Bulletin of the World Health Organization, **60** (2): 253 – 260 (1982)

Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method

R. H. HENDERSON[1] & T. SUNDARESAN[2]

A simplified cluster sampling method, involving the random selection of 210 children in 30 clusters of 7 children each, has been used by the Expanded Programme on Immunization to estimate immunization coverage. This paper analyses the performance of the method in 60 actual surveys and 1500 computer simulated surveys. Although the method gives a proportion of results with confidence limits exceeding the desired maximum of ± 10 absolute percentage points, it is concluded that it performs satisfactorily.

The Expanded Programme on Immunization (EPI) seeks to reduce morbidity and mortality by providing immunization against diphtheria, pertussis, tetanus, measles, poliomyelitis, and tuberculosis for all children of the world by 1990. The special concern of EPI is the strengthening of immunization services for children in developing countries, and the programme seeks methods of programme implementation and evaluation that are effective, simple, and inexpensive.

An example of such a method is the estimation of immunization coverage through the examination of approximately 210 children, selected randomly as 30 groups of 7 children each. This is based on a survey technique originally used in the United States of America (*1*) and later updated for use in West Africa (*2*). This paper analyses the results of the method in actual and computer simulated surveys, and discusses its strengths and weaknesses.

DESCRIPTION OF THE METHOD

The method used by EPI consists of:

— identification of the geographical area(s) of interest;

— identification of the age group(s) of interest;

— random selection of 30 sites (termed "clusters") from within each geographical area for which individual results are desired;

— random selection of a starting point ("household") within each site; and

— selection of 7 individuals of the appropriate age from within each of the 30 sites. Selection begins in the starting household and then continues to the next nearest household until a total of 7 individuals is obtained. All individuals of the appropriate age living in the last household falling into the sample are included, even if this means including 8 – 10 individuals in the cluster rather than the required minimum number of 7.

A full description of the theoretical basis for sampling procedures is beyond the scope of this paper, and may be found elsewhere (*3*). In brief, however, the EPI method treats immunization status as a binomial variable, meaning that for each vaccine or vaccine dose of interest all individuals are classified as belonging to one of two categories: immunized or not immunized. For binomial variables the size of a random sample required to produce results of given accuracy and precision can be determined from the formula (*3*, page 74):

$$n = (z^2pq)/d^2$$

To solve the formula for 'n', the number of persons required in the sample, values of the other variables must be provided. The value for 'd' corresponds to the precision of the result desired. For EPI, it was decided that an estimate that lay within ± 10% from the population mean would be adequate, and 'd' was assigned a value of 0.10. It should be noted that this value signifies 10 percentage points on a scale running from 0% to 100% and not to a percentage of the survey estimate itself: for a survey result of 30%, this formula predicts that the true result will lie between 20% and 40%, not 27% and 33%.

The value of 'z' corresponds to the confidence limits of the survey result. It was decided that for EPI

[1] Director, Expanded Programme on Immunization, World Health Organization, 1211 Geneva 27, Switzerland.

[2] Formerly: Statistician, Health Statistical Methodology, World Health Organization, Geneva, Switzerland.

the confidence limits of 95% would be adequate. This means that a range of ± 10% from the sample result should include the true result in 95 of every 100 surveys performed. Standard statistical tables can be consulted to find values for 'z' the normal deviate, corresponding to the confidence limits desired. The value of 'z' for confidence limits of 95% is 1.96.

The values for 'p' and 'q' correspond to the proportion of persons in the population who are immunized (p) and not immunized (q). The addition of 'p' and 'q' must equal 1.0. Of course, if these proportions were actually known beforehand, there would be no need to take a sample. The survey planner, therefore, is faced with making a best guess. As 'n' in the formula is maximized when a value of 0.5 is assigned to both 'p' and 'q', this value was adopted.

