36th WEDC International Conference, Nakuru, Kenya, 2013

DELIVERING WATER, SANITATION AND HYGIENE SERVICES
IN AN UNCERTAIN ENVIRONMENT

The issue of the design effect in water, sanitation and hygiene studies

R. Giné, A. Jiménez & A. Pérez-Foguet, Spain

BRIEFING PAPER 1737

Cluster sampling is commonly used in water, sanitation and hygiene (WASH) surveys, as in the Multiple Indicator Cluster Survey developed by the United Nations Children's Fund (UNICEF) for the assessment of development-related goals. In cluster survey techniques, despite a good approximation of the design effect is essential for efficient sample size determination and for obtaining accurate precision of survey estimates; the assessment of this parameter has often been overlooked. This study computes the design effects for three core WASH outcomes at two different administrative scales. We use the database of a Kenyan case study for this purpose. We show that design effects differ greatly, and large differences have been found for different variables, different regional setting, and different scale of analysis. We recommend that survey planners should keep in mind such differences when defining the objectives of the survey and the required precision of survey estimates.

Introduction

In order to assess and monitor the access to safe water and basic sanitation, accurate and reliable information is required. In low-income countries, the information needed has to be provided by means of cost-effective methodologies. An example of such a method is the cluster sampling survey. Specifically, the Multiple Indicator Cluster Survey (MICS) developed by UNICEF (United Nations Children's Fund, 2006) has been widely applied by governments and international agencies to collect social data. A cluster-sampling design, indeed may be the only practical solution where populations are large and geographically scattered (Bennett *et al.*, 1991), or where no exhaustive sampling frames can be constructed to permit simple random sampling (Lemeshow & Stroh, 1988).

Cluster sampling can be defined as any sampling plan in which sampling units are clusters of population elements. Typically, the population is divided into mutually exclusive and exhaustive groups (clusters) based upon geographic or operational criteria. In a household survey, for example, the basic sampling unit is the household (HH), and a cluster may be defined as a community, village or as any other administrative subunit. The process by which a sample of households is selected is stepwise. First, a sample of clusters (e.g. communities) is chosen; and then from each cluster, a sample of households is identified through random probabilistic techniques. Because cluster samples are not as varied as they would be in a random sample, the effective sample size must be enhanced by a factor called 'design effect' (Bennett *et al.*, 1991; Lemeshow & Stroh, 1988; Kaiser *et al.*, 2006). The design effect "*deff*" is specifically the ratio of the variance of the estimate under the sampling method used to the variance of the estimate computed under the assumption of simple random sampling. In practice, the survey practitioner can determine the sample size for simple random sampling and then multiply it by the estimated design effect. A good estimate of this statistical measure is thus crucial in cluster surveys to calculate the most efficient sample size (Kaiser *et al.*, 2006). The value of *deff*, however, varies from region to region, from survey to survey, and from variable to variable; and there is also evidence that *deff* declines with cluster size (Kish, 1980). In consequence, design effects from prior surveys may not be appropriate for a planned survey when any of the key features of the sample design differ (Katz & Zeger, 1994). As a relevant sector reference, the MICS assumes that design effect of most water and sanitation variables may range from 2 to 10 (United Nations Children's Fund,

2006; Howard *et al.*, 2003, draft), which is coherent with the values suggested in a parallel study (Bostoen, 2002). More precisely, the value 4 seems to be widely accepted by experts (United Nations Children's Fund, 2009).

It is noteworthy, however, that the estimates cited above have been computed for national surveys, and determination of *deff* for local surveys remains elusive. The aim of this paper is to tackle this shortcoming, as small contribution to the immense challenge of improving data collection methods in the WASH sector. Based on a dataset produced through a comprehensive household survey in two rural districts of Kenya, we determine the design effects of three core WASH indicators (Joint Monitoring Programme, 2006), namely i) the main source of drinking-water for members of the household, ii) the type of sanitation facility used by adults in the dwelling, and iii) the method employed for water treatment at the point-of-use. This information may be employed to calculate sample sizes more accurately.

