Survey Methodology, June 1989
Vol. 15, No. 1, pp. 71–81
Statistics Canada

Quality Assurance Sampling for Evaluating Health Parameters in Developing Countries

STANLEY LEMESHOW and GEORGE STROH, JR.[1]

ABSTRACT

A typical goal of health workers in the developing world is to ascertain whether or not a population meets certain standards, such as the proportion vaccinated against a certain disease. Because populations tend to be large, and resources and time available for studies limited, it is usually necessary to select a sample from the population and then make estimates regarding the entire population. Depending upon the proportion of the sample individuals who were not vaccinated, a decision will be made as to whether the coverage is adequate or whether additional efforts must be initiated to improve coverage in the population. Several sampling methods are currently in use. Among these is a modified method of cluster sampling recommended by the Expanded Programme on Immunization (EPI) of the World Health Organization. More recently, quality assurance sampling (QAS), a method commonly used for inspecting manufactured products, has been proposed as a potentially useful method for continually monitoring health service programs. In this paper, the QAS method is described and an example of how this type of sampling might be used is provided.

KEY WORDS: Lot sampling; Quality assurance; Acceptance sampling; Vaccination coverage.

1. INTRODUCTION

One of the problems continually confronting managers of health service programs is the identification and application of cost-effective and practical methods to monitor and evaluate operations. In developing countries the solution to such problems is usually complicated because records are often poorly maintained, reports from dispersed health facilities are usually received late or not submitted at all, and accurate target population sizes are not available. Consequently, community-based surveys are often the only means to obtain reliable numerator (*i.e.*, number of individuals with a characteristic) and denominator (*i.e.*, number of individuals studied) data. However, such surveys can be difficult to organize and implement and are often too costly to be used to monitor program operations.

Perhaps the best example of a program in which community-based surveys have been routinely used to collect information is the Expanded Programme on Immunization (EPI) of the World Health Organization (WHO) (see Henderson and Sundaresan 1982). The EPI, from its inception, has employed a cluster sampling method designed to measure immunization coverage in young children (see Serfling and Sherman 1975 and Henderson *et al.* 1973). The particular survey methodology was kept as simple in concept and application as possible to allow program managers and supervisors, often with minimal background in sampling techniques, to organize and implement the surveys (see WHO 1979). These surveys, which have been termed "30 by 7" surveys, typically involve 30 clusters and 7 individuals studied per cluster. Indeed, the strength of the EPI survey method lies in the simplicity of the design,

[1] Stanley Lemeshow, Ph.D., is a Professor of Biostatistics and Chair of Biostatistics/Epidemiology, Division of Public Health, University of Massachusetts, Amherst, MA. George Stroh, Jr., MPH, Centers for Disease Control, Atlanta,GA.

the standardized rules for implementation, and the uncomplicated procedure for compiling and interpreting results. Discussion and criticisms of the method on theoretical grounds are available elsewhere (Lemeshow *et al.* 1985 and Lemeshow and Robinson 1985).

Recently, EPI officials have recognized several practical limitations of the survey methodology. The first concern is that the results obtained with the survey method are relatively imprecise – estimates of coverage obtained can only be expected to be within 10 percentage points of the actual level of coverage in the population sampled. In developing countries where high levels of coverage have been attained, the method is too imprecise to identify significant changes between sequential surveys, or between different strata of a population being evaluated.

The second concern about the use of the EPI surveys is that, even though they are relatively easy to implement, they are still too great an undertaking for most local managers to use to assess operations in their areas of responsibility. Consequently, it is still most common for an EPI survey to be done for the entire population of a country, or for population units of relatively large size (*e.g.*: millions). Although the results are useful for managers at higher program levels, local managers and supervisors are unable to use the results at their levels of responsibility.

EPI surveys usually measure the percentage of children in an age cohort (usually 12 to 23 months of age) that should have received the entire series of vaccines that are provided in the EPI. The third concern is that this results in measurement of operations that preceded the date of the survey by more than a year; operations may have changed considerably during that interval.