Solving the formula:

$$n = (z^2pq)/d^2 = (1.96)^2(0.5)(0.5)/0.1^2$$
$$= (3.84)(0.25)/0.01 = 96$$

In survey work, much time and effort are spent in following a method of simple random sampling until an individual is actually identified and enrolled in the sample. Savings can be achieved if, instead of performing this 96 times for 96 individuals as required by the above calculations, the sampling process can be performed fewer times, each time taking several persons. This is called cluster sampling. Because there is usually a tendency for individuals found within a cluster to share characteristics, however, use of cluster sampling can be expected to decrease the precision of the sample result. In applying a method of cluster sampling, then, two questions must be answered:

— what is the miminum number of clusters that can be selected and still fulfil the requirements on which the theory of binomial sampling is based? and

— what should be done to compensate for the bias introduced when one samples persons in groups rather than individuals?

The first of these questions may be answered on the basis of theory: over ranges of values for 'p' from at least 5% to at least 95%, and for samples that contain a total of at least 96 individuals, selection of an equal number of these individuals from at least 30 randomly selected clusters is sufficient. As a working rule, the number 30 can be taken as being large enough to ensure that the cluster means will tend to have a normal distribution, thus permitting statistical theory based on the normal distribution to be used to analyse the data (*3*, page 157).

The second question can be answered only on the basis of experience. Following the example set by the smallpox immunization surveys conducted in West Africa in 1968 and 1969 (*2*), estimates made by Serfling & Sherman of the association within clusters of the status of smallpox immunization in children aged 1 – 4 years in the United States of America (*1*) were used. This had the effect of almost doubling the number of children required, and to permit an equal number of children to be selected from each of the 30 clusters, a sample size of 210 was adopted, 7 children to be selected from each cluster.

ANALYSIS OF RESULTS FROM SURVEYS IN DEVELOPING COUNTRIES

Since 1978, surveys using this method have been used in developing countries to estimate immunization coverage. Results from 60 such surveys performed in 25 countries (Table 1) were available for analysis at the time this paper was prepared. Many of these surveys have been presented in the *Weekly epidemiological record* (*4*).

Table 1. Number of EPI immunization coverage surveys analysed, by WHO Region and country

Region/Country	No. of surveys analysed
African Region	
Benin	1
Botswana	1
Gambia	1
Ivory Coast	1
Kenya	2
Malawi	3
Nigeria	1
Zaire	1
Eastern Mediterranean Region	
Bahrain	1
Pakistan	1
Somalia	8
Sudan	4
Syrian Arab Republic	1
United Arab Emirates	4
Yemen Arab Republic	1
European Region	
Algeria	1
Morocco	2
South-East Asia Region	
Bangladesh	1
Burma	3
India	14
Nepal	1
Sri Lanka	1
Thailand	4
Western Pacific Region	
Fiji	1
Philippines	1

Table 2. Distribution of 446 sample survey estimates by level of coverage and precision

Immunization coverage (%)	No. of estimates	No. of estimates with 95% confidence limits within:					
		± 10%		± 10.1 – 12.0%		± 12.1 – 13.0%	
		No.	%	No.	%	No.	%
< 5	0	0	0	0	0	0	0
5 – 14	32	32	100	0	0	0	0
15 – 24	36	35	97.2	1	2.8	0	0
25 – 34	37	26	70.3	8	21.6	3	8.1
35 – 44	50	33	66.0	14	28.0	3	6.0
45 – 54	58	29	50.0	20	34.5	9	15.5
55 – 64	56	42	75.0	10	17.9	4	7.1
65 – 74	54	49	90.7	5	9.3	0	0
75 – 84	61	61	100	0	0	0	0
85 – 94	56	56	100	0	0	0	0
⩾ 95	6	6	100	0	0	0	0
Total	446	369	82.7	58	13.0	19	4.3

Most surveys included determinations of coverage for a single dose of BCG vaccine and for first, second, and third doses of DPT vaccine. Several surveys also included determinations of immunization coverage of smallpox (1 dose), measles (1 dose), and poliomyelitis (3 doses).