Methodology

This study builds on the data from a case study carried out in 2011 by researchers from the Universitat Politècnica de Catalunya (Spain) jointly with UNICEF – Kenya Country Office. Two Kenyan rural districts were initially selected, Homa Bay and Suba, and a separate data collection campaign was planned to evaluate WASH-related variables at the dwelling. In terms of technique, the design and selection of the sample drew on the MICS. When sampling, however, a sample of households was selected from each stratum (stratified sampling), rather than selecting a reduced number of strata, from which identify a subsample of households (cluster sampling). In so doing, the risk of homogeneity within the strata ideally remains relatively low, thus reducing the need for applying large design effects in sample size determination. The sampling plan initially defined sought to achieve estimates to fall within 10 percentage points of the true proportion with 95% confidence, being the design effect estimated as 2. The resulting minimum sample size *n* required in each stratum was 192. In all, 1,157 households were surveyed to cover five targeted administrative divisions at Homa Bay District, while the sample at Suba District included 1,215 households (5 divisions).

The results of the survey were used to determine design effects for a reduced number of indicators at two different administrative scales, namely the division and the location. To do this, estimates of proportions were calculated together with standard errors of those estimates, so that design effects *deff* could be assessed on the basis of its original definition (Kish, 1965; as quoted in Bennett et al., 1991):

$$deff = \frac{s^2_{observed}}{s^2_{expected}} \qquad \text{Equation 1}$$

where:

$s^2_{observed}$ is the observed standard "s" error of the estimate (Equation 2), and
$s^2_{expected}$ is the expected standard error "s" when the data are assumed to come from a simple random sample (Equation 3)

$$s_{observed} = \frac{c}{\sum w_i x_i} \times \sqrt{\frac{\sum w_i^2 y_i^2 - 2p \sum w_i^2 x_i y_i + p^2 \sum w_i^2 x_i^2}{c(c-1)}} \qquad \text{Equation 2}$$

$$s_{expected} = \sqrt{\frac{p(p-1)}{\sum x_i}} \qquad \text{Equation 3}$$

where:

c is the number of clusters,
y_i is the number of households in the ith division / location with positive answer in the given variable
x_i is the number of surveyed households in the ith division / location, ,
p_i is the proportion of households in the ith division / location with positive answer in the given variable. Numerically, p_i is given by the ratio y_i / x_i, and
w_i is the weight of cluster i (proportional to the population of the cluster)

Study context

The Districts of Suba and Homa Bay are administratively located in Nyanza Province, in western Kenya, along the shores of Lake Victoria (Figure 1). Administratively, Homa Bay is divided into five divisions, and the divisions are further sub-divided into 25 locations. According to the 2009 census, the population is estimated at 366,620, and the district's density averages 313 persons per km^2. With regard to Suba, the district is made up of five administrative divisions and 20 locations. The total population is about 214,463, and the district's density stands at 202 persons per km^2.

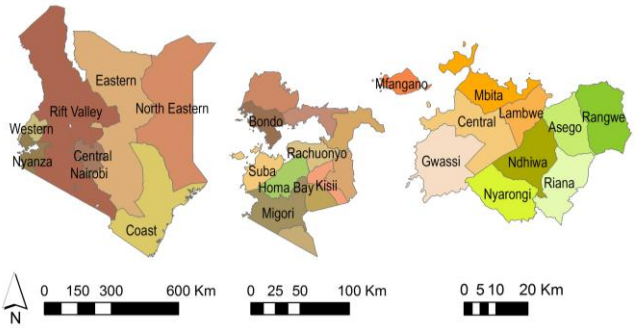

Figure 1. Administrative Units of Kenya, Nyanza Province, and Suba and Homa Bay Districts.

Results and discussion

The goal of the discussion first focuses on testing the statistics presented in previous section through an ad hoc case study. Specifically, we examine variability of *deff* in relation to i) the proportion *"p"* of the estimates and to ii) the administrative scale in which information is disaggregated. Second, and with the aim of showing to what extent the sample variances depend on the nature of the indicator studied, we use the data base from one Kenyan case study to compute *deff* for three core variables at two different administrative scales.

To begin with, we suggest a simple case study, in which only two clusters (A and B) are selected. As might be expected, it is observed from Table 1 that increased heterogeneity in the studied variable within clusters result in large values of *deff* - ten and hundred-fold -. In contrast, in a more homogeneous scenario, where p_A and p_B tend to same value, small *deff* might be applied.

