TABLE 3 *Variance of different sampling schemes with sample size 30, and (below) household design effect (*hdeff*), or ratio to srs variance*

	Sampling scheme[a]					
	SRS	EPI	EPI3	EPI5	QTR	PERI
Nutrition						
Height for age	2.2	1.9	2.3	2.2	2.1	1.8
		0.89	1.07	1.00	0.99	0.83
Weight for age	1.5	1.4	1.4	1.4	1.4	1.1
		0.92	0.97	0.98	0.96	0.74[c]
Weight for height	0.34	0.28	0.34	0.33	0.28	0.25
		0.82	1.00	0.96	0.83	0.74[c]
Mid-upper arm circumference	1.4	1.2	1.5	1.4	1.4	1.2
		0.83	1.05	1.00	1.01	0.85
Morbidity						
Fever	0.81	0.60	0.73	0.69	0.72	0.64
		0.75[c]	0.91	0.86	0.90	0.80[c]
Diarrhoea	0.69	0.56	0.62	0.62	0.67	0.53
		0.81	0.90	0.89	0.97	0.77[c]
Respiratory infection	1.5	1.5	1.6	1.4	1.3	1.3
		0.96	1.06	0.92	0.85	0.86
Health care						
Breastfed	1.5	1.2	1.5	1.4	1.1	0.9
		0.81	1.00	0.93	0.73[c]	0.61[c]
Pregnant	1.7	1.7	1.9	1.9	1.6	1.6
		0.99	1.10	1.12	0.93	0.93
Weighed in last 3 months	0.94	1.15	1.00	0.93	0.94	0.95
		1.22[b]	1.06	0.99	1.00	1.01
Growth chart available	2.6	3.5	3.5	3.1	3.2	2.9
		1.33[b]	1.32[b]	1.17	1.21[b]	1.12
Interpret chart	1.6	2.1	1.7	1.7	1.9	1.8
		1.29[b]	1.06	1.01	1.15	1.09
Socioeconomic						
Mother's education	3.6	3.5	3.9	3.7	3.5	3.0
		0.97	1.07	1.03	0.96	0.82
Father's education	2.9	2.8	3.2	2.8	2.7	2.6
		0.93	1.08	0.94	0.92	0.87
Ethnic group	1.8	2.2	2.0	1.8	2.0	1.9
		1.23[b]	1.14	1.00	1.15	1.07
Religion	3.4	4.4	3.8	3.3	3.6	4.0
		1.30[b]	1.13	0.98	1.06	1.19
Subsistence farmer	3.7	4.4	4.4	3.8	3.4	4.2
		1.20	1.20	1.04	0.93	1.14
Keeps cattle	2.4	2.9	2.8	2.5	2.5	2.3
		1.18	1.12	1.01	1.03	0.92
Grows a cash crop	1.6	2.1	1.9	1.8	1.6	1.7
		1.30[b]	1.18	1.10	0.95	1.01

[a] For an explanation of the sampling schemes see text.
[b] *hdeff* >1.2
[c] *hdeff* <0.8