The primary objective of the analysis was to determine what proportion of the results had 95% confidence limits within ± 10%, as had been desired in designing the method. It was found that the major variable associated with the size of the confidence limits was immunization coverage. This far outweighed any tendency for the confidence limits within a single survey, or for a given vaccine within a survey, to be influenced by the cluster design. For this reason, results have been analysed in terms of the immunization coverage observed, grouping all observations from all surveys pertaining to a given level of coverage.

In all, 446 sample estimations of immunization coverage were analysed (Table 2). The results are reassuring: 83% of the sample results had 95% confidence limits within ± 10%, and none of the results had 95% confidence limits exceeding ± 13%. Several of the surveys analysed included numbers of children that did not conform to the sample size prescribed, however, suggesting the possibility of methodological error. A sub-analysis was therefore performed using only those surveys in which a minimum of 7 children had been chosen in each cluster and in which a total of not more than 212 children had been chosen. There were 28 such surveys, comprising 209 observations (Table 3). Among this subgroup of surveys, 86% of the results had 95% confidence limits within ± 10%.

ANALYSIS OF RESULTS OF COMPUTER SIMULATED SURVEYS

Computer simulation was used to explore how well this survey method might perform. For this purpose, 12 hypothetical population strata were established with immunization coverage rates ranging from 99% to 10%, and 10 hypothetical communities were established by allocating to them various proportions of each of the strata (Table 4). The overall immunization coverage rates in these 10 communities varied between 60% and 87%.

By computer, 150 sample surveys were conducted in each of these 10 hypothetical communities. Each sample consisted of 30 clusters, with 7 individuals in each cluster. The probability of a given cluster falling within one of the 12 population strata was made proportional to the size of that stratum in the general population. The probability of an individual in a cluster being immunized was made proportional to the immunization coverage of that stratum. This simulation was therefore designed to test the survey method under circumstances in which it might be expected to perform least well; circumstances in which there is a wide variation in the immunization coverage among different subgroups of the popula-

Table 3. Distribution of 209 sample survey estimates by level of coverage and precision, in surveys with 210 – 212 children, in which all clusters contained at least 7 children

Immunization coverage (%)	No. of estimates	No. of estimates with 95% confidence limits within:					
		± 10%		± 10.1 – 12.0%		± 12.1 – 13.0%	
		No.	%	No.	%	No.	%
< 5	0	0	0	0	0	0	0
5 – 14	30	30	100	0	0	0	0
15 – 24	21	21	100	0	0	0	0
25 – 34	19	13	68.4	6	31.6	0	0
35 – 44	31	22	71.0	9	29.0	0	0
45 – 54	20	8	40.0	9	45.0	3	15.0
55 – 64	22	21	95.5	1	4.5	0	0
65 – 74	19	18	94.7	1	5.3	0	0
75 – 84	27	27	100	0	0	0	0
85 – 94	19	19	100	0	0	0	0
⩾ 95	1	1	100	0	0	0	0
Total	209	180	86.1	26	12.4	3	1.4

tion and in which there is a tendency for the immunization status of persons falling within one cluster to be correlated.

Fig. 1 illustrates how the results of the 150 samples obtained from 4 of the hypothetical communities are distributed, and Table 5 summarizes the proportion of survey results that fell within 10%, ± 10.1 – 12%, and ± 12.1 – 13% of the true results.

These simulated surveys support the validity of the method: more than 95% of the results were less than ± 10% from the actual population mean.

DISCUSSION

The precision of this method, as estimated from the results of both actual and simulated surveys, is considered satisfactory for the requirements of the Expanded Programme on Immunization. Among the actual surveys, the proportion of results whose confidence limits exceeded ± 10% was greatest (50%) when the immunization coverage observed in the sample was between 45% and 54%. But even within

Table 4. Description of the pattern of coverage in ten hypothetical communities, each with different percentages of the population in the different strata

