

# Increasing the Accuracy of the Expanded Programme on Immunization's Cluster Survey Design

DONNA BROGAN, PHD, ELAINE W. FLAGG, PHD, MICHAEL DEMING, MD, MPH,  
AND RONALD WALDMAN, MD, MPH

## ABSTRACT

The Expanded Programme on Immunization (EPI) of the World Health Organization developed a cluster survey design for the rapid assessment of children's vaccination coverage in developing countries. Because of its simplicity and familiarity, the EPI methodology has been used for additional purposes, including the detection of relatively small changes in vaccination coverage levels that are already high and the estimation of other population parameters when moderate accuracy is required. This article suggests techniques for improving the accuracy of the EPI cluster survey method, including (a) segmenting sample clusters, if necessary, so that reasonably sized areas can be counted, mapped, and listed; (b) selecting an equal probability sample of housing units within a cluster; (c) selecting a fixed number of housing units per sample cluster; and (d) performing weighted analyses. These procedures will produce accurate estimators and corresponding standard errors in EPI cluster surveys. *Ann Epidemiol* 1994;4:302-311.

KEY WORDS: Health surveys, sampling studies, immunization, statistics, survey methods.

## USE OF CLUSTER SURVEYS

The Expanded Programme on Immunization (EPI) of the World Health Organization has popularized the use of a particular cluster sample survey design in developing countries for rapid assessment of vaccination coverage. The survey was designed for ease of use and to provide cross-sectional estimates of coverage with approximate accuracy and fairly wide confidence limits. Even under difficult field conditions, the survey has proved to be affordable and relatively easy to implement and analyze. In fact, by March 1992, 4502 such surveys have been reported to the World Health Organization for the years 1978 to 1991; 229 of them were conducted in 1991 (1).

The health goals for countries endorsed by the World Summit for Children in 1990 include "the maintenance of a high level of immunization coverage (at least 90 percent of children under one year of age by the year 2000) against diphtheria, pertussis, tetanus, measles, poliomyelitis, tuberculosis, and against tetanus for women of child-bearing age" (2). Many developing countries are close to achieving 90% immunization coverage with antigens given to children, or

have already achieved it (3). A number of countries with relatively high levels of coverage now wish to use EPI surveys to evaluate efforts to achieve relatively small increases in coverage beyond levels that are already high or to maintain a high level of coverage. Because of its familiarity, the EPI survey method is also being used when relatively accurate estimates of other parameters are needed, particularly when public health programs are evaluated to determine if they have achieved their objectives.

We are concerned that the EPI survey method, as originally developed, may not be adequate for its newer uses. Hence, we suggest specific, feasible changes that can be made in the EPI survey design when greater accuracy is needed. The term "accuracy" is used to denote how close the sample estimator is, on the average, to the population parameter being estimated (4). An unbiased (or approximately unbiased) estimator is desirable because on the average, its value will equal the population parameter. We show that current EPI survey methods may produce biased estimators, and the proposed changes will result in approximately unbiased estimators.

Surveyors are also concerned about the "precision" of an estimator, that is, its variance, standard error, or mean square error (4, pp. 15-16). EPI surveyors already are familiar with the need to increase total sample size (more sample clusters and/or more persons per cluster) when more precision is desired (5). Some recent examples are surveys on nutritional status (6), neonatal tetanus mortality (7), infant mortality, maternal mortality, and contraceptive prevalence. Hence, this article focuses on increasing the accuracy

From the Division of Biostatistics (D.B.) and the Division of Epidemiology (E.W.F.), Emory University School of Public Health, Atlanta, and Technical Support Division, International Health Program Office, Centers for Disease Control (M.D., R.W.), Atlanta, GA.

Address reprint requests to: Donna Brogan, PhD, Division of Biostatistics, Emory University School of Public Health, 1599 Clifton Road NE, Atlanta, GA 30329.

Received March 25, 1993; revised December 8, 1993.

of estimators and estimated standard errors. As such, it is a contribution to Anker's call for "... further work in modifying and validating the EPI methodology..." (8, p. 95).

### FEATURES OF THE EPI CLUSTER DESIGN

The EPI cluster survey design was adapted from previous work by Serfling and Sherman (9) on designing immunization surveys to be conducted by local health departments in the United States. The primary objective of the EPI cluster survey, as originally proposed, is to estimate the proportion  $P$  (a population parameter) of children of a given age in a given geographic area who are vaccinated. A 95% confidence interval on  $P$  is desired such that its half-width is 0.10 or smaller or, equivalently, that  $\hat{P}$ , the point estimate of  $P$ , should be within  $\pm 0.10$  of  $P$  at the 95% confidence level.  $P$  is assumed to equal 0.50 for the purpose of calculating sample size, yielding a conservative (maximum) estimate of 96 as the required sample size under the assumption that a simple random sample of children is drawn from the population of inference (4, p. 75).

However, since simple random sampling of child populations is not feasible, the EPI surveys use cluster sampling, which requires the simple random sample size of 96 children to be inflated by an estimated design effect (4). With an assumed design effect of 2, the required sample size is 192 children. Thus, 30 sample clusters, each containing seven sample children, is the typical EPI design.

The first-stage sampling frame for the EPI cluster design is a list of mutually exclusive and exhaustive primary sampling units (PSUs) or clusters defined as geographic and/or political areas such as villages, cities, parts of cities, and rural zones. The measure of size of the PSU usually is estimated total population. The PSUs generally are ordered on the sampling frame by geographic proximity. Systematic probability proportional to estimated size (PPES) sampling is used to select a fixed number (at least 30) of sample PSUs from the ordered frame.

Various within-PSU sampling procedures are used in EPI surveys. A commonly used procedure is the "more or less" random selection of one starting housing unit (HU) from a sample PSU without doing the required counting, mapping, and listing to construct a sampling frame for HUs (5). Spinning a bottle or pencil at the center of a PSU (e.g., the market) is often done to select a direction for choosing the first sample HU. The HUs that lie along this line from the center to the border of the PSU are listed, and one is chosen at random. The "next-nearest" HU, as judged by the interviewer, is selected as the second HU in the sample and subsequent "next-nearest" HUs are visited until a HU that contains the seventh child for that PSU is selected. Since all children are selected in each HU, more than seven children may be

selected in a sample PSU if the last HU contains more than one eligible child. Selection of more than seven children per PSU is infrequent since the eligibility age range generally is narrow.