TABLE 4 *Bias (upper) and mean square error (lower) for sample size 7*

	Sampling scheme[a]					
	SRS	EPI	EPI3	EPI5	QTR	PERI
Nutrition						
Height for age	0.01	1.25	1.53	0.82	0.03	1.11
	11.9	12.1	13.4	11.8	10.7	9.9
Weight for age	−0.06	0.17	0.57	0.57	0.34	−0.20
	7.5	6.6	8.1	7.7	7.3	5.2[c]
Weight for height	0.05	0.23	0.29	0.46	−0.20	−0.23
	1.8	1.5	1.8	2.0	1.5	1.1[c]
Mid-upper arm circumference	−0.03	−0.65	0.20	−0.19	−0.36	−0.09
	6.6	7.5	7.8	6.8	7.1	6.0
Morbidity						
Fever	−0.02	−0.72	−0.39	−0.36	−0.10	−1.19
	3.5	4.1	4.0	4.0	3.3	4.3[b]
Diarrhoea	0.10	−0.32	−0.17	0.09	0.09	−0.36
	3.8	3.6	3.4	3.5	3.4	3.2
Respiratory infection	0.00	−0.22	−0.27	−0.25	0.12	−0.24
	7.6	8.1	7.8	6.9	7.0	6.9
Health care						
Breastfed	0.09	0.87	0.02	0.33	−0.08	−0.03
	7.7	6.1[c]	7.0	7.3	7.1	4.7[c]
Pregnant	0.11	−0.84	−0.27	−0.18	0.41	−2.12
	9.1	10.4	9.7	8.7	8.3	11.9[b]
Weighed in last 3 months	0.04	−0.14	−0.55	−0.25	0.45	−0.96
	4.8	5.5	5.2	4.6	4.3	4.5
Growth chart available	0.26	−0.38	−0.68	−0.90	1.34	0.04
	13.6	14.9	14.1	14.7	16.8[b]	10.9[c]
Interpret chart	0.14	−0.70	−0.15	0.17	1.15	−0.77
	7.8	11.1[b]	8.1	7.4	8.6	7.0
Socioeconomic						
Mother's education	0.01	1.23	−0.03	0.01	−0.93	2.16
	17.3	19.3	18.8	18.4	18.7	15.7
Father's education	−0.05	1.90	0.63	0.78	0.26	2.49
	15.1	19.1[b]	14.8	16.0	14.0	17.9
Ethnic group	0.11	2.62	0.69	0.66	0.05	2.69
	9.1	18.2[b]	9.5	8.8	9.6	14.2[b]
Religion	−0.12	−1.79	−0.70	−0.68	0.38	−0.79
	15.8	21.6[b]	17.8	16.1	17.7	13.3
Subsistence farmer	0.01	1.19	1.32	1.18	0.57	1.90
	18.7	22.5[b]	21.3	21.4	19.0	18.5
Keeps cattle	0.02	−3.01	−1.53	−1.44	−0.05	−3.45
	11.9	20.3[b]	14.0	15.0[b]	11.5	20.0[b]
Grows a cash crop	−0.00	−0.01	−0.93	−0.39	0.64	−1.67
	8.3	8.3	8.2	8.7	8.0	9.5

[a] For an explanation of the sampling schemes see text.
[b] MSE >1.2 × MSE of SRS
[c] MSE <0.8 × MSE of SRS

TABLE 5 *Bias (upper) and mean square error (lower) for sample size 30*

	Sampling scheme[a]					
	SRS	EPI	EPI3	EPI5	QTR	PERI
Nutrition						
Height for age	0.08	0.81	0.72	0.44	−0.38	1.27
	2.2	2.6	2.8[b]	2.4	2.3	3.4[b]
Weight for age	0.02	0.08	0.32	0.37	−0.33	0.35
	1.5	1.4	1.5	1.6	1.5	1.2
Weight for height	0.00	0.10	0.17	0.25	−0.10	0.32
	0.34	0.29	0.37	0.39	0.29	0.35
Mid-upper arm circumference	0.04	−0.37	−0.05	0.02	−0.78	0.54
	1.4	1.3	1.5	1.4	2.0[b]	1.5
Morbidity						
Fever	0.00	−0.06	0.08	−0.08	0.09	0.06
	0.81	0.61[c]	0.74	0.70	0.73	0.64[c]
Diarrhoea	−0.02	−0.83	−0.13	0.09	0.03	−0.47
	0.69	1.26[b]	0.63	0.62	0.67	0.75
Respiratory infection	0.00	−0.58	−0.34	−0.40	−0.03	−0.13
	1.5	1.8	1.8	1.6	1.3	1.3
Health care						
Breastfed	−0.02	0.40	0.29	0.16	0.26	0.81
	1.5	1.4	1.6	1.4	1.2[c]	1.6
Pregnant	−0.02	0.19	0.22	0.24	−0.00	0.40
	1.7	1.7	1.9	2.0	1.6	1.8
Weighed in last 3 months	0.02	0.38	0.03	0.20	−0.01	−0.10
	0.94	1.30[b]	1.00	0.97	0.94	0.96
Growth chart available	0.12	−0.18	−0.41	−0.29	0.58	−0.23
	2.7	3.5[b]	3.7[b]	3.2	3.5[b]	3.0
Interpret chart	−0.01	0.26	0.05	0.21	0.69	−0.49
	1.6	2.2[b]	1.7	1.7	2.4[b]	2.0[b]
Socioeconomic						
Mother's education	−0.07	0.34	−0.20	−0.21	−0.80	−0.16
	3.6	3.6	3.9	3.8	4.1	3.0
Father's education	−0.15	1.25	0.47	0.75	0.24	2.28
	3.0	4.3[b]	3.4	3.3	2.8	7.7[b]
Ethnic group	−0.07	0.55	0.16	0.43	−0.20	1.13
	1.8	2.5[b]	2.0	2.0	2.0	3.2[b]
Religion	0.09	−1.93	−0.16	−0.50	−0.49	−1.41
	3.4	8.1[b]	3.9	3.6	3.8	6.0[b]
Subsistence farmer	0.09	0.09	0.70	0.82	−0.71	0.14
	3.7	4.4	4.9[b]	4.5[b]	3.9	4.2
Keeps cattle	0.07	−2.49	−0.89	−1.00	−0.39	−2.85
	2.5	9.1[b]	3.5[b]	3.5[b]	2.7	10.4[b]
Grows a cash crop	−0.18	−0.13	−0.29	0.05	0.87	−0.54
	1.7	2.1[b]	2.0[b]	1.8	2.3[b]	2.0