Stratum No.	1	2	3	4	5	6	7	8	9	10	11	12	Overall coverage (%)
Coverage (%)	99	90	80	70	60	50	40	30	20	15	14	10	
Community 1	0	0	11	33	38	0	11	2	0	5	0	0	60.45
Community 2	15	0	13	30	17	3	13	6	0	0	3	0	65.37
Community 3	6	0	31	32	13	3	9	6	0	0	0	0	67.84
Community 4	6	0	41	18	22	0	5	5	0	3	0	0	68.49
Community 5	6	8	45	0	22	5	11	3	0	0	0	0	70.14
Community 6	12	0	39	21	15	2	7	4	0	0	0	0	71.78
Community 7	19	0	32	28	11	0	6	3	0	1	0	0	74.06
Community 8	38	0	35	10	10	2	4	1	0	0	0	0	81.52
Community 9	46	0	29	14	2	4	3	2	0	0	0	0	83.54
Community 10	51	0	40	0	5	0	3	1	0	0	0	0	86.99

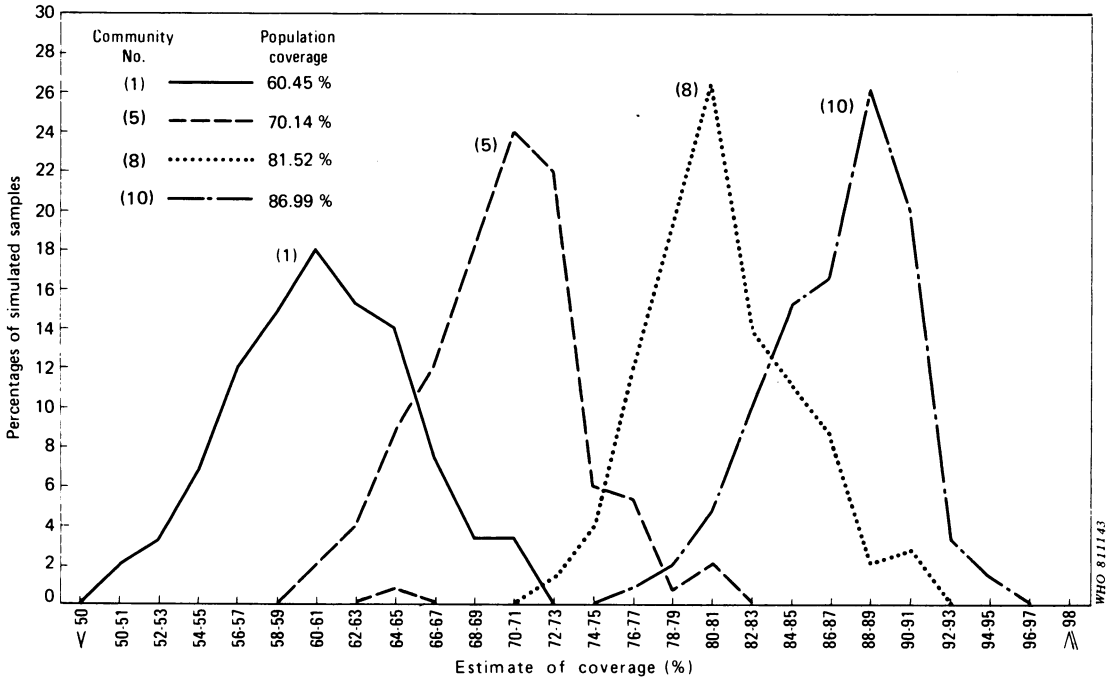

Fig. 1. Percentage of simulated samples giving various estimates of immunization coverage, in four hypothetical communities (see Table 4) with different patterns and levels of immunization coverage.