Finally, an additional objective of the EPI is to develop accurate record keeping that can be used to monitor and evaluate coverage – the surveys are the primary means of assessing the validity of records. However, with the current age groups surveyed, it is often difficult to identify the set of records that correspond to the period during which immunizations were given to the children surveyed.

In this paper, we present a method which has been proposed to continually monitor a health service program and can be used to assess whether operations are maintained at an acceptable, specified level. To do this, a particular type of stratified random sampling (Cochran 1977; Hansen *et al.* 1953; Kish 1965; Levy and Lemeshow 1980) is employed that uses very small samples obtained from operationally defined units of the population. Not only can this type of community-based sampling permit monitoring of operations within relatively small populations or small areas of operation, but the results will permit managers at virtually all levels to obtain estimates to continually evaluate program operations with sufficient precision. In areas where record systems have been developed that can be used to monitor program operations, the same sampling method can be used to validate the records and ensure that an accurate numerator and denominator are available from records. Once validated these records can then be relied upon as the major source of information for program monitoring and evaluation. The general term applied to this method of sampling, which we propose as a useful alternative to more traditional methods applied in the area of public health program evaluation, is Quality Assurance Sampling (QAS) – a term well known in the areas of engineering, manufacturing and business.

2. THE QAS METHOD

The origin of QAS is in sampling and inspecting manufactured products (Dodge and Romig (1959)) where it was developed to keep labor and other sampling costs at minimal levels. One type of QAS sampling, Lot Quality Acceptance Sampling (LQAS) is identical to stratified

sampling, but the samples are too small to provide what are usually considered acceptably narrow confidence intervals for estimates for a specific stratum (usually called a "batch" or "lot" in industry). Rather, a decision is made about the quality of a particular batch or lot based on the probability that the number of defective items in the sample is less than or equal to a specified number. The results of the samples taken from all the mutually exclusive and exhaustive batches can be combined to provide a precise overall estimate of the average quality of the total product.

The strategy and goals of QAS in the health field would be similar to those in the manufacturing field. The purchaser of goods does not want to accept a batch with more than a certain percentage (P_1) defective whereas the manufacturer wants to continually monitor production to identify products with more than an expected percentage (P_2) of defectives. It is not unusual for P_1 and P_2 to be unequal. It is not difficult to see the similarities between the objectives of a manufacturer and a health manager or supervisor. The latter "produces" immunized children rather than a manufactured item.

Generally, a lot is an "operationally useful" unit. For example, in an industrial application, if there were several machines producing the same part and three operators assigned to each machine, then "lots" could be chosen that are produced by the same machine – particularly if any variation in the parts produced is most likely to be due to machine drift as opposed to operator input.

For public health work, a manager might define "lots" as recipients of services from a single operational unit – such as a health post (HP) immunization team – over a specified period of time. The amount of time between sampling could coincide with the interval between "high incidence" seasons for immunizeable diseases, but would more likely be related to the amount of time and cost associated with the sampling than any other single consideration.

In public health work a serious error would be made if the population were judged to be adequately covered ("accept the lot") when, in fact, it is not. In order to control for this possibility, we design the procedure as a one-sided test.

The null hypothesis, illustrated at the 50% level, is

$$H_o\text{: } P \geq P_o \text{ (}i.e.\text{, proportion of unvaccinated children } \geq 0.50\text{)}$$

versus

$$H_a\text{: } P < P_o \text{ (}i.e.\text{, proportion of unvaccinated children } < 0.50\text{).}$$

The four-celled table presented in Figure 1 describes the consequences of the testing procedure. Because the test is set up as one-sided, and because we assume the population is not adequately covered unless we reject H_o, the type I error, *i.e.*, accepting the lot when it is defective (false negative), is the most serious error. That is, if (using the example of immunization) a population (lot) of children is thought to have an acceptable proportion immunized when, in fact, it does not, the larger number of susceptibles in the population increases the risk of transmission of the disease. Hence, we consider the "cost" of declaring that the population is adequately vaccinated, when it is not, to be high. On the other hand, the type II error, rejection of an acceptable lot, is not as serious since the result of a false-positive decision would be to concentrate efforts on an already adequately vaccinated population.

