Using design effects from previous cluster surveys to guide sample size calculation in emergency settings

Reinhard Kaiser MD MPH, Bradley A. Woodruff MD MPH, Oleg Bilukha MD PHD, Paul B. Spiegel MD MPH and Peter Salama MBBS MPH[1]

A good estimate of the design effect is critical for calculating the most efficient sample size for cluster surveys. We reviewed the design effects for seven nutrition and health outcomes from nine population-based cluster surveys conducted in emergency settings. Most of the design effects for outcomes in children, and one-half of the design effects for crude mortality, were below two. A reassessment of mortality data from Kosovo and Badghis, Afghanistan revealed that, given the same number of clusters, changing sample size had a relatively small impact on the precision of the estimate of mortality. We concluded that, in most surveys, assuming a design effect of 1.5 for acute malnutrition in children and two or less for crude mortality would produce a more efficient sample size. In addition, enhancing the sample size in cluster surveys without increasing the number of clusters may not result in substantial improvements in precision.

Keywords: cluster surveys, cross-sectional studies, heterogeneity, mortality, nutrition surveys, sample size

Introduction

Cluster sample surveys have become a standard method of assessing health and nutrition status in emergency situations when populations are large and no sampling frame exists to permit simple or systematic random sampling (Lemeshow et al., 1985). Since the 30 cluster by 30 children (30 by 30) standard survey design for measuring acute malnutrition was recommended in 1994 (ACC/SCN, 1995), humanitarian agencies increasingly have used this method in emergency environments, and its application has been expanded to measure morbidity (Rothenberg et al., 1985; Cutts, 1988) and mortality outcomes (Spiegel and Salama, 2000). Such use raises concern about the precision of the estimate when sample sizes are calculated.

Cluster sampling is defined as probability sampling in which sampling units at some point in the selection process are collections, or clusters, of population elements (Kalton, 1983). Because cluster sampling results in greater statistical variance, and therefore less precision than simple random sampling when calculating sample size, the size of the sample calculated for simple random sampling must be enhanced by a factor called 'design effect'. Design effect is defined as the ratio of the variance of the estimate under the actual design used to produce the estimate to the variance of the estimate assuming the same data to have come from a simple random sample. In practice, the survey planner can calculate the sample size for simple random sampling and then multiply it by the estimated design effect. Consequently, the design effect is much more useful under field conditions than other measures of increased variance in cluster design, such as the rate of homogeneity, pair wise odds ratio or coefficient of variation (Bennett et al., 1991; Katz and Zeger, 1994; Katz, 1995).

The standard nutrition cluster survey design described above assumes a design effect of two for acute malnutrition in children less than five years of age. However, the design effects for acute malnutrition, as well as for other outcomes, have not been thoroughly evaluated. They may be lower than two for nutrition or substantially higher for other outcomes (Katz and Zeger, 1994). Despite this uncertainty, a reasonable estimate of design effect is necessary to calculate the most efficient sample sizes and to avoid wasting time and scarce resources in emergency situations. Yet, examples of design effects for relevant outcomes from emergency assessment surveys are not common. Calculations of design effects often are done poorly, design effects are not documented in survey reports, or analyses do not account for cluster methods, resulting in artificially narrow confidence intervals.

The objective of this paper is to review design effects in selected cluster surveys that examined selected nutrition and health outcomes in humanitarian emergency contexts. This information then can be used to guide decisions about sample size, number of clusters and mean cluster size vis-à-vis future cluster surveys and to improve assessment methods in humanitarian emergencies (SMART, 2002; Sphere Project, 2004).

Methods

We selected seven nutrition and health outcomes from nine population-based cluster surveys conducted between 1994 and 2003 in five emergency settings: Afghanistan; Ethiopia; Kosovo; Mongolia; and Zaire (now Democratic Republic of the Congo). Only two of the surveys are published in the biomedical literature (Spiegel and Salama, 2000; Salama et al., 2001). The outcomes examined in these surveys included acute malnutrition in children aged 0–59 months or 6–59 months (n=9); crude mortality (n=8); mortality in children under five years of age (n=6); acute respiratory disease in the two weeks before the survey (n=5); diarrhoea in the two weeks before the survey (n=5); measles vaccination coverage in children aged 0–59, 6–59 or 9–59 months (n=6); and anaemia in children aged 6–59 months (n=4).

