Soc. Sci. Med. Vol. 22, No. 7, pp. 781-783, 1986.

CLUSTER SAMPLING FOR IMMUNIZATION COVERAGE

PETER W. KOK
Medical Research Centre, Kenya Medical Research Institute, P.O. Box 20752, Nairobi, Kenya

Abstract—The WHO/EPI cluster sampling method for immunization coverage surveys is part of the course for management training in EPI programmes. The application of this method, based on a framework of villages, is impractical in dispersed populations common in many countries in Africa. To make the method work under those circumstances leads to unacceptable bias, which tends to overestimate the real coverage rate. A modified random cluster sampling method for dispersed populations is presented.

INTRODUCTION

The choice of an appropriate method of surveillance and evaluation before implementing an immunization programme is essential to ensure that the very scarce resources available to national programmes are used to its maximal gain.

Sample surveys at certain intervals are usually cheaper and more reliable than setting up and maintaining a continuous surveillance system. It is therefore important to have a critical look at the sampling methods used. Various ways of doing an immunization survey by cluster sampling in the community are used.

Ideally, one should randomly, i.e. by chance alone, select each sample unit separately and independently. A sample unit could be a household, a pregnant mother or, as in our example, the under 5 child. This would need a complete numerical inventory of all sampling units concerned within a defined geographical area. This is usually impractical in developing countries as an inventory of all units in a community does not exist, while the costs of making such an inventory could be prohibitive.

For this reason, cluster sampling has become a popular method to perform immunization coverage surveys [1], even to the extent that oversimplification of sound methodological principles has occurred.

CRITIQUE REGARDING EPI SAMPLING METHOD

The cluster sampling method as used in the World Health Organization's Expanded Programme on Immunization (EPI) [1] and Breastfeeding Surveys [2] as applied in Kenya and other African countries, suffers from serious bias and presents practical problems.

The main limitation in using the EPI cluster sampling method is that the sampling framework is based on the assumed presence of *geographically* more or less distinct villages and towns.

However, in many parts of Africa, and Kenya in particular, the population is not living in distinguishable villages, but lives rather dispersed in areas which are not easily defined geographically.

In Kenya, the village as sample frame has been replaced by 'sublocation' as the smallest administrative unit for which population data are available. The WHO/EPI method of taking 30 clusters of 7 children each, clusters being selected from a cumulative village population list, has been changed to the selection of 30 clusters from a cumulative sublocation list, thus ensuring a proportional representation of clusters in a given area, usually a district. So far there is no problem, but the next stage of selecting the first household to be visited is very difficult. In the WHO/EPI village model, the instruction requires the field worker to go to the 'centre' of the village, select randomly a direction and count the number of houses between 'centre' and the village boundary, and select from this number of households randomly the first to be visited. If the village is large and heterogeneous, or grossly unequal in its dimensions, this procedure may lead to bias in selecting the first household, as the chance in one direction of being selected can be twice as big as in any other direction, or at least disproportionate. Counting of houses in larger villages or towns is often too difficult and time consuming.

In the case of proportional selection of sublocations instead of villages, those problems are even bigger as distances are usually greater while boundaries are less clearly defined in the field situation. As the 'centre' of a sublocation is usually taken to be one of the trading centres, with the chance that the area's immunization centre is at the same place (bias 1), and to avoid the problem of counting all households from the 'centre' to the boundary of the sublocation, the first household is randomly selected amongst the first 10 households counted from the 'centre', intentionally excluding the majority of the households from possible inclusion into the sample (bias 2).

As the housing density tends to be higher toward the 'centre', the method of taking the 'next nearest house' as the systematic sampling method, adds the third bias of being centripetal, with a good chance of ending up in and around the centre of the chosen village or trading centre, where the socio-economic and transport situation may differ from the more outlying parts of the sublocation or village, and thus an overestimation may occur of the immunization coverage.

To identify and prepare the villages or trading centres, extensive and very expensive previsiting by a district health worker is usually done in the Kenyan situation.

MODIFIED CLUSTER SAMPLING

The WHO/EPI cluster sampling method was therefore modified in such a way as to remove to a large

extent the three sources of bias mentioned before, and greatly reduced the cost by simplifying preparations in the field. The method proved to be very suitable in surveying dispersed rural populations in Kenya, and could be particularly useful in an urban situation.

The basis of this sampling method is the assumption, and therefore also its limitation, that the Standard I primary school child is the most randomly and proportionally distributed registered sampling unit in the community. The random selection of 30 Standard I children out of the total Standard I school population for the selection of 30 starting points (households) in a given area is a relatively simple procedure.

In Kenya, good census data area available from the Central Bureau of Statistics (CBS), for district, division, location and sublocation.

From the Ministry of Education or from the District Education Officer (DEO) a list of all primary schools with the number of pupils at each school can be obtained. On the basis of these two sets of data, a simple sample framework can be designed.

The assumption made before has to be tested, because if the schooling level is, arbitrarily, below 70% and certain areas are excluded from schooling, the sampling method would be defective. If that were the case, not every child of Standard I school-age would have the same chance of being included in the survey. The level of schooling can be estimated by dividing the primary school population as obtained from the DEO with the 5–14 year age group obtained from the CBS census data. The distribution of school attendance rates between divisions within the district can be assessed in the same way. As in the WHO/EPI sampling method, instead of making a cumulative list of village populations, a cumulative list of school populations is made.