The fundamental problem in LQAS sampling, is not so much one of simply determining sample size, but of choosing an appropriate balance between sample size and critical region. In all cases, the computation of β will depend upon the actual value of P when it is assumed to be different from P_o.

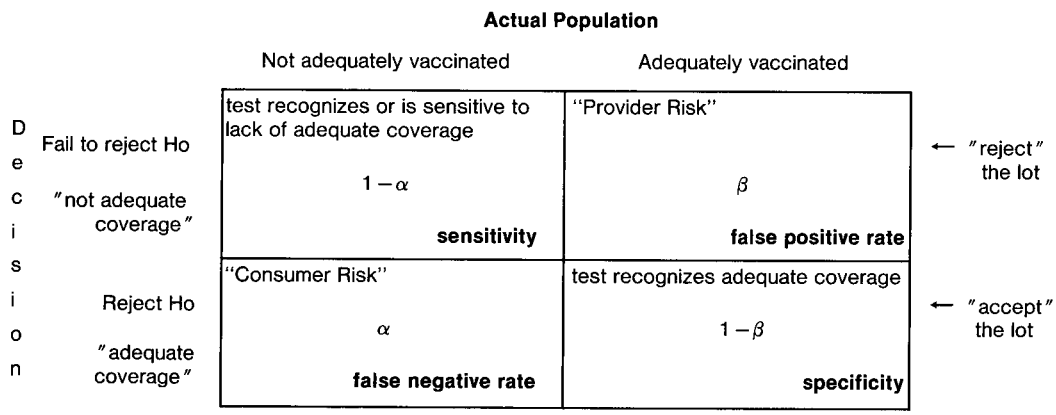


Figure 1. Consequences of Hypothesis Testing in LQAS Procedure

In practice, initially a minimal level for delivery of a service would be defined on the basis of the probable distribution of service levels across lots as well as in terms of practicality (*i.e.*, a level that could be achieved). Once this level is defined, sample size options are considered relative to the number of lots that would be misclassified with stated type I and type II errors. If the sample size were too large to be practical, there would be several options including: retaining the sampling scheme, but lengthening the time interval between sampling; choosing another critical level that would allow use of a smaller sample size; choosing another QAS sampling scheme (such as double sampling or sequential sampling) that would meet the objectives of classifying the lots and still be operationally feasible; and abandoning a QAS scheme.

One means of computing probabilities and determining necessary sample sizes can be accomplished using the binomial distribution. We will assume, as is usually the case, that N is very large relative to n; with large N, the Poisson can be practically substituted for the binomial. However, if it happens that N is not large relative to n, then the hypergeometric distribution can be used as described in Brownlee (1965) (Sec. 3.15). Letting p denote the probability of observing the characteristic, then the chance of observing exactly d individuals with the characteristic in a sample of size n is given by

$$p(d) = \binom{n}{d} P^d (1-P)^{n-d}.$$

Suppose we decide that 7 is the sample size we wish to use. The rejection region for the test states that we should reject H_o (and "accept the lot" as adequately vaccinated) if $d \leq d^*$. To determine the value of d^* such that $Pr(d \leq d^*) = \alpha$, we must compute $Pr(d \leq d^*)$ for a number of values of d^*. Clearly if we decide to use $d^* = 1$ then $Pr(d \leq d^*)$ would equal 0.0625 and the power of the test, if 70% of the population is actually unvaccinated, would equal 0.0038.

Results of a particular choice of n and d^* may be graphed as an **operating characteristic (OC) curve** where the variable on the horizontal axis is the proportion, P, in the population who have not been vaccinated. The vertical axis presents the probability of rejecting the null hypothesis H_o: $P = P_o$ and concluding that the vaccination coverage in the population is adequate. Each combination of n and d^* will generate a unique curve. Figure 2 presents a typical OC curve for $n = 7$, $d^* = 1$.