[a] For an explanation of the sampling schemes see text.
[b] MSE >1.2 × MSE of SRS
[c] MSE <0.8 × MSE of SRS

Proportion outside Range

The proportions of estimates which lie outside a given range (± 10%, ± 5%, ± 1%) centred on the true value were also calculated (data not shown) and confirm the pattern shown by the MSE, with PERI and EPI faring particularly poorly for socioeconomic, cultural, educational and health care variables.

DISCUSSION

The EPI household sampling scheme was designed as part of a rapid and economical survey procedure whose aim was to estimate vaccination coverage in children within a narrow age range. It has since been adopted for many other purposes, and a number of modifications have been proposed and evaluations carried out.[7] Simulation of household selection from artifical populations by Lemeshow *et al.*[4] has indicated that the scheme is less efficient than SRS for realistic models of the distribution of population and of vaccine coverage, although it may perform adequately in its objective of estimation to within 10 percentage points. In this simulation generated from real data from 30 communities in Uganda we have tested the properties of EPI sampling and four variants of it for a range of variables commonly studied in surveys of health and nutritional status, for children aged 0–4 years.

There are important aspects of the comparison of these sampling schemes that are not amenable to study by computer simulation. If the population estimates used for selecting communities with probability proportional to size are out of date, the unweighted survey estimates may be biased. The complete enumeration of households necessary for a systematic or simple random sample will enable this to be corrected, by weighting or adjusting the number of households taken. Selection of specified households from a numbered list also enables supervisors to monitor the work of interviewers by revisiting a subsample of households. This is much more difficult in the EPI scheme. Set against this is the cost saving in not carrying out mapping and enumeration.

The simulations are limited in their capacity to represent the topography of the villages: in practice the choice of random direction to follow would be limited by the roads and paths that existed, and by the existence of barriers such as fences, streams etc. Nor could the simulations consider other practical realities such as interviewer motivation, time constraints, or the views of the head of the community.

A clear feature of the results is that the choice of household sampling scheme has little effect on variables representing nutritional status of children and their recent morbidity, except that the EPI and PERI schemes overestimate (by up to 1% or more) the proportion with low height for age. In estimating educational variables and mother's pregnancy status there is evidence of bias with both PERI and EPI, particularly with the smaller sample sizes.

The EPI scheme has a very large variance compared with SRS for the variables relating to child care (whether weighed or not, presence and understanding of growth chart), particularly for the larger sample sizes. For such variables EPI5 performs as well as SRS, with the other schemes somewhere in between. For the other socioeconomic variables (religion, ethnic groups and variables relating to farming) EPI performs poorly, both in terms of variance and bias. The lower variances achieved by PERI are more than outweighed by the large biases seen with this method. EPI5 and QTR perform as well as SRS, and EPI3 not quite as well.

In general, household design effects increase with the number of households sampled, while the bias decreases a little.

The difference in the performance of the sampling schemes for different variables depends on the way in which these variables are distributed within communities. With variables which show a gradient from the centre to the edge of the community, such as keeping cattle, the EPI and the PERI schemes will yield a similar bias in every community, leading to an overall bias, since they both oversample a small group of houses at the centre of the community. (The EPI method does this because, in selecting the first household, an imaginary radius sweeping 360° round the community will pass over fewer houses near the centre than at the periphery.) A bias which differs between communities will lead to increased variance. Thus variance of the EPI scheme increases for variables such as ethnic group which are likely to cluster within each community: the sample may be wholly within a group in some villages, and wholly outside it in others.

With the PERI scheme, households at the centre of the community are included in every sample, while households halfway between the centre and the edge of the community will not be included in any sample. This limitation in the range of possible samples is not only the cause of its large bias, but also of the smaller variance. The risk of bias outweighs the low variance and suggests that schemes such as this should not be used.