Acute malnutrition in children was defined as a weight-for-height z score of less than -2.0 or nutritional oedema. Z scores were derived from the National Center for Health Statistics/Centers for Disease Control and Prevention (CDC)/World Health Organization reference population. Crude mortality was defined as the number of people who died in a given time interval during the emergency divided by the total population at mid-interval. Mortality in children was defined as the number of children less than five years of age who died in a given time interval divided by the population less than five years of age at mid-interval. The cumulative incidence of acute respiratory disease and diarrhoea was measured retrospectively for the two weeks prior to the start of the survey. Measles vaccination coverage was defined as the percentage of children of the targeted age group who had ever been vaccinated against measles before the day of the survey, confirmed by either vaccination card or reported by parents or a caregiver. Anaemia in children aged 6–59 months was defined as haemoglobin concentration less than 11 grams per decilitre (g/dL).

Epidemiologists from the International Emergency and Refugee Health Branch, Division of Emergency and Environmental Health Services, National Center for Environmental Health, CDC, were directly involved in all of the surveys and reassessed the data for this study. The criteria for selection of surveys included a minimum of 25 clusters, well-defined survey populations, and a well described procedure for sample selection and sample size calculation.

Statistical analyses were conducted using Epi Info (CDC, 2001) and SAS Institute, Inc. (SAS, 1999–2000) software. All samples were self-weighting. Design effects and standard errors for cluster design were derived from C-sample in Epi Info (CDC, 2001), without using weighting or stratification. For stratified surveys, each stratum was analysed as a separate survey as long as it met the selection criteria described above.

To measure the effect of increasing the sample size on the statistical precision of an estimate of mortality rates in a methodological simulation, we took systematic random samples of 75 per cent, 50 per cent and 25 per cent of households included in the mortality survey in Kosovo, while leaving the number of clusters unchanged. In addition, we examined the impact on the design effect of changing the number of clusters by drawing systematic random samples of 40 and 30 clusters in the Kosovo survey and then randomly reducing the sample size to approximately 50 per cent of the original size to make them comparable to the 50 per cent sample with 50 clusters.

Finally, to examine whether the effects of reducing sample size in the Kosovo survey would be similar in a survey in an emergency setting in a developing country, we selected a systematic random sample of 50 per cent of the households that participated in the 2002 survey in Badghis, Afghanistan, and calculated the resulting precision around the estimate of crude mortality.

Results

All of the selected surveys had 30 clusters, except the Kosovo one, which comprised 50 clusters. Table 1 shows for each outcome the predicted and observed design effect, and the main factors affecting design effects. Because most surveys investigate more than one outcome, sample size calculation usually is limited to the main outcomes of interest, often malnutrition and anaemia in children, and crude mortality. The assumed design effects used to calculate sample sizes for the surveys included in this study differ for various reasons. Nutrition and mortality surveys conducted in Ethiopia from 2003 used the 30 clusters by 30 children standard method, as recommended by national guidelines (ENCU, 2003). In Afghanistan, because the observed design effect for acute malnutrition in children was less than two in Badghis Province, the sample size calculation for subsequent surveys assumed a design effect of 1.5. Assuming that mortality was more clustered than malnutrition, investigators in Kosovo and Ethiopia predicted design effects of four for crude mortality. The unusually high design effects of 10 and above in Mongolia for malnutrition and anaemia in children came from previous surveys; however, the sample sizes calculated from these assumed design effects were not feasible and were not used to determine the final sample size (Table 1).

The cluster sizes for crude mortality were generally much higher (mean of 120 respondents) than the cluster sizes for other outcomes measured in these surveys (means ranging from 10–21 respondents) because the sample from which crude mortality was calculated included all members of selected households. The sample from which the other outcomes were calculated included only household members who were in the age- and sex-specific target groups appropriate for each outcome, such as children aged 6–59 months and women aged 15–49 years.

Prevalence and incidence rates varied according to the nature of the outcome and the severity of the emergency. For example, death rates in Goma, Zaire, in 1994 (Group, 1995), and Gode, Ethiopia, in 2000 (Salama et al., 2001), substantially exceeded emergency benchmarks of 1/10,000/day for crude mortality and 2/10,000/day for mortality among children under five years of age (Sphere Project, 2004), whereas other surveys yielded mortality rates below these yardsticks (Table 1). Other outcomes also showed considerable variability; for example, prevalence rates of acute malnutrition in children ranged from 0.9 per cent in the areas of Mongolia not affected by severe winter weather in 2001 to 29 per cent in Gode, Ethiopia, in 2000.