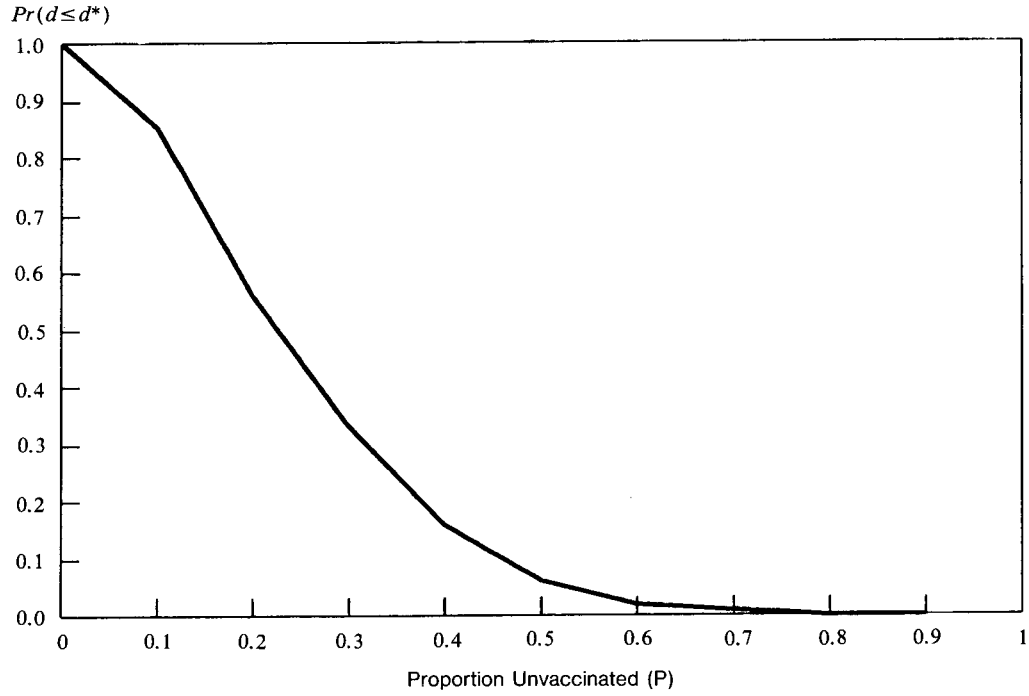


Figure 2: Operating Characteristic Curve for $n=7$ and $d^*=1$

The investigator will usually choose the value of d^* which yields a type I error less than α. Sometimes this strategy results in an extremely conservative test. For example, with $n = 7$, $d^* = 0$ and $P_o = 0.5$, α would equal 0.0078. Here the use of $d^* = 1$ with $\alpha = 0.0625$ as in Figure 2 might be justified. Table 1 presents values of d^* for small n (≤ 20) such that α will not exceed the stated type I error probability (0.01, 0.05 or 0.10) for various combinations of n and P_o. Details for the construction of this table are presented elsewhere (Dodge and Romig 1959).

The choice of the sampling scheme comes down to one of combining the desired power, $1 - \beta$, with the desired α level. Rather than providing curves which are difficult to read precisely, we developed Table 2 which presents values of (n,d^*) pairs for $\alpha = 0.05, \beta = 0.20$, and selected values of P under the null hypothesis (P_o) and P under the alternative hypothesis (P_a). In this table, (n,d^*) are chosen so that $Pr(d \leq d^* \mid n,P_o) \leq \alpha$ and $Pr(d \leq d^* + 1 \mid n,P_o) > \alpha$. More details are provided elsewhere (Lemeshow *et al.* 1987).

This table clearly shows the trade off one must make between power and sample size in LQAS surveys. For instance, it is essentially impossible to have $\alpha=0.05$, $\beta=0.20$ and use $n=5$ unless P_a under the alternative was actually close to 0. Hence investigators with limited resources must be ready to compromise on the value of β or the difference between P_o and P_a.