The statistical analysis techniques used for the EPI survey assume an equal probability sample of children from the population of inference. Hence, point estimates and their estimated variances are obtained using unweighted statistical analyses. Differential sampling weights are taken into account when surveys are stratified and sampling fractions vary from stratum to stratum; however, an equal probability sample of children within a stratum is assumed.

Formulas or descriptions for the point estimate  $\hat{P}$  and the estimated variance of  $\hat{P}$  appear in several publications that discuss the EPI cluster survey method (10-14). If the survey is not stratified and has the same number of children sampled per PSU, then

$$\hat{P} = \sum_{i=1}^n p_i/n \quad 1$$

where  $p_i$  is the proportion of surveyed children in PSU  $i$  who are vaccinated and  $n$  is the number of PSUs/clusters in the survey. The approximate estimated variance of  $\hat{P}$  with an equal number of children sampled per PSU is given by

$$v(\hat{P}) \doteq \sum_{i=1}^n (p_i - \hat{P})^2 / [n(n-1)]. \quad 2$$

Equations 1 and 2 may be generalized to account for an unequal number of children sampled per PSU by letting  $y_i$  denote the number of children sampled in PSU  $i$  who are vaccinated and letting  $m_i$  be the total number of children sampled in PSU  $i$ . Then

$$\hat{P} = \sum_{i=1}^n y_i / \sum_{i=1}^n m_i \quad 3$$

and the approximate estimated variance of  $\hat{P}$  is

$$v(\hat{P}) \doteq (\sum_{i=1}^n y_i^2 - 2\hat{P}\sum_{i=1}^n y_i m_i + \hat{P}^2 \sum_{i=1}^n m_i^2) / [n(n-1)\bar{m}^2] \quad 4$$

where  $\bar{m} = (1/n) \sum_{i=1}^n m_i$  is the mean number of children sampled per PSU. An approximate 95% confidence interval on  $P$  can be obtained from equations 1 and 2 or from equations 3 and 4 by using  $\hat{P} \pm 1.96 \sqrt{v(\hat{P})}$ .

### EPI SURVEY DESIGN ASSUMPTIONS

The EPI survey design makes two assumptions that allow the use of equations 1 through 4. First, the sample of children is assumed to be a probability sample, and further, it is assumed to be an equal probability sample of children. Second, nonresponse at the HU level and at the respondent level is assumed to be nonexistent. The discussion below maintains that the equal probability sample assumption is

unlikely to be true in EPI surveys, as well as in all cluster surveys. The second assumption may or may not hold, depending on local conditions. Each unsatisfied assumption leads to a biased estimator.

**Assumption 1: The Sample Obtained Is an Equal Probability Sample of Children**

Standard EPI procedures at both the first and second stages of sampling contribute to the violation of this assumption. At the first sampling stage, the typical measure of PSU size in EPI cluster surveys, estimated total population, may not yield an equal probability sample of children because it generally is not accurate, being out of date and/or based on population data of dubious quality. The preferred measure of PSU size for the first stage of sampling is actually the number of HUs per PSU rather than total population per PSU since HUs are the second-stage sampling units (5), but estimates of this measure of size may also be inaccurate.

At the second stage of sampling, the commonly used EPI field procedures for selecting the starting HU, as described earlier, are not likely to satisfy the assumption of equal probability sampling of HUs. In standard practice, the sample area is a randomly selected rectangle extending from the assumed center of the PSU to the boundary of the PSU. If all such possible rectangles are considered, the HUs near the center of the PSU fall into more rectangles, thus giving them a higher probability of selection. Another deviation of EPI second-stage sampling from equal probability sampling of HUs is that interviewer judgment is involved in determining which HU is the next nearest. In fact, this judgmental selection of HUs does not satisfy the definition of a probability sample of HUs at the second stage, leading to a sample of children that is not even a probability sample. Clustering of nearby HUs at the second stage of sampling is a standard technique to reduce field costs, but the technique yields a probability sample only if it is objectively applied, that is, not dependent on interviewers' subjective judgments.

To illustrate that the standard EPI method may not result in an equal probability sample of children, let  $T'_i$  be the estimated total population of PSU  $i$  and let

$$T' = \sum_{i=1}^N T'_i$$

be the estimated total population for the  $N$  clusters that comprise the geographic area of inference. The probability that PSU  $i$  is chosen for the sample is  $30 \times T'_i/T'$ . The number of HUs selected at the second stage in sample PSU  $i$  is not a predetermined constant; it equals the required number of HUs to visit in order to identify seven age-eligible children, approximately  $7 \times H_i/C_i$ , where  $H_i$  = the number of HUs in PSU  $i$  and  $C_i$  = the number of age-eligible children in PSU  $i$ . Assuming that an equal probability sample of

HUs is selected (which most likely is not true), then the probability of selecting a particular HU  $u$  in PSU  $i$  is approximately  $7/C_i$ . Hence, the probability  $\text{Pr}[i,u,j]$  of selecting child  $j$  in HU  $u$  in PSU  $i$  is given approximately by

$$\text{Pr}[i,u,j] = 30 \times (T'_i/T') \times (7/C_i) \times 1 \quad 5$$

This approximate selection probability in equation 5 is constant for every child within a PSU but varies across PSUs. This selection probability will be constant across PSUs, resulting in an equal probability sample of children, only if  $C_i/T'_i = d$  (a constant) for all  $i = 1, 2, \dots, N$ . This condition is satisfied only if the ratio of the current number of age-eligible children to the estimated total population has the same value in each of the  $N$  PSUs in the population. A sufficient set of conditions for this equality is that (a) the estimated population  $T'_i$  be a fixed percentage, over all PSUs, of the actual current population  $T_i$ , that is,  $T'_i = yT_i$  (for example,  $y = 80\%$ ); and (b) the percentage of the current population  $T_i$  that is in the eligible age range is constant over all PSUs, that is,  $C_i = zT_i$ . Under these two conditions  $C_i/T'_i = zT_i/yT_i = z/y$ , a constant.