Table 1. Variability of *deff* in relation to the proportion "p" of the estimate

	p_A / p_B	p	Std. Error	d
Case 1	0,5 / 0,5	0,5	0	*
Case 2	1 / 1	1	0	*
Case 3	0, 52 / 0,48	0,5	0,02	0,32
Case 4	0,55 / 0,45	0,5	0,05	2
Case 5	0,57 / 0,43	0,5	0,07	3,9
Case 6	0,60 / 0,40	0,5	0,1	8
Case 7	0,75 / 0,25	0,5	0,25	50
Case 8	1 / 0,5	0,75	0,25	66,67
Case 9	1 / 0,25	0,625	0,375	120
Case 10	1 / 0	0,5	0,5	200

Note: * Since Std. Error is equal to 0, D cannot be calculated

The second example presents the dependence of *deff* on the administrative scale in which information is analysed. Let the selected administrative unit (e.g. District of Homa Bay) have 2 divisions, 4 locations and 8 sublocations. Table 2 shows two different marked trends when *deff* is computed at lower administrative scales. The design effect diminishes when downscaling does not entail increased level of heterogeneity in

the studied variable (cases 1, 2 and 4). However, large values of *deff* are found in the increases of heterogeneity for lower administrative scales (cases 1, 3 and 5).

Table 2. Style: Variability of *deff* in relation to the administrative scale in which information is disaggregated

	p_{A1}	p_{A2}	p_{A3}	p_{A4}	p_{B1}	p_{B2}	p_{B3}	p_{B4}	p	d
Case 1 - 2 clusters	0,6	---	---	---	0,4	---	---	---	0,5	8
Case 2 - 4 clusters	0,6	0,6	---	---	0,4	0,4	---	---	0,5	2,67
Case 3 - 4 clusters	1	0,2	---	---	0,8	0	---	---	0,5	45,33
Case 4 - 8 clusters	0,6	0,6	0,6	0,6	0,4	0,4	0,4	0,4	0,5	1,14
Case 5 - 8 clusters	1	1	0,4	0	1	0,6	0	0	0,5	21,71

The trends described above may be observed if we analyse the data from the Kenyan case study, which are summarized in Table 3. Also, since WASH-based poverty may follow a highly heterogeneous pattern, widely varying between and within different administrative units; a set of maps are prepared and presented in Figures 2 to 7 to help understand and visualize such heterogeneity.

Table 3. Summary statistics of Kenyan case study

	Access to water				**Access to sanitation**				**Point-of-use water treatment**			
	P	*Std. Error*	*Conf. Inter.*	*d*	*P*	*Std. Error*	*Conf. Inter.*	*d*	*p*	*Std. Error*	*Conf. Inter.*	*d*
Homa Bay, at division level	0,478	0,049	0,379 - 0,576	11,34	0,130	0,017	0,096 - 0,162	2,82	0,702	0,028	0,645 - 0,758	4,40
Homa Bay, at location level	0,480	0,038	0,404 - 0,555	6,63	0,142	0,025	0,090 - 0,192	6,18	0,704	0,020	0,664 - 0,743	2,16
Suba, at division level	0,719	0,122	0,473 - 0,963	89,87	0,111	0,019	0,073 - 0,148	4,27	0,529	0,052	0,425 - 0,632	13,09
Suba, at location level	0,701	0,080	0,541 - 0,860	36,52	0,096	0,017	0,061 - 0,130	4,11	0,567	0,039	0,489 - 0,644	7,34

First, it can be seen that large standard errors of the estimates of proportions result in large values of *deff*. This is the case, for instance, of the variable related to the drinking-water source (Figures 2 and 3). It is also gleaned from the data that the sample variances depend on the nature of the case study. The two analysed districts show different *deff* for the three variables, regardless of the scale of analysis.

Similarly, when we move from the division to the location scale, the two trends identified in Table 2 are observed. As regards the type of household sanitation infrastructure used in Homa Bay, for instance, and at division level (Figure 4), high uniformity of the estimates p_i among divisions (Std. Error = 0.017) result in a relatively low value of *deff*, i.e. 2.82. At location scale (Figure 5), larger variability (Std. Error = 0.025) leads to higher *deff* values, i.e. 6.18. The contrary is true for the two other variables, where downscaling entails lower heterogeneity and thus smaller *deff*.