Table 5. Distribution of 1500 computer simulated survey estimates by level of coverage and precision

Hypothetical community	Immunization coverage (%)	No. of survey estimates	No. of estimates with 95% confidence limits within:					
			± 10%		± 10.1 – 12.0%		± 12.1 – 13.0%	
			No.	%	No.	%	No.	%
1	60.45	150	150	100.0	–	–	–	–
2	65.37	150	148	98.6	1	0.7	1	0.7
3	67.84	150	146	97.3	3	2.0	1	0.7
4	68.49	150	144	96.0	4	2.7	2	1.3
5	70.14	150	147	98.0	3	2.0	–	–
6	71.78	150	149	99.3	1	0.7	–	–
7	74.06	150	148	98.6	1	0.7	1	0.7
8	81.52	150	149	99.3	–	–	1	0.7
9	83.54	150	148	98.6	1	0.7	1	0.7
10	86.99	150	150	100.0	–	–	–	–
	Total	1500	1479	98.6	14	0.9	7	0.5

this range, only 16% of the results had confidence limits exceeding ± 12% and none of them had confidence limits exceeding ± 13% (Table 2). In surveys containing 210 – 212 children in which there were at least 7 children per cluster, a similar pattern was observed, although the proportion of results whose confidence limits exceeded ± 10% increased to 60% at immunization coverage levels of 45 – 54%.

As expected on theoretical grounds, the confidence limits narrowed as the observed immunization coverages approached either 100% or 0%. But, as observed in Tables 2 and 3, this narrowing was not symmetrical: when coverage was below 50% the results had wider confidence limits than when coverage was a corresponding amount above 50%. This may be either a chance phenomenon or may indicate that the distribution of immunized children in populations whose overall immunization coverage rate is below 50% tends to be less homogeneous than the distribution of unimmunized children in populations with immunization coverage rates above 50%. In other words, a highly selected group of children may be the beneficiaries of scarce immunization services, while never or partially immunized children may tend to be more evenly distributed in populations where services are more plentiful.

Because of particular interest in the performance of the survey method at high levels of coverage, the simulated surveys were limited to populations in which the coverage ranged from 60% to 87%. However, there is no reason to suppose that the simulation results would not be identical in symmetrically constructed populations with immunization coverages of 40 – 13% (that is to say, in populations whose proportion of unimmunized was 60 – 87%).

The simulated survey results had narrower confidence limits than did the actual surveys, 95% of them being with ± 10%, as opposed to 89% of the actual survey results in populations where coverage was similar (55 – 84% in the actual surveys). There are two main reasons for this. First, the 1500 simulated surveys reflected only 10 different populations, whereas the 60 actual surveys each reflected a separate population, giving results which might be expected to be more heterogeneous. In addition, the sampling under simulated conditions used absolutely uniform procedures throughout, whereas in the actual surveys it is probable that a number of departures from the recommended procedures occurred. While it would be hoped that errors introduced into the actual survey procedures would be random and would tend to cancel each other out in the final result, they would be expected to decrease the precision of the final result. This latter hypothesis is strengthened by the analysis of actual surveys in Table 3, where methodological errors may have been fewer. Among these, 97% of the results in the coverage range of 55 – 84% had confidence limits within ± 10%.

Both in the actual and simulated surveys, "clusters" of children were selected, as opposed to choosing every child at random. The consequence can be expressed quantitatively by a measure called the "design effect" which is the ratio of the variance of the cluster sample result to the variance of the result from a simple random sample of the same size. If each child within a cluster were to have the same chance of being immunized as every other child in that cluster, the value of the design effect would be equal to the cluster size (namely 7, in this instance). Under these circumstances, the precision of the sample estimate would be equivalent to a simple random sample in which 30 children (corresponding to one per cluster) were chosen. At the other extreme the "design effect" would have the value of "1" if the children within a given cluster had no tendency for their immunization status to be correlated with one another. Under these circumstances, the precision of the sample estimate would be equivalent to a simple random sample in which 210 children were selected. (Values of the

Table 6. Coverage levels required in a second survey to establish a statistically significant change from the level observed in the first survey ($P \leqslant 0.05$, two-tailed test)[a]