The median observed design effect for acute malnutrition in children in the nine surveys was 1.4, and the observed design effect was less than two in seven (78 per cent) of nine surveys. Four (50 per cent) of the eight surveys that assessed crude mortality revealed observed design effects of less than two, and the median observed design effect was 1.7. Observed design effects for mortality and anaemia in children were most frequently below two (median 1.25 and 1.5, respectively). However, the observed design effects for communicable disease outcomes and measles vaccination coverage were consistently much higher (Table 1); median observed design effects for the cumulative incidence of acute respiratory disease and diarrhoea and measles vaccination coverage were 3.6, 3.4 and 5.05, respectively.

Among 15 measurements of outcomes for which design effect had been assumed to calculate sample size, the resulting design effect was higher than anticipated for only two (13 per cent) (Table 1). No design effects, though, were assumed for the cumulative incidence rates of acute respiratory disease and diarrhoea and measles vaccination coverage—those outcomes with the highest observed design effects in this study. Observed design effects ranged from 0.8–4.4 across all outcomes except for measles vaccination coverage, which spanned a considerably larger range: 2.3–11.8. Observed design effects for anaemia and mortality in children yielded the lowest variability (range of 1.4–2.4 and 0.9–2.1, respectively).

In the survey in Kosovo in 1999, if simple random sampling was used instead of cluster sampling, the standard error for crude mortality decreased as the sample size increased, as expected (Figure 1). However, with the cluster sampling actually done in this survey, this potential increase in precision was almost completely offset by the

Table 1

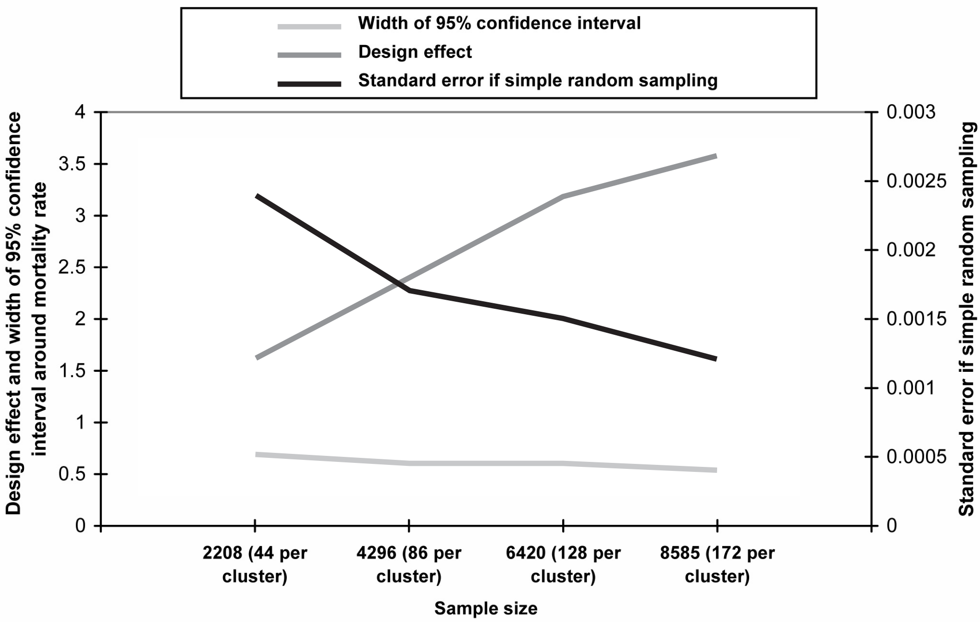
Predicted and observed design effect and main factors affecting design effect, by outcome and selected surveys in humanitarian emergencies, 1994–2003 (ranked by decreasing design effect within outcome)