These two conditions, however, seem unlikely since (a) fertility rate and age structure of the population are probably not exactly the same in all PSUs, and (b) it is unlikely that the proportional error of the estimated population relative to the true population is the same for all PSUs. Even if these conditions seemed plausible, however, the probability in equation 5 is only an approximation because (a) the sample size for HUs is a random variable, and (b) the sample may not even be a probability sample. In all, it seems unlikely that the EPI method really will yield an equal probability sample of children.

An unweighted analysis, based on the assumption of an equal probability sample of children, may give three misleading results. First, the estimator  $\hat{P}$  may be seriously biased for  $P$ . For example, if HUs close to the center of the village (PSU) are more likely to be chosen than outlying HUs, and if children in the closer-in HUs are more likely to be vaccinated, then the unweighted point estimate will be biased in the direction of overestimating coverage. Second, the estimated variance of  $\hat{P}$  will be underestimated since estimated variances with differential sampling weights (based on the inverse of the selection probability) are larger than when the sampling weights are all equal to each other (15). The extent of this potential bias and almost certain variance underestimation increases as the selection probabilities for the sample children become more heterogeneous. Third, the use of variance in the presence of bias, rather than mean square error, underestimates the overall precision of the point estimate.

In summary, use of total population as the measure of size, use of inaccurate measures of size, selection of the first HU in each cluster with unequal probabilities, and selection of subsequent HUs with some interviewer subjectivity do

not lead to an equal probability sample of HUs, and hence do not lead to an equal probability sample of children. Analysis of the data as though it were an equal probability sample can lead to biased point estimates and underestimated variability in the point estimates.

#### **Assumption 2: Not at Homes and Refusals Do Not Occur**

The earlier formulas for the point estimate and variance estimation assume no nonresponse. HU nonresponse occurs when an interviewer does not find anyone at home at an occupied sample HU or when the occupants refuse to identify HU members eligible for the survey. Respondent nonresponse occurs when a selected eligible respondent refuses to participate in the survey or is too sick or incapacitated to participate. Substituting another HU for a nonresponding HU or another respondent for a nonrespondent runs the risk of yielding biased estimators if the vaccination characteristics of subjects or HUs that are respondents differ from the vaccination characteristics of subjects or HUs that are nonrespondents.

If nonresponse (both HU and eligible respondent) has been found to occur rarely or not at all in the population being surveyed, the assumption of 100% response may be reasonable. However, if nonresponse is not negligible, and if respondents and nonrespondents are likely to differ on important survey variables (e.g., vaccination status), then the substitution method should not be used and standard field and/or analytic procedures to deal with nonresponse should be considered (16).

---

### **RECOMMENDED MODIFICATIONS IN THE EPI CLUSTER SURVEY DESIGN FOR GREATER ACCURACY**

The following recommended modifications in the EPI cluster design will result in a probability sample of children, most likely an unequal probability sample, and approximately unbiased estimators.

#### **1. Continue to Use PPES Sampling at the First Stage**

PPES (probability proportional to estimated size) sampling commonly is used at the first stage when it is desired to obtain an equal (or near-equal) probability sample of persons residing in a given geographic area. The preferred measure of PSU size is the estimated number of HUs, but the estimated number of children or estimated total population can be used if HU data are not available. However, if available estimates of number of HUs per PSU are less accurate than estimates of number of children or total population, then it may be better to use the population estimates. The

designer of the sample needs to take into account the accuracy of available size measures as well as the best match to the second stage of sampling.

#### **2. A Probability Sample, and Preferably an Equal Probability Sample, of HUs Should Be Selected within Each Sample PSU**

Selection of a probability sample of HUs ensures that (a) selection probabilities of children can be calculated and (b) formulas for point estimates and their estimated variances can be validly applied to the data. It is standard practice in area sampling to select an equal probability sample of HUs.

Understandably, EPI surveyors have been reluctant to adopt the recommendation of a probability sample of HUs because of the assumed necessity of counting, mapping, and listing all HUs in each sample PSU, an especially difficult and time-consuming task in places where transportation within sample PSUs is not easy or when the population of PSUs is large. Recommendations (a) and (b) below suggest a method to accomplish the necessary counting, mapping, and listing activities at minimum cost.

(a) **Small-sample PSUs can be totally counted, mapped, and listed.** In a small-sample PSU with few HUs or a small geographic area, it is feasible to use standard survey procedures to locate, count, map, and list all HUs. Counting is typically done first to ensure that it is feasible to map and list the entire PSU. A stage 2 sampling frame then is constructed by numbering and listing the HUs from 1 to  $H_i$ , where  $H_i$  is the number of HUs located in the  $i$ th sample PSU. The location of each listed HU is indicated on the map or sketch of the PSU.

All sample PSUs can be totally counted, mapped, and listed if they are appropriately defined as small enough PSUs (or clusters) on the stage 1 sampling frame. However, this procedure requires estimated measures of size for all of these smaller PSUs. In practice, it is probably unrealistic to assume that all PSUs can be defined small enough, with an accompanying measure of size, so that counting, mapping, and listing is feasible for all HUs in each PSU.

(b) **Segmenting and subsegmenting can be used to count, map, and list larger sample PSUs.** It is desirable to count all HUs in the PSU, even though the count may be rough or approximate in order to minimize cost. If a large number of HUs are counted or estimated, then the PSU can be divided into smaller segments. Geographic boundaries such as roads, rivers, and villages can be used to define segments. One segment is randomly chosen to map and list, using probability proportional to size (the count) sampling. The mapping and listing time will be considerably less for the chosen segment than for the entire PSU.

If a PSU covers a wide geographic area and/or has many HUs, it may take too many resources to count all the HUs.