The large *hdeffs* of the EPI scheme for educational levels, mother's pregnancy status, and variables relating to child care, indicate the presence of within-community clustering for these variables, and reflect the high design effects seen for complete surveys for socioeconomic factors and factors related to health care.[1,11] The low *hdeffs* shown by the morbidity and nutrition variables for all sampling schemes indicate that neither cases of common illnesses, nor children with poor nutritional status, form clusters within communities, being distributed rather randomly. Such a homogeneity has been observed in the relatively low whole-survey design effects for prevalence of common illnesses from surveys in Guinea[12] and Maldives (W Liyanage, unpublished MSc thesis), but contrasts with the high design effects for epidemic diseases such as measles.[13]

The sample size necessary under SRS sampling of households, assuming that one has already taken account of cluster sampling of communities, will need to be multiplied by a factor *hdeff* to obtain the same precision with EPI sampling of households. For a survey whose main interest is estimating levels of nutritional status and general morbidity, no increase in sample size is necessary. For surveys whose primary concern is to investigate child care practices and socioeconomic status, the sample size will need to be increased by between 20% and 40%, as shown by the values of *hdeff*. Whether the bias of 1–2% seen with the EPI scheme is important will depend on the objectives of the survey.

In studying all children aged 0–4 years we examined the effect of each sampling scheme in a situation where almost every household would be eligible for inclusion in the sample. Use of a narrower age range would entail visiting more households in order to find the required sample size. Experience shows that to find seven children aged 12–23 months may entail visiting 20 or more households, so that a sample taken using the EPI scheme may result in a child being selected from approximately every fourth household. The EPI sample would thus have the properties of the EPI3 or EPI5 scheme, and as we have shown, would be likely to perform as well as SRS.

The results of these simulations suggest that the EPI scheme, which is simple and quick, is suitable for rapid appraisal of the prevalence of morbidity or the nutritional status of communities where practical reasons preclude the use of systematic or simple random sampling. However, where surveys cover a wider range of topics, such as health care, or where a survey seeks to examine the association of health or nutritional status with explanatory variables such as education or socioeconomic factors, the unmodified EPI scheme will be inefficient and somewhat biased. Adapting the scheme by taking every fifth house, or by splitting the community into quarters and carrying out the EPI scheme in each quarter, gives results as good as SRS and much better than the unmodified EPI scheme in such cases. The extra work involved is small: a little more walking and, if sampling in quarters, the repetition four times of the selection of the starting household. There is a need for field studies to evaluate both the costs and benefits of such adaptations, and the costs and additional benefits (interviewer control, corrected population estimates) of complete enumeration.

ACKNOWLEDGEMENTS

This work was supported by UNICEF, Uganda. S Bennett and A Radalowicz are supported by the UK Medical Research Council.

REFERENCES

1 Bennett S, Woods A J, Liyanage W M, Smith D L. A simplified general method for cluster sample surveys of health in developing countries. *World Health Stat Q* 1991; **44**: 98–106.

2 World Health Organization. *Training for Mid-level Managers: The EPI Coverage Survey.* Geneva: WHO Expanded Programme on Immunization, 1991, WHO/EPI/MLM/91.10.

3 Henderson R H, Sundaresan T. Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. *Bull World Health Organ* 1982; **60**: 253–60.

4 Lemeshow S, Tserkovnyi A G, Tulloch J L, Dowd J E, Lwanga S K, Keja J. A computer simulation of the EPI survey strategy. *Int. J Epidemiol.* 1985; **14**: 473–81.

5 El Bindari-Hammad A, Smith D L. *Primary Health Care Reviews. Guidelines and Methods.* Geneva: WHO, 1989.

6 D'Alessandro U, Aikins M K, Langerock P, Bennett S, Greenwood B M. Nationwide survey of bednet use in rural Gambia. *Bull World Health Organ* 1994; **72:** 391–94.

7 Bennett S. The EPI cluster sampling method: a critical appraisal. *Bull Int Statist Inst* 1993; **55**: Book 2, 21–35.

8 Vella V, Tomkins A, Nidku J, Marshall T. Determinants of child mortality in south-west Uganda. *J Biosoc Sci* 1992; **24**: 103–12.

9 Vella V. *An Epidemiological Analysis of Predictors of Childhood Malnutrition and Mortality in Southwest Uganda.* London: University of London, 1991. PhD thesis.

10 Kish L. *Survey Sampling.* London: Wiley, 1965.

11 Kish L, Groves R M, Krotki K P. *Sampling Errors for Fertility Surveys.* London: World Fertility Survey, 1976. Occasional paper no. 17.

12 Dabis F, Breman J G, Roisin A J, Haba F, the ASCI-CCCD team. Monitoring selective components of primary health care: methodology and community assessment of vaccination, diarrhoea, and malaria practices in Conakry, Guinea. *Bull World Health Organ* 1989; **67**: 675–84.

13 Rothenberg R B, Lobanov A, Singh K B Stroh G. Observations on the application of EPI cluster survey methods for estimating disease incidence. *Bull World Health Organ* 1985; **63**: 93–99.

(Revised version received June 1994)