Outcome	Survey	Design effect predicted (%)	Mean cluster size	Prevalence (%) (for mortality deaths per 10,000/day), 95% confidence interval	Standard error	Design effect observed
Acute malnutrition in children aged 6–59 months (weight-for-height z-score < -2.0)	Ethiopia, Boricha 2003	2	27	7.1, 4.4–9.8	0.0138	2.4
	Ethiopia, Gode 2000	2	29	29.0, 24.6–33.3	0.0219	2.0
	Afghanistan, Badghis 2002 (0–59 months)	2	19	6.5, 3.9–9.1	0.0133	1.6
	Zaire, Goma Mugunga, 1994	N/A*	25	17.8, 13.7–21.9	0.0207	1.6
	Mongolia, 2001—not winter affected	10	15	0.9, 0–1.9	0.0051	1.4
	Mongolia, 2001—winter affected	10	16	1.1, 0–2.2	0.0056	1.3
	Afghanistan, Panjwaye 2002 (0–59 months)	1.5	10	10.4, 6.6–14.2	0.0194	1.2
	Ethiopia, Dubti 2003	2	21	10.1, 7.6–12.6	0.0127	1.1
	Afghanistan, Zabul 2002 (0–59 months)	1.5	11	11.9, 8.8–12.6	0.0158	0.8
Crude mortality	Ethiopia, Gode 2000	4	136	3.2, 2.4–3.8	0.0082	4.0
	Kosovo, 1999	4	172	0.2, 0.1–0.3	0.0023	3.6
	Afghanistan, Badghis 2002	2	105	0.7, 0.5–1.0	0.0045	2.9
	Zaire, Goma Mugunga 1994	N/A	135	45.8, 38.3–53.2	0.0114	2.0
	Afghanistan, Zabul 2002	N/A	74	0.3, 0.1–0.5	0.0025	1.4
	Ethiopia, Dubti 2003	N/A	107	0.6, 0.3–1.0	0.0017	1.3
	Afghanistan, Panjwaye 2002	N/A	49	0.5, 0.3–0.9	0.0038	1.3
	Ethiopia, Boricha 2003	N/A	181	0.4, 0.3–0.5	0.0015	1.1
Mortality in children aged 0–59 months	Ethiopia, Gode 2000	N/A	34	6.8, 5.4–8.2	0.0163	2.1
	Afghanistan, Badghis 2002	N/A	20	2.8, 1.8–3.9	0.0154	1.7
	Ethiopia, Boricha 2003	N/A	33	0.4, 0.3–0.5	0.0066	1.3

	Afghanistan, Zabul 2002	N/A	13	1.2, 0.5–1.9	0.0108	1.2
	Kosovo, 1999	N/A	13	0.2, 0–0.3	0.0031	0.9
	Afghanistan, Panjwaye 2002	N/A	12	1.1, 0.4–1.7	0.0084	0.9
Acute respiratory disease within two weeks of the survey	Afghanistan, Panjwaye 2002	N/A	11	56.4, 45.0–67.8	0.0582	4.4
	Afghanistan, Zabul 2002	N/A	12	54.6, 44.6–64.6	0.0510	3.7
	Kosovo, 1999	N/A	16	31.7, 25.7–37.7	0.0306	3.6
	Afghanistan, Badghis 2002	N/A	18	33.8, 26.3–41.3	0.0383	3.6
	Ethiopia, Gode 2000	N/A	29	21.6, 17.5–25.7	0.0209	2.2
Diarrhoea within two weeks of the survey	Afghanistan, Badghis 2002	N/A	18	29.8, 22.3–37.2	0.0378	3.7
	Afghanistan, Zabul 2002	N/A	12	36.3, 26.8–45.8	0.0485	3.5
	Afghanistan, Panjwaye 2002	N/A	11	52.5, 42.5–62.5	0.0510	3.4
	Ethiopia, Gode 2000	N/A	29	21.2, 16.8–25.5	0.0219	2.6
	Kosovo, 1999	N/A	16	33.5, 28.8–38.3	0.0245	2.2
Measles vaccination coverage in children	Ethiopia, Gode 2000 (0–59 months)	N/A	29	59.7, 48.5–71.0	0.0577	11.8
	Ethiopia, Dubti 2003 (9–59 months)	N/A	20	67.5, 56.3–78.7	0.0571	8.9
	Afghanistan, Badghis 2002 (6–59 months)	N/A	16	52.5, 42.9–62.1	0.0489	5.1
	Afghanistan, Zabul 2002 (0–59 months)	N/A	12	81.0, 71.8–90.2	0.0469	5.0
	Ethiopia, Boricha 2003 (0–59 months)	N/A	27	80.0, 74.8–85.2	0.0255	3.4
	Afghanistan, Panjwaye 2002 (0–59 months)	N/A	11	83.1, 76.9–89.3	0.0316	2.3
Anaemia in children aged 6–59 months (Haemoglobin < 11 g/dL)	Mongolia, 2001—not winter affected	25	16	28.0, 21.8–34.3	0.0321	2.4
	Mongolia, 2001—winter affected	25	15	32.7, 27.3–38.0	0.0270	1.6
	Afghanistan, Panjwaye 2002	1.75	5	63.7, 54.9–72.5	0.0449	1.4
	Afghanistan, Zabul 2002	1.75	6	48.0, 39.3–56.7	0.0444	1.4

Note: * Not available.