The total school population is then divided by the number of clusters needed (30) to determine the sampling interval (a). Likewise a random number (n) is selected within the size of sampling interval. The first school selected is the one attended by the nth child, the second school is the one attended by the nth $+ a$ child, the third by the nth $+ 2a$ etc., till 30 schools have been identified.

On the (school) day of the survey the investigation team, preferably with added local members of the community, visits the school, examines the attendance list of Standard I kept by the headmaster of the school, and randomly chooses a Standard I child from amongst all the listed Standard I children. This child or his elder sibling directs the team to his or her house. To remove the bias 'household with school child', the next nearest household is the starting point for sampling households till at least 7 children in the desired age group have been surveyed.

Illustration from Machakos District, Kenya

An immunization coverage study carried out in Machakos District in Kenya may serve as an example. The population estimate (1981) was 1,091,000 and the number of children 5–14 was 327,000 (CBS). A list of *schools with the number of children in each* grade was obtained from the DEO, showing a primary school population of 314,000 children in 788 schools. The proportion of children attending school is estimated as 314,000/327,000 = 96%. Further analysis of the distribution rate by geographical division showed that no division had a school attendance rate of below 89%. In view of the *increasing drop-out rate with increasing grade in schooling*, the Standard I child was taken as the basis for our calculations.

Table 1. Cumulative listing schools, Machakos District, Kenya, 1981

School	Enrollment	Cumulative all pupils	Cumulative Standard I
1. Ianguni	494	494	130
2. Kithimani	1034	1528	313
3. Kwakulu	421	1949	399
4. Kaumoni	483	2432	496
5. Kaseve	363	2795	570
Total		314,000	60,220

Analysis for the Machakos District showed that the mean proportion of Standard I school children in the total school age population (CBS) by division was 18.5% with a standard deviation of 1.5%.

The assumption therefore that the Standard I school child is randomly and proportionally distributed in Machakos District seems valid.

The 788 schools in the Machakos District were listed, with the number of pupils attending, and if available, the number of Standard I children (Table 1).

The *sampling interval* for Standard I pupils was obtained by dividing the total number of children (60,220) by the number of clusters (30), giving 2007. Next a random number between 0000 and 2007 was chosen, in our example 0404, representing the first child. This 404th child was attending school No. 4 Kaumoni. The second school was obtained by adding the sample interval to the first random number, 0404 + 2007 = 2411. This school at which the 2411th child was attending could be identified from the list. This process was repeated till all 30 schools were identified. A list of these schools and their address was compiled, and the survey could start. Each school was subsequently visited, the Standard I school attendance list was obtained from the headmaster, and from this list a child was randomly chosen. Permission was obtained for the child or elder sibling to bring the team to the pupil's house, from where the next nearest house was taken as the starting point for the cluster survey. The child was then returned to the school.

Table 2. School attendance rates in some districts in Kenya, 1981

District	School population	No. of schools	School attendance rate (%)
Kilifi (1)	81,500	256	84* 44†
Kwale (1)	52,300	201	73* 49†
Machakos (2)	327,000	788	96
Kiambu (3)	187,000	305	78
Uasin Gishu (4)	81,400	215	84
Trans Nzoia (4)	73,400	121	85
Busia (5)	86,400	263	92
Kisumu (6)	131,718	400	91
Siaya (6)	149,500	472	102

*Boys; †girls.
1—Coast; 2—Eastern; 3—Central; 4—Rift Valley; 5—Western Province; 6—Nyanza Province.
Source: Ministry of Education.

PROBLEMS AND LIMITATIONS

The sampling method is dependent on the school, and therefore cannot be done during school holidays and weekends. In case the school attendance list for the year in which the survey takes place is not yet available, the school list of the previous year could be used to identify schools, as it might be expected that changes in the number and relative attendance of schools are small over such a limited period. The actual child is always selected physically at the school during the survey.

A special case is the boarding school with pupils from very far away. Their wider distribution and overlap of other schools might not disturb the representiveness if the number is small. A limit as to the distance from the school a pupil is living can be fixed beforehand, depending on the nature and population densities of the area to be surveyed.

This method using the schools as sampling framework is widely applicable in Kenya as school attendance rates are generally high (Table 2).

The assumption of random proportional distribution of school children in the community could not be made readily in the (partially Islamic) Coastal Districts of Kenya as less than the *arbitrary* 70% of the children were going to school, however further examination revealed that a strong discrepancy in school attendance rates between boys and girls existed; less than 50% of the girls were going to school, while well over 70% of the boys were attending. The sampling method using schools could still be used, taking for the cumulative list of school attendances only the number of boys per school.

REFERENCES

1. Henderson R. H. and Sundaresen T. Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. *Bull. Wld Hlth Org.* **60,** 253–260, 1982.
2. World Health Organization. Methodology for breastfeeding survey. Document MCH/BF/SUR/81, 1981.
3. Kok P. W. and Kwamanga D. A simple random survey method for the expanded programme on immunization. *Current Medical Research in Eastern Africa,* pp. 82–84. Africascience International, Nairobi, 1982.