In this case, it may be possible to gain information from local residents that allows segmenting and estimation of HUs per segment. One of the segments is selected with probability proportional to estimated size. If the chosen segment is small enough to count, map, and list in its entirety, then the mapping and listing procedures in recommendation (a) above are followed.

If the chosen segment is still too large to map and list, then it in turn can be counted and subsegmented as described above. One subsegment is chosen with probability proportional to size (the count), and the chosen subsegment can be mapped and listed as in (a) above.

Estimated sizes of segments and subsegments can be rough and approximate if necessary resources for accurate counting are not available. Inaccuracies in these estimated sizes will contribute to an unequal probability sample of children, just as inaccurate measures of size at the first stage of sampling contribute to an unequal probability sample. Even though the estimated measures of size at the various stages of sampling may be inaccurate, it is possible and straightforward to calculate a selection probability and sampling weight for each sample child.

A given child's sampling weight, to be used in a weighted analysis of the survey data, is defined as the inverse of the selection probability for that child. The sampling weight can be interpreted as the number of children in the population that the sample child represents. When the sample of children is not an equal probability sample, then a weighted estimator  $\hat{P}$  of  $P$  can be constructed which is approximately unbiased (17). Equations 1 and 3 for  $\hat{P}$  should not be used when a weighted estimator is needed, that is, when the sampled children have unequal selection probabilities and hence, unequal weights.

Unequal selection probabilities and unequal weights also contribute to an increased standard error of  $\hat{P}$ , but this generally is an acceptable trade-off for avoiding the counting, mapping, and listing of large PSUs and for obtaining an approximately unbiased estimator of  $P$  by using a weighted analysis.

The result of counting, mapping, and listing all or part of the sample PSU is the construction of a stage 2 sampling frame of HUs, a prerequisite for obtaining a probability sample of HUs. Note that segmenting and subsegmenting procedures avoid travel over the entire PSU for mapping and listing and also concentrate the sample of HUs in one geographic area of the sample PSU. A detailed discussion of mapping and listing issues in developing countries is given by Sunter (18).

(c) **An equal probability sample of a fixed number ( $k$ ) of housing units should be selected from each sample PSU.** An equal probability sample of HUs from the stage 2 sampling frame may be obtained via simple random sampling, systematic random sampling, or cluster sampling of HUs. With simple random sampling, the sample of HUs

will be spread over the mapped and listed area of the sample PSU. This dispersion should not result in high travel costs for small-sample PSUs or for large PSUs that are segmented and subsegmented.

Systematic random sampling of HUs generally is easier to implement in the field, compared to simple random sampling. Systematic sampling of HUs also will yield a sample of HUs dispersed over the mapped and listed area of the PSU.

A cluster sample of HUs from the stage 2 sampling frame can be selected by randomly choosing one HU with equal selection probabilities, followed by the selection of the next ( $k - 1$ ) HUs above or below the starting HU on the frame. Since HUs adjacent on the ground are adjacent on the sampling frame, the cluster sample should incur less travel time than either a simple random sample or a systematic random sample.

The selection of a cluster sample from the HU sampling frame is conceptually similar to the current EPI procedure of next-nearest HU selection. However, there are two major differences between the cluster sample recommendation here and current EPI procedure. First, the starting HU is selected from a HU sampling frame based on mapping and listing, and every HU on the frame has a known and equal probability of selection. Second, interviewer judgment is absent from the selection of any HUs for the sample.

Census or other data can be used to determine  $k$ , the number of HUs to be selected in each sample PSU. To obtain an average of seven age-eligible children within each sample PSU, the value for  $k$  will be  $7H'/C'$ , where  $H'$  is the estimated number of HUs and  $C'$  is the estimated number of age-eligible children in the population of inference. The determination of  $k$  may also be influenced by the number of HUs the field team can visit in 1 or 2 days (5).

The recommendation of selecting  $k$  HUs per sample PSU will result in an unequal probability sample of children when measures of size are not accurate or exactly proportional at any stage of sampling. It is possible to calculate a fixed sample size of HUs separately for each sample PSU, based on the mapping, counting, and listing activities in that PSU, so that an equal probability sample of children will be maintained over all sample PSUs. However, we recommend  $k$  HUs per sample PSU because we believe it is easier to conduct a weighted analysis on an unequal probability sample of children than it is to vary HU sample size and amount of field work over sample PSUs.

(d) **All eligible children should be selected within each housing unit sampled.** The selection of all eligible children within each sample HU, in conjunction with the above, will produce an equal or near-equal probability sample of children under very particular conditions. The probability of child  $j$  in HU  $u$  in PSU  $i$  being selected for the sample, assuming no segmenting in any PSUs, is

$$\Pr[i, u, j] = 30 \times (H_i'/H) \times (k/H_i) \times 1 \quad 6$$

A near-equal probability sample of children will be obtained if  $H_i'$  (the estimated number of HUs in PSU  $i$ ) is close to  $H_i$  (the actual number of HUs in PSU  $i$ ). An equal probability sample of children will be obtained if  $H_i' = H_i$  or if  $H_i = rH_i'$  for all  $i$ , where  $r$  is a constant.

If no HU size measure is available and estimated population is used, then, with no segmenting,

$$\Pr[i, u, j] = 30 \times (T_i'/T') \times (k/H_i) \times 1 \quad 7$$

An equal probability sample of children will be obtained if  $T_i'/H_i$  is constant over all PSUs, that is, if the mean number of estimated persons per HU is constant over all PSUs.

When a fixed number of HUs is selected within each sample PSU and all age-eligible children within a sample HU are selected, the number of sample children may vary among sample PSUs. However, this variation will not be dramatic if the mean number of age-eligible children per HU is fairly constant over all PSUs in the population.

If the mean number of children per HU varies dramatically among PSUs, and an estimate of this mean number is available for each PSU, then it may be desirable to calculate the fixed number of sample HUs separately for each sample PSU, so that an anticipated seven (or other value) children are obtained within each sample PSU. This is certainly a valid survey design, although it may introduce unwanted complexity into the planning of field work and it definitely will yield an unequal probability sample of children.

