. Here we present their specific meaning in the context of mortality-data collection.

Age-adjusted mortality rate: mortality rate that takes into account the age structure of the population to which it refers. Used to compare mortality in populations with very different age structures.

Age-sex pyramid: graph of the sex and age-group distribution of the population. Used in mortality studies to observe possible alterations in the demographics of the population as a result of high mortality or population loss in a particular age group or sex.

Age-specific mortality rate: mortality rate in a specific age group. See **under-5 mortality rate** for an example.

Baseline mortality rate: mortality rate before the crisis (similar to **non-crisis mortality rate**).

Bias: systematic error during data collection which results in a distortion of the findings (in mortality studies, an over- or under-estimation of mortality).

Case-fatality ratio or **rate:** the proportion of cases of a given disease that result in death. Often abbreviated to CFR.

Cause-specific mortality rate: the **mortality rate** due to a specific disease (e.g. cholera) or phenomenon (e.g. violence).

Child mortality rate: the number of children under five years of age dying per 1,000 live births in a given year.

Cluster sampling: a **sampling design** commonly used in retrospective mortality surveys when comprehensive lists of individual households cannot be obtained. Clusters are groups of households of which the first is chosen at random, and the remainder by a rule of proximity (e.g. second closest). In a cluster mortality survey, 30 or more clusters are usually sampled from the target study population, and each cluster usually contains at least 30 households.

CMR: see **crude mortality rate**.

Confidence interval: a range that expresses the level of approximation, or **imprecision**, around the **point estimate**. Also known as a margin of error. 95% confidence intervals are usually presented: we are thus 95% confident that the true population estimate lies within the range of the confidence interval.

Convenience survey: survey that is not based on a randomly selected, representative sample, but rather on data from households/individuals that can easily be reached or observed (e.g. people standing in a food-distribution queue).

Crude mortality rate: mortality rate among all age groups and due to all causes. Often abbreviated to **CMR**.

Death rate: equivalent to **mortality rate** (some authors prefer the former).

Design effect: phenomenon caused by **cluster sampling**, and which increases the **sampling error** or **imprecision**. Households/individuals within a cluster resemble each other because of their proximity, thus resulting in an overall loss in sampling variability.

Emergency threshold: mortality rate above which an emergency is said to be occurring. Usually taken as a **crude mortality rate** of 1 per 10,000 per day, or as an **under-5 mortality rate** of 2 per 10,000 per day.

Excess mortality, excess mortality rate: mortality above what would be expected based on the **non-crisis mortality rate** in the population of interest. Excess mortality is thus mortality that is attributable to the crisis conditions. It can be expressed as a rate (the difference between observed and non-crisis mortality rates), or as a total number of excess deaths.

Extrapolation: mathematical attempt to extend the findings of a mortality study to a population and/or period that was not represented by the sample.

Imprecision: phenomenon whereby there is a lot of uncertainty or approximation around a **point estimate** obtained from a sample (does not apply to surveillance data which reflect the entire population, or to studies in which each household is interviewed). Imprecision is reflected in the width of the **confidence interval** around the **point estimate**. There is always some imprecision in a result based on a sample. The degree of imprecision of a mortality estimate is determined by **sample size**, length of the **recall period**, **sampling design** and **design effect** (if any), and the **mortality rate** itself.

Infant mortality rate: number of infants below one year old dying per 1,000 live births in a given year.

Interpolation: mathematical process by which mortality-rate findings are applied to the entire population that the study is representative of, so as to obtain total numbers of deaths or percentages of the population that died over the **recall period** investigated.

Maternal mortality ratio: number of women dying of pregnancy-related causes out of 100,000 live births in a given year.

Mid-period population: estimated population at risk at the middle point of the recall period. Used in the routine simplified expression of **mortality rate**.

Mid-point population: see **mid-period population**.

Mortality rate: number of deaths occurring in a given population at risk during a specified time period (also known as the **recall period**). In emergencies, usually expressed as deaths per 10,000 persons per day; alternatively, as deaths per 1,000 persons per month or per year.

Non-crisis mortality rate: mortality rate which would be expected to occur in a given population if there were no crisis.

Non-sampling error: see **bias**.

Non-systematic error: see **imprecision**.