First survey observed coverage (%)	Assumed design effect[b]	Second survey: Minimum coverage to detect improvement (%)	Second survey: Maximum coverage to detect deterioration (%)
10	2.30	20.6	2.8
15	2.20	26.7	6.2
20	2.64	33.9	8.9
25	2.38	38.9	13.3
30	2.74	45.5	16.5
35	3.10	51.9	19.8
40	2.72	55.9	25.1
45	3.06	61.9	28.7
50	4.08	69.1	30.9
55	2.92	71.0	38.5
60	3.00	75.6	43.3
65	2.83	79.5	48.9
70	2.60	83.2	54.9
75	2.70	87.4	60.1
80	2.40	90.6	66.7
85	2.50	94.3	72.5
90	2.40	97.3	79.2
95	1.84	99.4	87.6

[a] 30 clusters of 7 children assumed in both surveys. The figures given should be taken only as approximate indications, as the intra-cluster correlation is unlikely to be constant (see footnote[b]).

[b] The intra-cluster correlation for corresponding coverage levels is assumed to be the same for the 2 surveys and gives rise to the "design effect" as indicated. These "design effects" have been calculated from the 60 actual surveys analysed in this paper. For each level of coverage shown, the design effect would be at, or smaller than, the level shown for 90% of the results analysed.

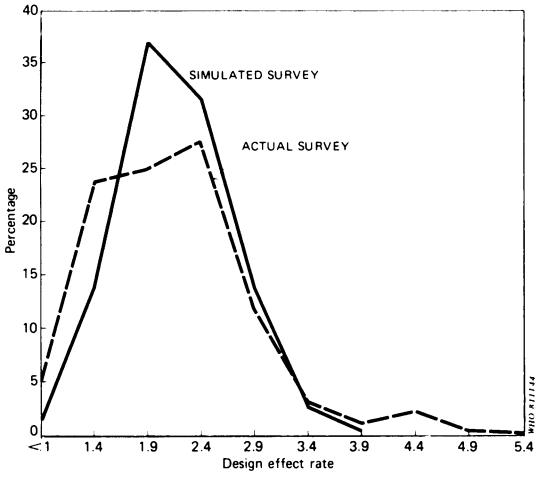

Fig. 2. Distribution of "design effects" in simulated and actual surveys.

design effect less than "1" can occur rarely , and indicate a high degree of uniformity *between* clusters. Values of less than "1" were observed in 5% of the actual survey results and 1% of the simulated survey results.)

Fig. 2 illustrates the distribution of the values of the "design effect" in relation to coverage estimates for both the actual[a] and simulated surveys. These values generally lie between 1.5 and 2.5 with more than 90% being less than "3". This reflects a moderate tendency for children within a given cluster to have similar immunization histories.

As immunization programmes become established, primary concern will shift from estimation of an initial coverage rate to documenting trends in coverage. Table 6 indicates for various initial immunization coverage rates, the amount of difference that two surveys, performed at two different times in the same population according to the currently recommended EPI method, would need to show before one could conclude that the results were significantly different. This table is based on the 60 actual surveys included in this analysis, and can serve as a general guide for programme managers. For a specific determination of significance, analysis of the two surveys in question would be required.

[a] The distribution of values for the "design effect" in the subsample of surveys analysed in Table 3 was not significantly different from the distribution of values from all actual surveys, and the values of the latter are used in the remaining analyses.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the help of national and international EPI staff in providing results of immunization coverage surveys. We are also grateful to Ms Tonya Price-Jones, formerly short-term consultant with the EPI, for the help in preparing the early drafts of this paper, and to Mr Donald Eddins, Immunization Division, Centers for Disease Control, for help with several technical questions. We should also like to thank specially Miss D. Magnin, Health Statistics Methodology, WHO, Geneva, for help in analysing the actual surveys, in performing the simulations, and in preparing the summary tables and graphs.

RÉSUMÉ

SONDAGE PAR GRAPPE POUR ÉVALUER LA COUVERTURE VACCINALE: EXAMEN D'UNE EXPÉRIENCE SUR UNE MÉTHODE SIMPLIFIÉE DE SONDAGE

Pour estimer la couverture de son programme élargi de vaccination, l'OMS a utilisé une méthode de sondage par grappe faisant appel à la sélection aléatoire de 210 enfants, soit 30 grappes de 7 enfants chacune. Cette méthode a été conçue pour fournir des limites de confiance ne dépassant pas ± 10% en valeur absolue de la couverture estimée, avec un risque d'erreur de 5%.