### 3. Anticipate That an Unequal Probability Sample of Children May Be Selected

Although the desire for an equal probability sample of children is understandable so that statistical analysis can be simplified, imperfect measures of size at all stages of sampling make it unlikely that the EPI cluster survey method (or any multistage and cluster survey!) achieves this objective. Unequal probability samples are common and quite acceptable, and they require that weighted statistical analyses be done in order to obtain approximately unbiased estimators.

With the recommended procedures above, the selection probabilities are constant within a sample PSU, although they most likely differ across sample PSUs. Hence, the regular formulas for weighted analyses simplify considerably and are not significantly more complicated than equations 3 and 4. It is necessary to record the selection probabilities at each stage of sampling (PSU, segment (if done), subsegment (if done), and HU) so that appropriate selection probabilities and sampling weights for analysis can be calculated.

Thus, the recommendations above plan for an equal probability sample of children, using the best measures of size available. However, it should be recognized that an equal probability sample is not likely to be obtained, and records should be kept during all sampling stages so that

selection probabilities and sampling weights can be calculated.

### 4. Calculate Selection Probabilities and Sampling Weights of Children within Each PSU

The result of the three recommendations above is a probability sample of children for whom the selection probabilities may be unequal but can be easily calculated. Large deviations from an equal probability sample of children are unexpected if reasonably accurate measures of size are used at each stage of sampling. Since the sampling at all stages of selection is carried out independently, the selection probability for each child is the product of that child's selection probabilities at each stage of sampling. See Appendix I for examples of how to calculate these selection probabilities.

### 5. Minimize Nonresponse; Substantial Nonresponse Should Be Adjusted for during the Analysis

Standard practice in sample surveys is to call back several times in order to contact "not-at-homes" (17, pp. 536-558). If repeated callbacks fail to find anyone at home, it may be possible to obtain information on the occupants of the sample HU from neighbors in order to determine if the HU contains any eligible respondents.

Generally, it is not a good idea to allow interviewers to replace "not-at-homes" or "housing unit refusals" or "eligible respondent refusals" with occupied HUs where the occupants agree to participate in the survey; this procedure runs the risk of yielding biased estimators. It is recommended that the fixed sample size of HUs be inflated beyond  $k$  so that with the anticipated nonresponse, the desired sample size of children is attained. The outcome of each sample HU is determined and noted by the interviewer to be one of the following categories: (a) vacant HU; (b) occupied HU, but interviewer is unable to enumerate eligible respondents due to occupants never being home, occupants' refusal to cooperate, or neighbors' refusal to give information; (c) HU contains no eligible respondents; and (d) HU contains one or more eligible respondents who are enumerated. Repeated callbacks can be used to minimize the number of selected HUs reported in category (b).

For each eligible subject in the sample, the interviewer may complete an interview, or the subject may refuse to be interviewed or be too sick or incapacitated to be interviewed. Again, the interviewer should note the outcome for each selected subject.

If necessary, imputation techniques (16) can be used prior to analysis to adjust for HU and eligible respondent nonresponse. Most techniques for HU nonresponse inflate the sampling weight of respondents so that the nonrespondents are represented by a subset of the respondents. A discussion of nonresponse and other nonsampling errors in household surveys in Africa is given by Kiregyera (19). EPI surveys,

especially in developing countries, generally have high response rates and imputation generally is not required.

#### 6. Calculate $k$ , the Number of HUs to Be Selected per Sample PSU, Based on Required Precision, Demography of Inference Population and Anticipated Nonresponse

Let  $r$  be the desired mean number of respondents (children) per PSU in the survey, as determined by precision requirements and anticipated design effect. Let  $IR$  be the anticipated response rate among enumerated children (or their informants). Let  $C/H$  be the mean number of age-eligible children per HU in the population of inference. Let  $HR$  be the HU response rate, that is, the proportion of selected occupied HUs for which all age-eligible children are enumerated. Let  $OR$  be the occupancy rate of HUs in the population of inference. Then,  $k$  is given by

$$k = r \times 1/IR \times H/C \times 1/HR \times 1/OR \quad 8$$

If one can assume an occupancy rate ( $OR$ ), HU response rate ( $HR$ ), and informant response rate ( $IR$ ) of 100% or 1.0, then equation 8 for  $k$  reduces to  $r \times H/C$ . That is, the number of sample HUs to select per PSU is equal to the desired mean number of children interviewed per PSU multiplied by the expected number of HUs required to find one age-eligible child.

#### 7. Do a Weighted Analysis Using a Calculator, Spreadsheet Software, or Sample Survey Software

The weighted estimator  $\hat{P}$  can be easily computed with a hand-held calculator or any software package that performs summations, such as standard spreadsheets, using the following formula:

$$\hat{P} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i m_i} \quad 9$$

where  $m_i$  is the number of sample children in PSU  $i$ ,  $y_i$  is the number of vaccinated children in PSU  $i$ ,  $w_i$  is the common sampling weight for all children in PSU  $i$ , and  $n$  is the number of sample PSUs in the survey.

Since several sample PSUs may be segmented and even subsegmented before selection of the sample HUs, the sample PSUs are likely to vary on the number of stages of sampling carried out therein. Thus, the survey is no longer a simple two-stage cluster survey, where the first stage of sampling is PSU selection and the second (final) stage is HU selection. Because of this, it is difficult or impossible to obtain an exact formula for the estimated variance of the estimator  $\hat{P}$  in equation 9. One standard sample survey practice in this situation, supported by sample survey theory, is to calculate the estimated variance of  $\hat{P}$  by considering only variation at the first sampling stage; that is, each child is considered to be a member of its sample PSU regardless

of how many stages of sampling within the PSU were used to identify the child. The standard approximate formula for the estimated variance of  $\hat{P}$  using this "ultimate cluster approach" (20) is

$$v(\hat{P}) \doteq [n/(\hat{M}^2(n-1))] \sum_{i=1}^n w_i^2 (y_i - \hat{P} m_i)^2 \quad 10$$

where  $\hat{M} = \sum_{i=1}^n w_i m_i$  is the estimated total number of age-eligible children in the population of inference.