In practice, such large variability in design effects for most of the important sector-related variables urges survey planners to fine-tune the statistical precision of the estimates, especially when they are produced for sector planning, monitoring and evaluation support. It is recalled that a *deff* value of 4 has been widely accepted (United Nations Children's Fund, 2006, 2009), which is not in line with the results presented herein. A simple alternative for the calculation of the precision of an estimate is proposed by Bennet et al, (1991), who employs the standard error to construct a 95% confidence interval for the true value (from

estimate - 2 standard errors to estimate +2 standard errors). In so doing, we show that the precision initially expected (prior to data collection), i.e. 10 percentage points, is only achieved where *deff* values are not significantly large. For remaining variables, ignoring the design effect of the study would lead to assign estimates a narrower confidence interval than the correct value.

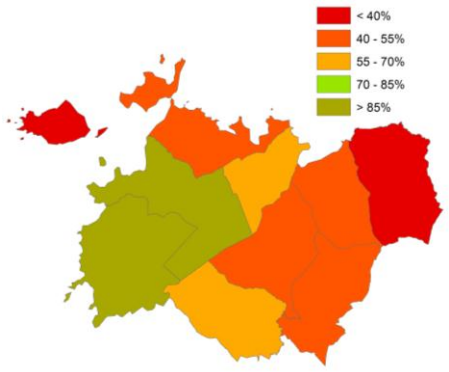

Figure 2. Access to improved water supplies (division)

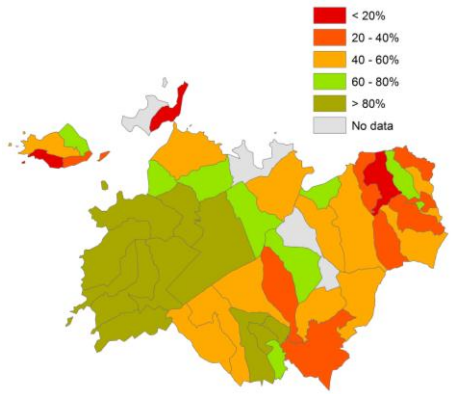

Figure 3. Access to improved water supplies (location)

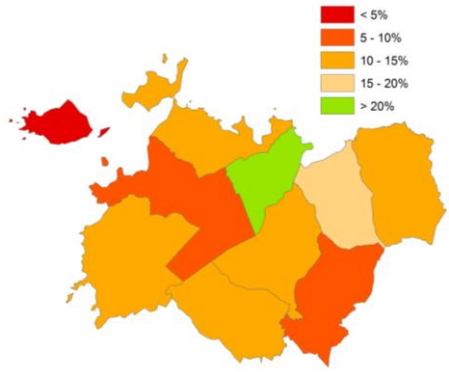

Figure 4. Access to improved sanitation (division)

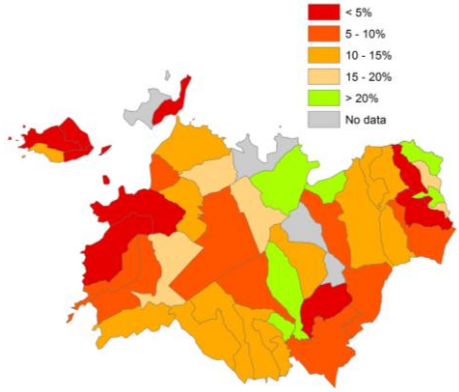

Figure 5. Access to improved sanitation (location)

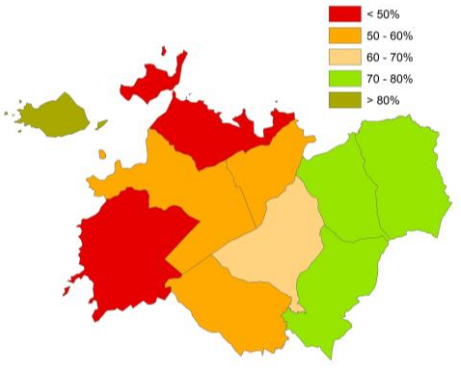

Figure 6. Household water treatment (division)