The method of quality assurance sampling described to this point is known as "single sampling" since only one sample is taken before a decision is reached regarding the disposition of the lot. A modification of this LQAS procedure, which may be useful under certain field conditions, incorporates a "double sampling" strategy. With this method, a sample is first selected of size n_1. If this sample fails, a second sample of size n_2 may be selected. This requires the specification of two acceptance numbers. The first, d_1, applies to the observed number of defectives in the first sample alone and the second, d_2, applies to the total number of defectives in the first and second samples combined. In practice, the principal advantage

Table 1

Values of d* for Combinations of P_o and n to Achieve alpha ≤ 0.01, 0.05, or 0.10

	P_o, alpha ≤ 0.01					P_o, alpha ≤ 0.05					P_o, LPH ≤ 0.10				
n	0.50	0.60	0.70	0.80	0.90	0.50	0.60	0.70	0.80	0.90	0.50	0.60	0.70	0.80	0.90
5	×	×	0	1	2	0	0	1	1	2	0	1	1	2	3
6	×	0	0	1	2	0	1	1	2	3	0	1	2	3	3
7	0	0	1	2	3	0	1	2	3	4	1	2	2	3	4
8	0	1	1	2	4	1	2	2	3	5	1	2	3	4	5
9	0	1	2	3	5	1	2	3	3	5	2	3	4	5	6
10	0	1	2	4	5	1	2	4	5	6	2	3	4	5	6
11	1	2	3	4	6	2	3	4	5	7	2	4	5	6	8
12	1	2	4	5	7	2	3	5	6	8	3	4	5	7	8
13	1	3	4	6	8	3	4	5	7	9	3	5	6	8	9
14	2	3	5	6	9	3	4	6	8	10	4	5	7	8	10
15	2	4	5	7	9	3	5	7	8	10	4	6	7	9	11
16	2	4	6	8	10	4	5	7	9	11	4	6	8	10	12
17	3	4	6	8	11	4	6	8	10	12	5	7	8	10	13
18	3	5	7	9	12	5	6	8	10	13	5	7	9	11	14
19	4	5	7	10	13	5	7	9	11	14	6	8	10	12	14
20	4	6	8	11	13	5	7	10	12	15	6	8	10	13	15

× No test for this sample size.

Table 2

Sample Size and Decision Rule for LQAS, Alpha = 0.05, Beta = 0.20, One-sided Test

P_a	P_o				
	0.50	0.60	0.70	0.80	0.90
	n , d*	n , d*	n , d*	n , d*	n , d*
0.05	5 , 0	×	×	×	×
0.10	8 , 1	5 , 0	×	×	×
0.15	11 , 2	7 , 1	×	×	×
0.20	15 , 3	9 , 2	5 , 1	×	×
0.25	23 , 7	12 , 3	7 , 2	×	×
0.30	37 , 13	16 , 5	9 , 3	5 , 1	×
0.35	67 , 26	24 , 10	11 , 4	6 , 2	×
0.40	153 , 66	38 , 17	16 , 7	8 , 3	×
0.45	617 , 288	67 , 33	23 , 12	10 , 5	5 , 2
0.50		151 , 80	35 , 20	13 , 7	6 , 3
0.55		601 , 340	62 , 37	19 , 11	7 , 4
0.60			137 , 86	29 , 19	10 , 6
0.65			535 , 356	50 , 35	13 , 9
0.70				109 , 80	20 , 15
0.75				419 , 321	33 , 27
0.80					69 , 58
0.85					253 , 219

× Sample size less than 5.

of double sampling is that, if the defective rate is relatively low, it may be possible to study fewer subjects than with single sampling since n_1 is typically less than the n required in single sampling. However, if it becomes necessary to go to the second sample in many of the lots, the procedure may require a larger overall sample size. In most cases, the total sample size would be less than $n_1 + n_2$ since sampling stops as soon as the critical value, d_2, is exceeded in the second sample. (The first sample is always completed to provide the information to be combined and used to compute the overall proportion acceptable in the population). Details for this procedure are presented elsewhere (Dodge and Romig 1959) and an example will be presented in Section IV.

3. ESTIMATING THE OVERALL POPULATION PROPORTION WITH QAS SAMPLING

In addition to the binary decision to "accept" or "reject" the lot, the simple random samples within each HP may be considered a stratified sample and an overall population estimate constructed.