Computation of  $v(\hat{P})$  may be easily accomplished using microcomputer spreadsheet software. Microcomputer software packages for analysis of "complex sample surveys" include the ultimate cluster approach as one option to approximate the estimated variance of survey point estimates; these packages also incorporate sampling weights into their calculations (21, 22).

Although there may be initial resistance among some surveyors to the assumed complexity of doing a weighted analysis with spreadsheet software or with a specialized survey software package, a weighted analysis usually will be necessary because (a) imperfect measures of size are used at all stages of sampling and (b) field work is simplified by having the same number of sample HUs selected in each sample PSU. The use of unweighted analyses when weighted analyses are required may yield serious bias in  $\hat{P}$  and almost certainly will underestimate the variability of  $\hat{P}$ . On the other hand, if the obtained sample does not deviate significantly from an equal probability sample of children (the selection probabilities can be calculated and compared in order to know this), then analyzing the data unweighted using formulas 3 and 4 will yield reasonable estimates of the population parameters of interest and their estimated standard errors.

If there are more than two stages of sampling, survey software packages may be able to incorporate these additional stages of sampling in estimating the variance. However, common survey practice in using these software packages is to ignore all stages of sampling within the PSU and use the ultimate cluster approach. Using the two survey software packages PC-CARP and PC-SUDAAN (21, 22), with the ultimate cluster approach and an unweighted analysis, yields calculations equivalent to equations 3 and 4. Using PC-CARP and PC-SUDAAN with the ultimate cluster approach with sampling weights yields calculations equivalent to equations 9 and 10.

#### 8. Extension to Stratified Surveys and to Larger Sample Sizes

The recommendations above also can be carried out in stratified cluster surveys, for example, where the inference population is stratified into rural and urban areas or into provinces or districts before first-stage sampling is done.

EPI surveyors usually select 30 PSUs per stratum because generally they desire to estimate coverage for each stratum, as well as estimate coverage for the entire inference population. Equations 9 and 10 can be modified for stratified cluster sampling as follows (21).

The estimator  $\hat{P}$  of  $P$  is given by

$$\hat{P} = \frac{\sum_{h=1}^L \sum_{i=1}^{n_h} w_{hi} y_{hi}}{\sum_{h=1}^L \sum_{i=1}^{n_h} w_{hi} m_{hi}} \quad 11$$

where  $y_{hi}$  is the number of vaccinated children in cluster  $i$  within stratum  $h$ ,  $m_{hi}$  is the number of sample children in cluster  $i$  within stratum  $h$ ,  $w_{hi}$  is the sampling weight for all children in cluster  $i$  within stratum  $h$ ,  $n_h$  is the number of sample clusters selected in stratum  $h$  (often  $n_h = 30$ ), and  $L$  is the total number of strata in the survey.

The estimated variance of  $\hat{P}$  is given by

$$v(\hat{P}) = \sum_{h=1}^L \frac{n_h}{\hat{M}^2(n_h - 1)} \sum_{i=1}^{n_h} \left[ w_{hi} (y_{hi} - \hat{P} m_{hi}) - \frac{(\hat{P}_h - \hat{P}) \hat{M}_h}{n_h} \right]^2 \quad 12$$

where

$$\begin{aligned} \hat{M} &= \sum_{h=1}^L \sum_{i=1}^{n_h} w_{hi} m_{hi} \\ \hat{P}_h &= \frac{\sum_{i=1}^{n_h} w_{hi} y_{hi}}{\sum_{i=1}^{n_h} w_{hi} m_{hi}} \\ \hat{M}_h &= \sum_{i=1}^{n_h} w_{hi} m_{hi} \end{aligned}$$

## INVESTIGATION OF THE EPI CLUSTER SURVEY METHODOLOGY

Considering the widespread use of the EPI survey methodology, it is surprising that few published materials exist on the performance of this technique (8). A computer simulation study by Lemeshow and associates (23) compared cluster sampling of HUs (adjacent HUs) to simple random sampling of HUs at the second stage of sampling. For cluster sampling, the computer simulation selected the starting HU with equal probability among all HUs in the sample PSU, a technique not used in the field by EPI surveyors. Hence, the simulated cluster sampling of this study does not adequately represent the EPI survey method, which uses interviewer judgment. The simulation study showed that simple random sampling of HUs was better than cluster sampling of HUs, producing smaller relative bias and smaller mean square error. This finding is not surprising and would be anticipated from sample survey theory. Simple random sampling of HUs was especially better than cluster sampling of HUs when "pockets" of vaccinated children existed in sample PSUs. When vaccinated children were distributed uniformly throughout the sample PSUs, then simple random sampling of HUs yielded equivalent results to cluster sampling of HUs. Although the simulation showed evidence

for the superiority of simple random sampling of HUs rather than cluster sampling of HUs, some degree of clustering of HUs is done in all sample surveys of resident populations for practical reasons of time and cost.

The simulation study by Lemeshow and associates (23) has been widely quoted as validating the EPI survey methodology (5, 8, 24, 25). However, this conclusion goes beyond the actual study since the authors did not simulate on a computer the judgmental, nonprobabilistic HU selection procedures used by interviewers in the field. The empirical results of the simulation study support sampling theory, which states that cluster sampling of HUs generally is less precise than simple random sampling of HUs, where both are equal probability samples of HUs and of children.

Henderson and Sundaresan (26) showed that most of 300 EPI cluster surveys that they reviewed did yield acceptable precision, that is, confidence intervals on vaccination coverage that had a half-width less than 0.10. This finding is not surprising since the sample size calculations for the  $30 \times 7$  survey assume the worst case scenario of  $P = 0.50$ , which yields the widest confidence intervals on  $P$ . The vaccination coverage surveys reviewed by Henderson and Sundaresan (26) estimated  $P$  to be anywhere between 5 and 95%. A major finding of their study, however, was that the assumed design effect of 2 has been a reasonable assumption for many EPI surveys. Although these authors did show that the half-width of the assumed 95% confidence interval does not exceed 0.10 in most surveys, the actual confidence level of the interval may be significantly below 95% if bias is present in the point estimates. Their review of the precision of EPI surveys did not address the accuracy and potential bias of such surveys, which is the focus of this article.