Figure 1
Design effect, width of 95% confidence interval, and standard error if simple random sampling was used, as a function of change in sample size, for crude mortality, Kosovo mortality survey, 1999



rise in the observed design effect because of the larger cluster size. This resulted in only a minimal increase in actual precision and only a slight decrease in the width of the confidence intervals around the crude mortality rate as the sample size grew. Analysis of the data for violence-related mortality, which also had a high observed design effect in this survey (results not shown), yielded the same effect. In contrast, non-violence-related mortality did not show clustering, as indicated by an observed design effect of 1.0. For this mortality outcome, increasing the sample size from 2,208 to 8,585 substantially enhanced overall precision, as signified by a narrowing of the 95 per cent confidence interval from 0.06–0.37 to 0.18–0.34, because the observed design effect did not change.

When the sample size remained constant, observed design effects and standard errors for crude mortality and violence-related mortality in Kosovo decreased, with an increase in the number of clusters reducing the mean cluster size (Table 2). Selection of a systematic random sample of 50 per cent of households in the Badghis (Afghanistan) survey cut the sample size from 3,160 to 1,572 and the mean cluster size from 105 to 52. Analysis of this sub-sample demonstrated a decrease in observed design effect from 2.9 to 1.5, whereas the standard error remained unchanged (0.0049), thus confirming the lack of effect on precision of changes in sample size in cluster surveys when the number of clusters remains constant.

Table 2

Effect of number of clusters (sample size constant at 50% of original size) on mortality rate, precision and observed design effect, Kosovo mortality survey, 1999

Outcome	Sample size	Number of clusters	Mean cluster size	Deaths per 1,000/ month*, 95% confidence interval	Standard error	Design effect observed
Crude mortality	4,253 (approximately 50%)	30	142	0.69, 0.29–1.10	0.0035	4.5
	4,344 (approximately 50%)	40	109	0.76, 0.37–1.14	0.0033	3.8
	4,296 (50%)	50	86	0.71, 0.41–1.01	0.0026	2.4
Violence-related mortality	4,253 (approximately 50%)	30	142	0.41, 0.04–0.79	0.0033	6.5
	4,344 (approximately 50%)	40	109	0.50, 0.15–0.86	0.0031	4.9
	4,296 (50%)	50	86	0.48, 0.19–0.77	0.0026	3.5
Non-violence-related mortality	4,253 (approximately 50%)	30	142	0.28, 0.14–0.41	0.0012	1.2
	4,344 (approximately 50%)	40	109	0.26, 0.14–0.37	0.0010	1.0
	4,296 (50%)	50	86	0.23, 0.12–0.35	0.0010	1.0

Note: * To be consistent with the original paper, we report here deaths per 1,000 per month.

Discussion

The surveys included in this paper demonstrate that the observed design effect for acute malnutrition among children in emergency-affected populations is often substantially less than two, the value assumed in the standard nutrition survey design recommended and used by many organisations to evaluate nutritional status in humanitarian emergencies. Consequently, health and nutrition assessment surveys routinely use an unnecessarily large sample size, resulting in the use of time, personnel and resources that might be better employed for other purposes. The assumption in the standard survey design of a prevalence of 50 per cent for acute malnutrition in children exacerbates this apparent inefficiency. Such prevalence rates are seen only in the most severe humanitarian emergencies (CDC, 1993) and may not be appropriate for widespread employment as an assumption in sample size calculations.

Nonetheless, use of the standard 30 by 30 survey design is occasionally necessary in emergency settings where organisations with little epidemiological expertise undertake assessment surveys (Boss, Toole and Yip, 1994; Garfield, 2000; Spiegel et al., 2004). In Ethiopia after the famine of 2002–03, the Ministry of Health promulgated standard guidelines for nutrition surveys, which recommended the 30 by 30 design (ENCU, 2003). Even in this situation, though, these guidelines allowed calculation of a sample size other than 900 children if the necessary technical expertise were available.