Person-time: cumulative time spent by each individual at risk in the population. Used in the formal expression of **mortality rates**, and necessary when comparing mortality rates in periods that are different in length for each individual in the population (e.g. before versus after arrival to a camp).

Point estimate: most likely value for the parameter of interest (e.g. **crude mortality rate**) obtained through a sample survey. A point estimate should always be accompanied by a **confidence interval**.

Population-proportional sampling: approach to selection of clusters or households to be sampled, whereby more populous sections of the study area are allocated proportionately more clusters or households.

Proportionate mortality: fraction of all deaths due to a specific cause.

Recall bias: bias due to imperfect recall by questionnaire respondents of events in their households. Usually results in an under-estimation of mortality.

Recall period: period of interest in the measurement of a mortality rate, i.e. the interval of time to which the mortality rate in a given population refers.

Reporting bias: bias due to (often intentional) under- or over-reporting of information, such as number of deaths or household size.

Sample size: number of clusters/households/individuals that a survey sets out to include, i.e. interview.

Sampling design: method by which households to be sampled are selected within the target population.

Sampling error: see **imprecision**.

Sampling frame: list of households, or sub-sections of the study area/population, used to allocate clusters or select households to be sampled.

Sampling step: distance between one sampled cluster and the next, or one sampled household and the next, on the **sampling frame**.

Selection bias: type of **bias** whereby a specific kind of household is systematically excluded from the survey, and thus not represented in the results. May result in both over- and under-estimation of mortality.

Simple random sampling: sampling design whereby an individual sampling frame of households is established, and households to be sampled are selected using random numbers.

Spatial sampling: approach to selection of clusters or households to be sampled, whereby clusters and/or households are allocated proportionately to surface area within the study area. Alternative to **population-proportional sampling**.

Stratification: sampling and analysis of sub-groups or sub-periods.

Surveillance (prospective): ongoing collection of epidemiological data, with real-time analysis. Mortality surveillance systems usually rely on home visitors who record deaths in households on a weekly basis.

Survey (retrospective): study of past mortality in a population using a standardised questionnaire that is administered to the entire population or, more commonly, to a randomly selected sample.

Survival bias: type of **selection bias** specific to retrospective surveys, whereby households that disappear during the **recall period** because of the death of all members and consequent disintegration are not represented in the sample. It occurs when high and/or very clustered mortality persists for a long period. Survival bias always results in an under-estimation of mortality.

Systematic error: see **bias**.

Systematic random sampling: sampling design whereby an individual sampling frame of households is established, and households to be sampled are selected using a constant **sampling step** (i.e. every nth household).

Under-5 mortality rate: number of deaths occurring in a given population of under-five children during a specified time period. Often abbreviated to **U5MR**.

Notes

1 R. F. Mollica et al., 'Mental Health in Complex Emergencies', *The Lancet*, 2004;364(9450):2058–67.

2 See 'Niger Leader "Ignorant" of Hunger', *BBC News*, 12 August 2005, http://news.bbc.co.uk/2/hi/africa/4144642.stm.

3 P. Salama, P. Spiegel, L. Talley and R. Waldman, 'Lessons Learned from Complex Emergencies Over Past Decade', *The Lancet*, 2004;364(9447):1801–13.

4 'Niger Leader "Ignorant" of Hunger'.

5 Centers for Disease Control, 'Famine-affected, Refugee, and Displaced Populations: Recommendations for Public Health Issues', *MMWR Recomm Rep*, 1992;41(RR-13):1-76; Sphere Project, *Sphere Handbook*, revised ed., 2004, www.sphereproject.org.

6 M. J. Toole and R. J. Waldman, 'Prevention of Excess Mortality in Refugee and Displaced Populations in Developing Countries', *JAMA*, 1990;263(24):3296–302.

7 UNHCR, *Handbook for Emergencies*, second ed. (Geneva: UNHCR, 2000).

8 D. Guha-Sapir and O. Degomme, *Darfur: Counting the Deaths* (Brussels: Centre for Research on the Epidemiology of Disasters, 2005).