Bennett and coworkers (5) recently offered suggested modifications to the EPI survey, although their focus is not the same as that of this article (i.e., the accuracy of point estimates). Their recommendations, which are similar to ours, include the following: (a) use number of households as the measure of size; (b) select a fixed number of HUs per PSU; (c) spread out the selected HUs in a general-purpose health survey, for example, use simple random or systematic sampling; (d) use a weighted analysis in the presence of differing selection probabilities and/or nonresponse; and (e) use the ultimate cluster approach for variance estimation when there are more than two stages of sampling. Bennett and coworkers (5) did not emphasize the selection of a probability sample of HUs and the potential bias that can result from a nonprobability sample.

## CONCLUSION

Although the EPI survey method has been extremely useful in determining rough estimates of vaccination coverage in

developing countries, its continued widespread use for an increasing number of purposes requiring more accuracy and precision necessitates a careful consideration of the assumptions and potential biases of the methodology. The review herein suggests the following techniques to increase the accuracy of estimates derived from the basic EPI survey design: (a) Do counting, mapping, and listing for all or part of each sample PSU so that an equal probability sample of HUs can be chosen; (b) minimize nonresponse or adjust for it, and do not allow substitution; (c) calculate a selection probability and sampling weight for each child, and assess the degree of deviation from an equal probability sample of children; and (d) if an unequal probability sample of children has been obtained, then perform a weighted analysis to obtain an approximately unbiased estimator of  $P$  and a more accurate estimated standard error.

The suggested modifications of mapping the PSU and listing and counting HUs obviously will not improve the accuracy of survey estimates if the mapping, listing, and counting procedures are not implemented correctly and accurately. Consequently, it is necessary to train EPI survey field personnel in these tasks. To our knowledge to date, these suggested modifications to the EPI survey technique have been used only under the supervision of a survey statistician.

Even when survey field personnel are adequately trained for these activities, however, it certainly is possible to miss HUs in counting and listing procedures. We believe that the potential bias or error introduced by possibly missing HUs in counting and listing procedures is far less than the potential bias or error from not counting and listing at all in the current EPI survey design. Furthermore, if desired, standard survey field procedures can be used to correct for possibly missing HUs in counting and listing, for example, the open interval technique.

Note that these suggested modifications to the EPI survey methodology do not address the fact that the EPI survey, as well as virtually all population-based health surveys, omits the homeless population from the sampling frames. Omission of this population generally is a minor or nonexistent problem in standard applications of the EPI survey methodology since the subjects for many surveys are infants and young children. Also, homelessness has not emerged as a major problem in most developing countries where a majority of the population is rural. Of course, special situations such as war or natural disasters may force the EPI survey researcher to adjust the design for a substantial homeless population. Recent discussion on sampling the homeless population can be found in articles by Sudman and colleagues (27), Burnam and Koegel (28) and Pilivin and coauthors (29).

## APPENDIX I. CALCULATION OF SELECTION PROBABILITIES AND SAMPLING WEIGHTS

The following two examples illustrate the calculation of selection probabilities and sampling weights for children in the adaptation of the EPI survey we have suggested.

### Example 1: No Segmenting or Subsegmenting

The PSUs are census enumeration areas, estimated total population is the measure of size, and 30 PSUs are selected using PPES sampling. If the fifth PSU selected for the sample had an estimated population of  $T_5'$ , its probability of selection is  $30 \times (T_5'/T')$ , where  $T'$  is the estimated total population of all PSUs in the population of inference. The PSU is small enough to list and map in its entirety, yielding  $H_5$  listed HUs. One HU is selected at random from the list, and a cluster sample of  $k$  HUs is selected by including the  $(k - 1)$  HUs listed below this starting HU. The second-stage selection probability for each HU in PSU 5 is  $(k/H_5)$ . In each of the  $k$  HUs selected, all children of a specified age range are selected. Therefore, the selection probability for each sample child in PSU 5 is

$$\Pr(\text{select child } j \text{ in PSU } 5) = \frac{13}{30 \times (T_5'/T) \times (k/H_5) \times 1}$$

Hence, the sampling weight for each child in PSU 5 is  $(1/30) \times (T'/T_5') \times (H_5/k) \times 1$ .

### Example 2: Segmenting and Subsegmenting

As a second example, let the 17th selected PSU, with an estimated total population of  $T_{17}'$ , be too spread out to conveniently map and list in its entirety. A local official or other knowledgeable person describes physical attributes in the PSU such as roads, fields, or sections within a village, allowing the team leader to create a map of the PSU with these attributes as segment boundaries. The official confirms the segment boundaries and estimates the number of HUs (or persons if HUs cannot be estimated) in each segment. At the second stage of sampling, one of the segments in PSU 17 is selected using probability proportional to estimated size. For example, if PSU 17 is divided into six segments each of size  $S_\ell$  ( $\ell = 1, 2, \dots, 6$ ) and if segment 2 is selected, then the second-stage selection probability for segment 2 is  $(S_2/S)$ , where  $S$  is the total estimated size of the PSU, that is,

$$S = \sum_{\ell=1}^6 S_\ell.$$

The field team travels to segment 2 within PSU 17, where another quick inspection shows that segment 2 is spread out and sparsely populated. While the field leader could subsegment this segment, it is decided that some inconvenience in listing and mapping the entire segment is prefer-

able to subsegmenting and then risking the choice of a subsegment that has less than  $k$  HUs. The total number of HUs found in segment 2 is  $H_{17,2}$ . An equal probability cluster sample of  $k$  HUs is selected from the  $H_{17,2}$  HUs at the third stage of sampling, with all age-eligible children sampled in each of the  $k$  HUs selected. Thus, the probability of selection for each sample child in PSU 17 is

$$\Pr(\text{select child } j \text{ in PSU } 17) = 30 \times \frac{1}{(T_{17}'/T')} \times \frac{1}{(S_2/S)} \times \frac{1}{(k/H_{17,2})} \times 1 \quad 14$$

Hence, the sampling weight for each child in PSU 17 is  $(1/30) \times (T/T_{17}') \times (S/S_2) \times (H_{17,2}/k) \times 1$ .