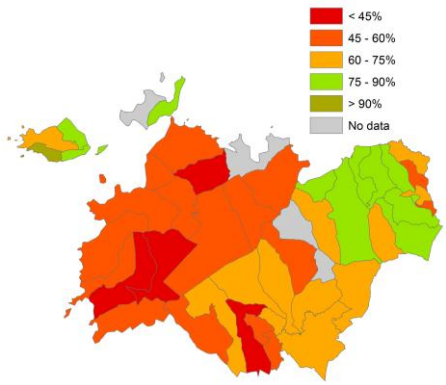

Figure 7. Household water treatment (location)

Conclusions

This study aims to discuss about the effect of cluster sampling strategies on the overall precision of produced estimates. In the WASH sector, it is an approach widely adopted by international agencies, e.g. the MICS, since it improves the efficiency of the data collection exercise by reducing costs and resources. In cluster survey techniques, as the process of clustering increases the risk of homogeneity within clusters, the size of the sample calculated has to be increased by the design effect. Literature elsewhere suggests for WASH variables a range of design effects between 2 to 10.

Based on data from a Kenyan case study, we show that design effects differ greatly within variables, and also depend on the regional setting. We therefore conclude that survey planners should carefully define the objectives of the survey and the required accuracy of collected data, keeping in mind that the sample variance may substantially impact on the final precision of survey estimates.

Acknowledgements

The authors would like to extend thanks to all families who participated in the study. Further thanks go to UNICEF (Kenya Country Office) and to Homa Bay and Suba District Water Office and District Public Health Office for their support to undertake the survey. This study has been partially funded by the Centre de Cooperació per al Desenvolupament (UPC). The support of the Agencia General d'Ajuts Universitaris I de Recerca (AGAUR, Catalan Government), provided through the Beatriu de Pinos grant (BP-DGR 2011) to Dr Jiménez, is also acknowledged.

References

United Nations Children's Fund (2006). Multiple Indicator Cluster Survey Manual 2005. New York: UNICEF, Division of Policy and Planning.

Bennett, S., Woods, T., Liyanage, W. M., & Smith, D. L. (1991). A simplified general method for cluster-sample surveys of health in developing countries. *World Health Stat Q*, 44(3), 98-106.

Lemeshow, S., & Stroh, G. (1988). Sampling Techniques for Evaluating Health Parameters in Developing Countries. Washington, D.C.: National Research Council.

Kaiser, R., Woodruff, B. A., Bilukha, O., Spiegel, P. B., & Salama, P. (2006). Using design effects from previous cluster surveys to guide sample size calculation in emergency settings. *Disasters*, 30(2), 199-211.

Kish, L. (1980). Design and Estimation for Domains. *Statistician*, 29(4), 209-222.

Katz, J., & Zeger, S. L. (1994). Estimation of design effects in cluster surveys. *Annals of Epidemiology*, 4(4), 295-301.

Howard, G., Ince, M., & Smith, M. (2003, draft). Rapid assessment of drinking-water quality: a handbook for implementation. Geneva: World Health Organization and United Nations Children Fund.

Bostoen, K. (2002). Measuring Access and Practice: Designing a Survey Methodology for the Hygiene, Sanitation and Water Sector. *Department of Infectious and Tropical Diseases*. London: London School of Hygiene and Tropical Medicine.

United Nations Children's Fund (2009). A WASH Baseline Survey under the NL - UNICEF partnership for WASH. One Million Initiative (Mozambique). Maputo: UNICEF and WE Consultant.

Joint Monitoring Programme (2006). Core questions on drinking-water and sanitation for household surveys. Geneva / New York: WHO / UNICEF.

Contact details

Ricard Giné Garriga, Alejandro Jiménez Fdez. De Palencia and Agustí Pérez Foguet
Research Group on Cooperation and Human Development, University Research Institute for Sustainability Science and Technology, Universitat Politècnica de Catalunya, Spain
Email: ricard.gine@upc.edu / alejandro.jimenez.fdez.de.pale@upc.edu / agusti.perez@upc.edu
www: http://grecdh.upc.edu
http://is.upc.edu