9 P. Salama et al., 'Lessons Learned from Complex Emergencies over Past Decade', *The Lancet*, 2004;364(9447):1801–13; Médecins Sans Frontières, *Refugee Health: An Approach to Emergency Situations* (London: Macmillan Education Ltd, 1997); Epicentre, *Rapid Health Assessment of Refugee or Displaced Populations* (Paris: Médecins Sans Frontières, 1999); UNHCR, *Handbook for Emergencies*, second ed. (Geneva: UNHCR, 2000); Sphere Project, *Sphere Handbook*, revised ed., 2004, www.sphereproject.org.

10 L. Roberts and M. J. Toole, 'Cholera Deaths in Goma', *The Lancet*, 1995;346(8987):1431 ; T. Grein, F. Checchi, J. M. Escriba et al., 'Mortality Among Displaced Former UNITA Members and Their Families in Angola: A Retrospective Cluster Survey', *BMJ*, 2003;327(7416):650.

11 Centers for Disease Control, 'Public Health Consequences of Acute Displacement of Iraqi Citizens – March–May 1991', *MMWR Morb Mortal Wkly Rep*, 1991;40(26):443–7.

12 World Health Organisation, *Health and Mortality Survey among Internally Displaced Persons in Gulu, Kitgum and Pader Districts, Northern Uganda* (Geneva: WHO, 2005).

13 P. S Moore et al., 'Mortality Rates in Displaced and Resident Populations of Central Somalia During 1992 Famine', *The Lancet*, 1993;341(8850):935–8; Goma Epidemiology Group, 'Public Health Impact of Rwandan Refugee Crisis: What Happened in Goma, Zaire, in July, 1994?', *The Lancet*, 1995;345(8946):339–44; P. Nabeth, M. J. Michelet, G. Le Gallais and W. Claus, 'Demographic and Nutritional Consequences of Civil War in Liberia', *The Lancet*, 1997;349(9044):59–60; H. Creusvaux et al., 'Famine in Southern Sudan', *The Lancet*, 1999;354(9181):832; P. Salama et al., 'Malnutrition, Measles, Mortality, and the Humanitarian Response During a Famine in Ethiopia', *JAMA*, 2001;286(5):563–71; N. Ford 'Slow Response to Angola's Food Crisis', *The Lancet*, 2002;360(9329):334–61; E. Depoortere, F. Checchi, F. Broillet et al., 'Violence and Mortality in West Darfur, Sudan (2003–04): Epidemiological Evidence from Four Surveys', *The Lancet*, 2004;364(9442):1315–20.

14 *Sphere Handbook*, 2004; MSF, *Refugee Health: An Approach to Emergency Situations*.

15 P. Salama et al., 'Malnutrition, Measles, Mortality, and the Humanitarian Response During a Famine in Ehiopia', *JAMA*, 2001;286(5):563–71.

16 Moore et al., 'Mortality Rates in Displaced and Resident Populations of Central Somalia'.

17 Mike Toole, personal communication.

18 B. Coghlan et al., *Mortality in the Democratic Republic of Congo: Results from a Nationwide Survey (Conducted April–July 2004)* (New York: International Rescue Committee, 2004).

19 L. Roberts, unpublished observations.

20 L. Roberts, R. Garfield, J. Khudhairi and G. Burnham, 'Mortality Before and After the 2003 Invasion of Iraq: Cluster Sample Survey', *The Lancet*, 2004;364(9448):1857–64.

21 Depoortere, Checchi, Broillet et al., 'Violence and Mortality in West Darfur, Sudan'; F. Grandesso et al., 'Mortality and Malnutrition Among Populations Living in South Darfur, Sudan: Results of 3 Surveys, September 2004', *JAMA*, 2005;293(12):1490–4.

22 N. Nathan, M. Tatay, P. Piola, S. Lake and V. Brown, 'High Mortality in Displaced Populations of Northern Uganda', *The Lancet*, 2004; 363(9418):1402.

23 WHO, *Health and Mortality Survey Among Internally Displaced Persons in Gulu, Kitgum and Pader Districts*.

24 P. B. Spiegel and P. Salama, 'War and Mortality in Kosovo, 1998–99: An Epidemiological Testimony', *The Lancet*, 2000;355 (9222):2204–9.

25 Depoortere, Checchi, Broillet et al., 'Violence and Mortality in West Darfur, Sudan'.

26 Nathan et al., 'High Mortality in Displaced Populations of Northern Uganda'.

27 N. Binkin, K. Sullivan, N. Staehling and P. Nieburg, 'Rapid Nutrition Surveys: How Many Clusters Are Enough?', *Disasters*, 1995;16:97–103; J. Katz, 'Sample-size Implications for Population-based Cluster Surveys of Nutritional Status', *Am J Clin Nutr*, 1995; 61(1):155–60.