For example, suppose 294 HP's of known population size were sampled selecting 7 children from each. Using standard stratified sampling formulae, estimates may be obtained for P, $\text{Var}(\hat{P})$, and an appropriate confidence interval may be constructed. LQAS resembles stratified sampling in that it requires that an accurate sampling frame be established in each lot and that a **simple random sample** be selected from each of these lots. However, it does not provide more information than conventional stratified random sampling since confidence intervals could be established for each stratum (or lot) and decisions could be based on values covered by each such interval (if sample sizes were made large enough to provide useful confidence intervals).

Although the n for each stratum in LQAS are too small to provide useful confidence intervals for estimates for each stratum, an appropriately designed LQAS scheme may provide a means for continually testing strata and classifying them as "acceptable" or "unacceptable" in terms of a particular outcome. This results from the fact that LQAS sample sizes are relatively small, increasing the likelihood that sampling can be done more frequently. Among its benefits, the rules of LQAS sampling are simple to follow, requiring minimal retraining of the surveyor/classifier. Lastly, since LQAS samples are, in fact, stratified random samples, the results for strata can be combined to provide adequately precise estimates for groups of strata, such as for districts, regions, or a nation as a whole.

The potential benefits of use of an LQAS scheme must be weighed against the loss of precision expected with the small samples taken in each stratum. Perhaps the best way for the reader to judge whether LQAS might be useful is an example in which a conventional stratified random sample survey approach is compared with an LQAS scheme.

4. AN EXAMPLE OF THE APPLICATION OF QAS

The example is set in circumstances similar to those in Costa Rica, and is applied to immunization coverage of children which is provided by 294 HP that cover the population of the country. The manager of the EPI would like to know the percentage of children, 12-23 months of age that received all of the immunizations that should have been given during their first year of life. Based on the immunizations that have been reported by staff, the manager thinks that the coverage level for the nation is about 60%, but the coverage that has been reported by the 294 individual HP varies from 20% to 100%; it is thought that the distribution

of coverage rates is uniform across the range. The EPI manager suspects that the estimates of coverage provided on reports may not be completely accurate because of numerator and denominator errors. As a result, it is decided that a survey of HP areas should be made in order to obtain estimates of coverage for each of the 294 areas since it would be important to be able to concentrate supervision on those HPs that have "low" coverage.

The first plan for the survey that the EPI manager evaluates is a "conventional" stratified random sampling scheme. Coverage estimates are required for each of the 294 HP, and each estimate should have confidence bounds no larger than an absolute 10%, with $a = 0.05$. Since the average HP population is approximately 2500, and since it can be estimated that 3.5% of the population are children between the ages of 12 and 23 months, it is estimated that the number of children available for sampling in each HP will be approximately $2500 \times 0.035 = 88$. The formula for sample size determination which incorporates a finite population correction is given by Cochran (1977, p.75) and results in $n = 47$.

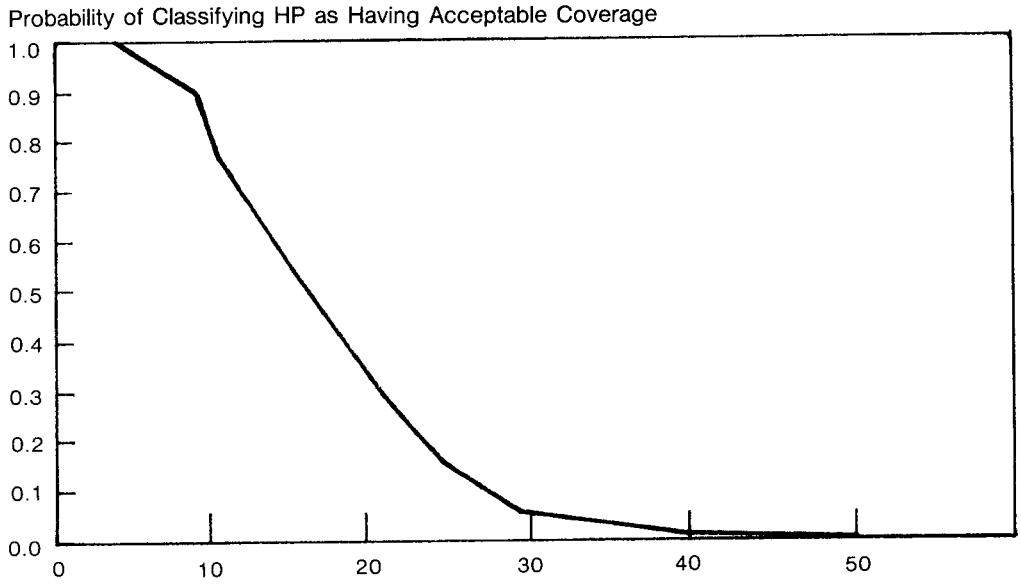
Thus, in each of the 294 HP areas, 47 (53%) of the 88 children between the ages of 12 and 23 months will be surveyed. In the entire country, 13,818 children in this age group will be surveyed. For the national estimate of coverage, P can be estimated to within 0.5% (assuming the worst level of coverage for precision (50%) and little variation in HP populations).