Retrospective measurement of mortality in cross-sectional surveys has become common. In surveys that include such measurement, assuming a design effect of two or more for mortality rates may be too high, especially for mortality in children under five years of age, and may lead to a larger sample size than required to achieve the desired precision. Yet, in some situations, mortality rates may display substantially higher design effects of four or more. For example, warfare in the Balkans in the 1990s was marked by ethnic cleansing, which frequently consisted of targeted attacks on specific towns and villages, while others were left intact (Spiegel and Salama, 2000). This pattern of violence caused considerable clustering of violence-related mortality. In addition, because of the lower level of communicable disease, violence was the predominant cause of mortality. Thus, as demonstrated by the results of the Kosovo survey, the design effect for the crude mortality rate was somewhat higher than in other surveys. Similar situations where design effects were affected by violence-related mortality have been recently reported in Darfur, Sudan (Depoortere et al., 2004) and Iraq (Roberts et al., 2004).

In contrast, mortality in humanitarian emergencies in developing countries, where the other surveys in our study were carried out, is more often caused by communicable disease and malnutrition. In such situations, mortality usually may not be as clustered, and mortality rates may not have design effects as high as in Kosovo. The exception may lie in some acute outbreak situations. High non-violence-related mortality characterised the 2000 survey in Gode, Ethiopia, of which a substantial proportion was attributed to measles. Measles incidence rates sometimes may be as heterogeneously distributed as violence-related mortality, potentially resulting in high design effects that may influence the design effects for estimates of crude mortality. Because a higher proportion of children than adults die in developing countries due to the routine causes of mortality (such as diarrhoea, acute respiratory infections, malaria and measles), even during an acute measles outbreak deaths rarely will be as clustered for children as they are for adults. Survey planners need to be aware that, in developing countries, design effects in general tend to be lower for children than for adults.

Unlike the other outcomes considered in this study, the sample size for mortality consists of person time, not just the number of persons. Overestimating the sample size for mortality may result in visiting too many households, which may waste time and resources. Alternatively, the sample size for mortality could be increased by expanding the length of the recall period. However, extending the recall period into the more distant past may produce an estimate of mortality that may not sufficiently represent the current situation. Survey managers must weigh these two methods of adjusting sample size to ensure the survey produces a mortality estimate with both the necessary precision and timeliness to be useful.

The two-week cumulative incidence of diarrhoea and respiratory infection is regularly measured in cross-sectional surveys in emergencies to estimate the contribution of these common illnesses to overall morbidity. In addition, because of the potentially catastrophic consequences of a measles outbreak (Toole et al., 1989), measles vaccination coverage is a critical health parameter to assess early in an emergency-affected population. These measures of morbidity and vaccination coverage appear to be more clustered than malnutrition, non-violence-related mortality and anaemia in children. Since most diarrhoeal diseases and some respiratory diseases are transmitted from person to person, individuals in the same cluster of a survey sample may be more likely to have similar disease status than those in different clusters. Hence, the differences between clusters are high relative to the differences between individuals within clusters, leading to heterogeneous distribution of disease and a high design effect. Measles vaccination coverage from a mass campaign may depend on the distance of the cluster from the vaccination site. In routine vaccination services, coverage will depend on access to health services. Both of these factors may be heterogeneously distributed. Survey planners appraising these outcomes must decide whether high precision is needed for these potentially important health conditions. If so, the sample size may need to be increased to account for the relative high design effect. Yet, in many surveys, these outcomes are considered secondary to malnutrition and mortality, and attempting to achieve high precision may not be an efficient use of survey resources.

Our findings indicate that enhancing the sample size without changing the number of clusters in a cluster survey may increase logistic requirements without substantially augmenting the statistical precision of an estimate of mortality rates. Decreasing the sample size of the Kosovo survey would have resulted in only a small loss of precision, and reducing the sample size in the Badghis survey by one-half would have yielded no loss of precision at all. In the latter survey, the measurement of mortality rates involved taking a complete household census at a specific point in the past and determining the current status of each of these household members. Such a procedure required substantial time at each household. It would have been possible to collect these data at every second selected household and still achieve the desired precision around the estimate of the crude mortality rate while saving substantial time during data collection.