28 Depoortere, Checchi, Broillet et al., 'Violence and Mortality in West Darfur, Sudan'.

29 Francesco Checchi, unpublished observations.

30 B. A. Woodruff, 'Measuring Mortality Rates in Cross-Sectional Surveys: A Commentary', *Field Exchange*, 2002;17:16–22.

31 Adapted from D. Legros, C. Paquet and P. Nabeth, 'The Evolution of Mortality Among Rwandan Refugees in Zaire Between 1994 and 1997', in National Research Council (ed.), *Forced Migration and Mortality* (Washington DC: National Academies Press, 2001): 52–68.

32 M. A. Myatt, 'A Method for Estimating Mortality Rates Using Previous Birth History', *Field Exchange*, 2002;17:13–15.

33 W. R. Taylor, A. Chahnazarian, J. Weinman et al., 'Mortality and Use of Health Services Surveys in Rural Zaire', *Int J Epidemiol*, 1993;22 Suppl 1; S. R. Becker, F. Diop and J. N. Thornton, 'Infant and Child Mortality in Two Counties of Liberia: Results of a Survey in 1988 and Trends Since 1984', *Int J Epidemiol*, 1993;22 Suppl 1:S56–63.

34 D. Chandramohan, G. H. Maude, L. C. Rodrigues and R. J. Hayes, 'Verbal Autopsies for Adult Deaths: Their Development and Validation in a Multicentre Study', *Trop Med Int Health*, 1998;3(6):436–46.

35 Standardised Monitoring and Assessment of Relief and Transitions programme (SMART), SMART Methodology, version 1, 2005.

36 Centers for Disease Control, 'Public Health Consequences of Acute Displacement of Iraqi Citizens'.

37 Amnesty International, 'Sudan, Darfur – Rape as a Weapon of War', 2004.

38 V. Brown et al., 'Rapid Assessment of Population Size By Area Sampling in Disaster Situations', *Disasters*, 2001;25(2):164–71.

39 L. P. Boss, M. J. Toole and R. Yip, 'Assessments of Mortality, Morbidity, and Nutritional Status in Somalia during the 1991–1992 Famine: Recommendations for Standardization of Methods', *JAMA*, 1994;272(5):371–6.

40 P. B. Spiegel, P. Salama, S. Maloney and A. van der Veen, 'Quality of Malnutrition Assessment Surveys Conducted During Famine in Ethiopia', *JAMA*, 2004;292(5):613–8.

41 Depoortere, Checchi, Broillet et al., 'Violence and Mortality in West Darfur, Sudan'.

42 Goma Epidemiology Group, 'Public Health Impact of Rwandan Refugee Crisis'.

43 Spiegel and Salama, 'War and Mortality in Kosovo'.

44 Coghlan et al., *Mortality in the Democratic Republic of Congo: Results from a Nationwide Survey*.

45 Roberts et al., 'Mortality Before and After the 2003 Invasion of Iraq'.

46 D. Nabarro, *Sudan: Mortality Projections for Darfur* (Geneva: World Health Organisation, 2004); Guha-Sapir and Degomme, *Darfur: Counting the Deaths*.

47 Grein et al., 'Mortality Among Displaced Former UNITA Members and Their Families in Angola'.

48 K. Kallander, J. Nsungwa-Sabiiti and S. Peterson, 'Symptom Overlap for Malaria and Pneumonia – Policy Implications for Home Management Strategies', *Acta Trop*, 2004;90:211–14.

49 R. E Black, S. S. Morris and J. Bryce, 'Where and Why Are 10 Million Children Dying Every Year?', *The Lancet*, 2003;361(9376):2226–34.

50 SMART Methodology, version 1.

51 R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford: Oxford University Press, 1991).

52 M. A. Connolly et al., 'Communicable Diseases in Complex Emergencies: Impact and Challenges, *The Lancet*, 2004;364(9449):1974–83.