In the two examples above, all sampled children within a given PSU have the same selection probability and hence, the same sampling weight. This is true for each sample PSU if the suggested modifications to the EPI methodology given above are followed. Note, however, that children in different PSUs generally will not have the same selection probability. If nonresponse adjustments are done to these sampling weights, children within the same PSU may have different sampling weights after the adjustment.

---

This research was funded by the African Child Survival Initiative/Combating Childhood Communicable Diseases Project (698-0421) of the United States Agency for International Development. Dr. Brogan's and Dr. Flagg's work was supported by an Intergovernmental Personnel Act agreement between Emory University and the Technical Support Division, International Health Program Office, Centers for Disease Control.

## REFERENCES

1. Information System. Geneva: World Health Organization, Expanded Programme on Immunization; April 1992. WHO/EPI/CEIS/92.1:68.
2. World Summit for Children. World Declaration on the Survival, Protection and Development of Children and Plan of Action for Implementing the World Declaration on the Survival, Protection and Development of Children in the 1990s. New York: United Nations; September 30, 1990:22.
3. Information System. Geneva: World Health Organization, Expanded Programme on Immunization; September 1992. WHO/EPI/CEIS/92.2:Tables 1.3.1-1.3.6.
4. Cochran WG. Sampling Techniques. 3rd ed. New York: Wiley; 1978:12-15.
5. Bennett S, Woods T, Liyanage WM, Smith DL. A simplified general method for cluster-sample surveys of health in developing countries, *World Health Stat Q*. 1991;44:98-106.
6. Stetler HC, Ayebowa A, Brink EW, Agle AN, Stachling NW. Nutritional status of preschool children in Togo, 1976-1977, *Bull WHO*. 1980;58:889-895.
7. Galazka A, Stroh G. Guidelines on the Community-Based Survey on Neonatal Tetanus Mortality. Geneva: World Health Organization; 1986. WHO/EPI/GEN/86/8.
8. Anker M. Epidemiological and statistical methods for rapid health assessment introduction, *World Health Stat Q*. 1991;44:98-106.
9. Serfling RE, Sherman IL. Attribute Sampling Methods for Local Health Departments with Special Reference to Immunization Surveys. Atlanta, GA: U.S. Centers for Disease Control, Epidemiology Program Office; 1965. DHEW publication no. (PHS)1230.
10. The EPI Coverage Survey. Geneva: World Health Organization, Expanded Programme on Immunization; October 1991. WHO/EPI/MLM/91.10.
11. Lemeshow S, Robinson D. Surveys to measure programme coverage and impact: A review of the methodology used by the Expanded Programme on Immunization, *World Health Stat Q*. 1985;38:65-75.
12. Lemeshow S, Stroh G. Sampling Techniques for Evaluating Health Parameters in Developing Countries. Washington, DC: National Research Council, Office of International Affairs, Board on Science and Technology for International Development, National Academy Press; 1988.
13. Household Survey Manual. Diarrhoea Morbidity, Mortality, and Treatment Practices. Geneva: World Health Organization, Diarrhoeal Diseases Control Programme; 1986. WHO/CDD/SER/86.2.
14. Rothenberg RB, Lobanov A, Singh KB, Stroh G. Observations on the application of EPI cluster survey methods for estimating disease incidence, *Bull WHO*. 1985;63:93-99.
15. Lee ES, Forthofer RN, Lorimer RJ. Analyzing complex survey data. In: Sage University Paper Series on Quantitative Applications in the Social Sciences. Beverly Hills: Sage; 1989. Publication 07-071.
16. Oh HL, Scheuren FJ. Weighting Adjustment for Unit Nonresponse. In: Madow WG, et al, eds. Incomplete Data in Sample Surveys. v. 2. New York: Academic; 1983:143-182.
17. Kish L. Survey Sampling. New York: Wiley; 1965:148-181.
18. Sunter A. On finding and delineating area sample units in the field, *J Official Stat*. 1989;5:171-182.
19. Kiregyera B. Types and some causes of nonsampling errors in household surveys in Africa, *J Official Stat*. 1987;3:349-358.
20. Frankel M. Sample theory. In: Rossi PH, et al, eds. Handbook of Survey Research. San Diego: Academic; 1983.
21. Fuller W. PC CARP. Ames, IA: Statistical Research Laboratory, Iowa State University; 1986.
22. Shah BV, Barnwell BG, Hunt PN, LaVange LM. SUDAAN User's Manual: Professional Software for Surey Data Analysis for Multi-stage Sample Designs. Release 6.0. Research Triangle Park, NC: Research Triangle Institute; 1992.
23. Lemeshow S, Tserkovnyi AG, Tulloch JL, Dowd JE, Lwanga SK, Keja J. A computer simulation of the EPI survey strategy, *Int J Epidemiol*. 1985;14:473-481.
24. Frerichs RR. Simple analytic procedures for rapid microcomputer-assisted cluster surveys in developing countries, *Public Health Rep*. 1984;104:24-35.
25. Frerichs RR, Tar Tar K. Computer-assisted rapid surveys in developing countries, *Public Health Rep*. 1989;104:14-23.
26. Henderson RH, Sundaresan T. Cluster sampling to assess immunization coverage: A review of experience with a simplified sampling method, *Bull WHO*. 1982;60:253-260.
27. Sudman S, Sirken MG, Cowan CD. Sampling rare and elusive populations, *Science*. 1988;240:991-996.
28. Burnam MA, Koegel P. Methodology for obtaining a representative sample of homeless persons: The Los Angeles skid row study, *Eval Rev*. 1988;12:117-152.
29. Piliavin I, Westerfelt H, Elliott E. Estimating mental illness among the homeless: The effects of choice-based sampling, *Soc Probl*. 1989;36:525-531.