Our results demonstrate that design effects for mortality rates increase, resulting in loss of precision, as the number of clusters for the same sample size decreased from 50 to 30. Previously, 25–30 clusters have been recommended for cluster surveys, considering the additional resources and time required to assess additional clusters (Binkin et al., 1995). However, if more precision is needed, survey planners should consider boosting the number of clusters when they enlarge sample size rather than raising the average cluster size.

This study was based on a limited number of surveys from only nine emergency settings and considered only seven different health and nutrition outcomes. Nevertheless, these outcomes are the most important and most frequently measured in emergency assessment surveys. More surveys need to be examined to confirm these findings, and additional health and nutrition outcomes, including the prevalence rates of malnutrition and anaemia in adults, post-traumatic stress disorder and other mental health indicators, and infection with human immunodeficiency virus (HIV) and other sexually transmitted pathogens, should be included in similar analyses.

Selecting a design effect to obtain the most efficient sample size is more likely if the variability of the design effect in previous surveys has been relatively low, as in this study with the prevalence of anaemia and the mortality rate in children. For outcomes where design effect has shown more variability, such as for measles vaccination coverage, using design effects calculated from other surveys may be less helpful, and a more thorough on-site investigation may be necessary for each survey to estimate the heterogeneity of the distribution of the outcome in the target population. Among other techniques, such an investigation may consist of asking key persons if specific disadvantaged population groups have more or less of the health outcome than more privileged groups or if persons living in remote areas are likely to differ from those living in less isolated areas. If an outcome with a highly variable design effect is critical to the survey, selection of a relatively high design effect may be prudent to ensure sufficient precision.

Reducing the sample size of a survey to the minimum required to attain the desired precision also may enhance the accuracy of data collected. Smaller sample sizes may allow the hiring of fewer more highly qualified survey workers. Furthermore, survey training may be more effectively targeted at fewer survey workers, and supervision during data collection may be better if few survey teams are used. The consequence of underestimating the design effect and employing too small a sample size is an increase in the width of the confidence intervals around the estimates of prevalence or incidence rates. Perhaps more importantly, though, having a substantially larger sample size may result in bias that produces an erroneous outcome, potentially leading to implementation of unnecessary or inappropriate interventions. Because precision can be quantified, survey managers often pay much more attention to precision than to bias. Overall, more realistic assumptions about the design effects used to calculate sample sizes for emergency health and nutrition assessments can save resources, reduce delays in acquiring results and improve the accuracy of such assessments. Enhancing the efficiency and accuracy of health and nutrition assessments is especially critical in humanitarian emergencies where resources frequently are relatively scarce and timeliness is crucial.

Initiatives to improve health and nutrition assessment methods in humanitarian emergencies should incorporate these findings in future cluster survey guidelines (SMART, 2002; Sphere Project, 2004).

Conclusion

We conclude that planners of nutrition and health surveys in emergency-affected populations may save resources and increase the accuracy of survey estimates if they follow three key recommendations. First, survey planners should calculate sample sizes of nutrition and health outcomes instead of using a standard 30 cluster by 30 children sample size. Second, they should carefully evaluate what precision is needed for the objectives of the survey, pay as much attention to preventing bias as to precision, and keep in mind that raising the average sample size in a cluster survey without increasing the number of clusters may not substantially augment the precision of survey estimates. Finally, survey planners should consider design effects lower than two for acute malnutrition among children and mortality in situations without major violence-related mortality or acute communicable disease outbreaks. In cases of violence-related mortality, design effects for crude mortality may be four or higher.

Correspondence

Reinhard Kaiser, MD MPH, Outbreak Preparedness and Response Unit, European Centre for Disease Prevention and Control, Tombtebodavägen 11A, 17183 Stockholm, Sweden. Telephone: +46 708 59 78 25; fax: +46 8 30 00 57; e-mail: reinhard.kaiser@ecdc.eu.int.

Endnotes

1 Reinhard Kaiser MD MPH, Bradley A. Woodruff MD MPH and Oleg Bilukha MD PHD represent the Division of Emergency and Environmental Health Services, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA; Paul B. Spiegel MD MPH represents the United Nations High Commissioner for Refugees, Geneva; and Peter Salama MBBS MPH represents the United Nations Children's Fund, New York, NY.

References

ACC/SCN (United Nations Administrative Committee on Coordination/Sub-Committee on Nutrition) (1995) *Report of a workshop on the improvement of the nutrition of refugees and displaced people in Africa*. ACC/SCN, Machakos.