53 F. Checchi, G. Elder, M. Schafer, E. Drouhin and D. Legros, 'Consequences of Armed Conflict for an Ethnic Karen Population', *The Lancet*, 2003;362(9377):74–5; W. C. Robinson, M. K. Lee, K. Hill and G. M. Burnham, 'Mortality in North Korean Migrant Households: A Retrospective Study', *The Lancet*, 1999;354(9175): 291–5.

54 Spiegel et al., 'War and Mortality in Kosovo, 1998–99'; Coghlan et al., *Mortality in the Democratic Republic of Congo*.

55 C. J. Murray, G. King, A. D. Lopez, N. Tomijima and E. G. Krug, 'Armed Conflict as a Public Health Problem', *BMJ*, 2002; 324(7333):346–9; D. Guha-Sapir and W. G. Panhuis, 'Conflict-related Mortality: An Analysis of 37 Datasets', *Disasters*, 2004;28(4): 418–28.

56 USAID, *Projected Mortality Rates in Darfur, Sudan 2004–2005* (Washington DC: United States Agency for International Development, 2004).

57 S. Zaidi and M. C. Smith Fawzi, 'Health of Baghdad's Children: Letter to the Editor', *The Lancet*, 1995;346(8988):1485.

58 L. Roberts and C. Labula, IRC, unpublished, 2001.

59 Taylor et al., 'Mortality and Use of Health Services Surveys in Rural Zaire'; Becker et al., 'Infant and Child Mortality in Two Counties of Liberia'.

60 L. Roberts, P. Ngoy, C. Mone et al., *Mortality in the Democratic Republic of the Congo: Results From a Nation-wide Survey* (New York: IRC, March 2003), http://www.theirc.org/DRCongo/index.cfm.

61 A. Natsios, *The Great North Korean Famine: Famine, Politics and Foreign Policy* (Washington DC: US Institute for Peace Press, 2002).

62 CDC, unpublished data.

63 L. Roberts et al., *Mortality in the Democratic Republic of the Congo*.

64 See http://www.iraqbodycount.org.

65 Roberts et al., 'Mortality Before and After the 2003 Invasion of Iraq'.

66 'Blair Defends Assault on Falluja', *BBC News*, 8 November 2004, http://news.bbc.co.uk/1/hi/uk_politics/3993277.stm.

67 F. Kaplan, '100,000 Dead – Or 8,000: How Many Iraqi Civilians *Have* Died as a Result of the War?', *Slate*, 29 October 2004, http://slate.msn.com/id/2108887.

68 Provided by R. Garfield, a consultant to the Iraqi Minister of Health, 3 June 2005. Confirmed by the *Washington Post*, 3 June 2005.

69 See *Iraq Living Conditions Survey 2004*, UNDP, http://www.iq.undp.org/ILCS/overview.htm.

70 Roberts et al., 'Mortality Before and After the 2003 Invasion of Iraq'.

71 C. W. Hoge et al., 'Combat Duty in Iraq and Afghanistan, Mental Health Problems, and Barriers to Care', *NEJM*, 1 July 2004(351):13–21.

72 C. Bialik, 'Counting the Civilian Dead in Iraq', *Wall Street Journal Online*, 5 August 2005.

73 'Lack of Access Muddies Death Toll in Darfur', *Washington Post*, 7 February 2005, http://www.washingtonpost.com/wp-dyn/articles/A6186-2005Feb7.html.

74 See, for example, E. Reeves, 'Darfur Mortality Update: June 30, 2005', 1 July 2005; Guha-Sapir and Degomme, *Darfur: Counting the Deaths*; 'UN's Darfur Death Estimate Soars', *BBC News*, 14 March 2005, http://news.bbc.co.uk/2/hi/africa/4349063.stm; J. Leaning and M. VanRooyen, 'An Assessment of Mortality Studies in Darfur, 2004–2005', *Humanitarian Exchange*, no. 30, June 2005; 'How Many Have Died in Darfur?', *BBC News*, 16 February 2005, http://news.bbc.co.uk/2/hi/africa/4268733.stm.

75 'Sudan Charges MSF Man Over Report', *BBC News*, 30 May 2005, http://news.bbc.co.uk/2/hi/africa/4593443.stm.

76 Médecins Sans Frontières, 'The Crushing Burden of Rape – Sexual Violence in Darfur', 2005.

77 See http://www.smartindicators.org.