On a analysé les résultats de 60 enquêtes sur la couverture vaccinale conduites dans 25 pays en développement. Au total, 446 échantillons ont fourni autant d'estimations individuelles; pour 83% d'entre elles, l'intervalle de confiance au seuil de probabilité de 95% est tombé dans les limites de ± 10%, alors que pour la totalité des estimations le même intervalle de confiance n'est pas sorti des limites de ± 13%. L'étendue de l'intervalle de confiance a largement dépendu du niveau de couverture observée, l'intervalle le plus large correspondant à des niveaux voisins de 50%. Lorsque la couverture vaccinale observée dans l'échantillon était comprise entre 45% et 54%, l'intervalle de confiance n'est tombé dans les limites de 10% en plus et en moins de l'estimation que dans la moitié des cas. C'est seulement lorsque la couverture de la vaccination atteignait au moins 75% que l'intervalle de confiance tombait sans exception dans ces limites de ± 10%.

On a par ailleurs analysé les résultats de 150 enquêtes simulées sur ordinateur et conduites dans 10 populations théoriques construites à cet effet, avec une couverture vaccinale variant de 60% à 87%. Les intervalles de confiance ont

été dans l'ensemble plus étroits pour les couvertures simulées que pour les couvertures réellement observées: dans le cas des estimations simulées, 95% de leurs intervalles de confiance sont restés à l'intérieur des limites de ± 10%, alors que cette condition n'a été remplie que pour 89% des estimations faites sur le terrain, à niveau de couverture vaccinale comparable. Cette différence est attribuée au fait que les simulations ont été produites à partir de 10 populations, alors que les enquêtes sur le terrain en ont couvert 60; de plus, les procédures de simulation sont d'une uniformité rigoureuse, alors que la sélection des échantillons et l'exécution des enquêtes sur le terrain sont sujettes à des erreurs occasionnelles inévitables.

Dans les 60 enquêtes analysées, la méthode de sondage par grappe a introduit un «effet de plan» égal à environ 3, ce qui signifie que, pour obtenir des résultats de même précision, le nombre des personnes à inclure dans l'enquête avec sondage par grappe doit être 3 fois plus élevé que dans le cas d'une sélection strictement aléatoire des individus. Cela reviendrait donc à porter à 10 enfants la dimension de chacune des 30 grappes choisies. Le nombre total des enfants à échantillonner passerait alors de 100 à 300 au lieu de 100 à 210 comme dans cette étude où «l'effet de plan» avait été supposé voisin de 2; cependant le gain en précision qui pourrait en résulter n'est, du moins pour l'instant, pas suffisant pour justifier le coût de l'opération.

REFERENCES

1. SERFLING, R. E. & SHERMAN, I. L. *Attribute sampling methods.* Washington Public Health Service, 1965 (Publication No. 1230).
2. HENDERSON, R. H. ET AL. Assessment of vaccination coverage, vaccination scar rates, and smallpox scarring in five areas of West Africa. *Bulletin of the World Health Organization*, **48:** 183 – 194 (1973).
3. COCHRAN, W. C. *Sampling techniques.* New York, Wiley, 1963.
4. *Weekly epidemiological record*, **54** (25): 193 – 194 (1979); **54** (29): 221 – 223 (1979); **54** (35): 265 – 266 (1979); **54** (50): 385 – 386 (1979); **55** (9): 66 – 67 (1980); **55** (18): 132 – 133 (1980); **55** (24): 177 – 179 (1980); **55** (32): 244 – 245 (1980); **55** (39): 297 – 300 (1980); **55** (42): 321 – 323 (1980); **55** (43): 329 – 330 (1980); **56** (12): 89 – 90 (1981); **56** (14: 109 – 110 (1981); **56** (15): 114 (1981); **56** (28): 220 – 221 (1981); **56** (29): 227 – 228 (1981).