Bennett, S., T. Woods, W.M. Liyanage and D. L. Smith (1991) 'A simplified general method for cluster-sample surveys of health in developing countries'. *World Health Statistics Quarterly—Rapport Trimestriel de Statistiques Sanitaires Mondiales.* 44(3). pp. 98–106.

Binkin, N., K. Sullivan, N. Staehling and P. Nieburg (1995) 'Rapid nutrition surveys: how many clusters are enough?' *Disasters.* 16(2). pp. 97–103.

Boss, L.P., M.J. Toole and R. Yip (1994) 'Assessments of mortality, morbidity, and nutritional status in Somalia during the 1991–1992 famine. Recommendations for standardization of methods'. *Journal of the American Medical Association.* 272(5). pp. 371–376.

CDC (Centers for Disease Control and Prevention) (1993) 'Nutrition and mortality assessment—southern Sudan'. *Morbidity and Mortality Weekly Report.* 42. pp. 304–308.

CDC (2001) *Epi Info version 6.04d*. Department of Health and Human Services (DHHS), Public Health Service (PHS), and CDC, Atlanta, GA.

Cutts, F.T. (1988) 'The use of the WHO cluster survey method for evaluating the impact *of the expanded programme on immunization on target disease incidence'. American Journal of Tropical Medicine and Hygiene.* 91(5). pp. 231–239.

Depoortere, E. et al. (2004) 'Violence and mortality in West Darfur, Sudan (2003–04): epidemiological evidence from four surveys'. *Lancet.* 364. pp. 1315–1320.

ENCU (Emergency Nutrition Coordination Unit) (2003) *Guidelines for nutrition and mortality surveys in Ethiopia*. ENCU, Addis Ababa.

Garfield, R. (2000) 'Studies on young child malnutrition in Iraq: problems and insights, 1990–1999'. *Nutrition Reviews.* 58(9). pp. 269–277.

Group, G.E. (1995) 'Public health impact of Rwandan refugee crisis. What happened in Goma, Zaire, in July 1994?' *Lancet.* 345. pp. 339–343.

Kalton, G. (1983) *Introduction to survey sampling*. Sage Publications, Thousand Oaks, CA.

Katz, J. (1995) 'Sample-size implications for population-based cluster surveys of nutritional status'. *The American Journal of Clinical Nutrition.* 61(1). pp. 155–160.

Katz, J. and S.L. Zeger (1994) 'Estimation of design effects in cluster surveys'. *Annals of Epidemiology.* 4(4). pp. 295–301.

Lemeshow, S. et al. (1985) 'A computer simulation of the EPI survey strategy'. *International Journal of Epidemiology.* 14(3). pp. 473–481.

Roberts, L., R. Lafta, R. Garfield, J. Khudhairi and G. Burnham (2004) 'Mortality before and after the 2003 invasion in Iraq: cluster sample survey'. *Lancet.* 364. pp. 1857–1864.

Rothenberg, R.B., A. Lobanov, K.B. Singh and G. Stroh (1985) 'Observations on the application of EPI cluster survey methods for estimating disease incidence'. *Bulletin of the World Health Organization.* 63(1). pp. 93–99.

Salama, P. et al. (2001) 'Malnutrition, measles, mortality and the humanitarian response during a famine in Ethiopia'. *Journal of the American Medical Association.* 286. pp. 563–571.

SAS Institute, Inc. (1999–2000) *Statistical Analysis System (SAS), Version 8.01*. SAS Institute, Inc., Cary, NC.

SMART (Standardized Monitoring and Assessment of Relief and Transitions) (2002) *Improving the Monitoring, Reporting and Evaluation of Humanitarian Assistance*. http://www.smartindicators.org/workshop/frame2.htm.

Sphere Project (2004) *Humanitarian charter and minimum standards in disaster response*. http://www.thesphereproject.org.

Spiegel, P.B. and P. Salama (2000) 'War and mortality in Kosovo, 1998–99: an epidemiological testimony'. *Lancet.* 355. pp. 2204–2209.

Spiegel, P.B., P. Salama, S. Maloney and A. van der Veen (2004) 'Quality of malnutrition assessment surveys conducted during famine in Ethiopia'. *Journal of the American Medical Association.* 292. pp. 613–618.

Toole, M.J., R.J. Steketee, R.J. Waldman and P. Nieburg (1989) 'Measles prevention and control in emergency settings'. *Bulletin of the World Health Organization.* 67. pp. 381–388.