The manager then considers a QAS scheme. It is decided that any HP that has a coverage level of 70% or lower is performing poorly, and should be identified for increased supervision. The manager wants to be able to identify a HP with coverage of 70% with a probability of about 0.95, and HPs with lower levels of coverage with even higher probability. Several QAS schemes are considered and a double sampling scheme is proposed.

The particular double sampling scheme proposed can be denoted as $n_1{:}d_1 = 10{:}0$ and $n_2{:}d_2 = 14{:}3$. This means that in each HP area an initial sample of 10 children will be surveyed for their immunization status. Regardless of how many children are found unimmunized, all 10 will be surveyed. The number of children found unimmunized among each HP sample of 10 children will be used to compute estimates for combined areas and ultimately for the national estimate of coverage. If upon completion of a survey of the first sample of 10 children, none are found unimmunized, the HP will be categorized as having "acceptable" coverage. If 4 or more children are found unimmunized, the HP will be classified as having "unacceptable" coverage. In either scenario, no further sampling is required in the HP area. However, if upon completing the initial survey, 1, 2, or 3 children are found unimmunized, a second sample of 14 additional children is drawn. During the survey of the second sample, whenever a total of 4 unimmunized children is reached (including those from the first sample of 10) the survey is stopped, and the HP area is classified as having "unacceptable" coverage. However, if upon completion of the second sample, a total of 3 or fewer unimmunized children have been found, the HP area is classified as "acceptable".

Figure 3 shows the operating characteristic curve for this particular sampling scheme. This curve allows one to predict what the probabilities are for correctly classifying HP areas on the basis of the level of coverage. We will assume that the distribution of the 294 HPs is uniform and that all HPs in each decile have a coverage that corresponds to the mid-point value for each decile. If the probabilities of accepting a HP as having acceptable coverage are read from the OC curve and are applied to the numbers of HPs in corresponding deciles, it is possible to predict the number of HPs that would be accepted and rejected as having acceptable levels of coverage. The results of this projection are shown in Table 3.

As can be quickly computed from the expected results shown in the table, greater than 99% (183 of 184) of the HPs that had coverage less than 70% would be "rejected" (*i.e.*, they are classified as having an unacceptable level of coverage). Of the 110 HPs that had coverage above



Adapted from Dodge and Rowing (1959) Appendix 2: OC Curves For All Double Sampling Plans — (N=51–100)

Figure 3: Operating Characteristic Curve for Double Sampling Scheme with $n_1{:}d_1 = 10{:}0$ and $n_2{:}d_2 = 14{:}3$

Table 3

Expected Classification of 294 HP with Use of Double Sampling Scheme $n_1{:}d_1 = 10{:}0$ and $n_2{:}d_2 = 14{:}3$

Percentage Coverage in HP Area	Number of HP	Number of HP Classified as:	
		>70% Coverage	≤70% Coverage
20- 30%	36	0	36
31- 40%	37	0	37
41- 50%	37	0	37
51- 60%	37	0	37
61- 70%	37	1	36
71- 80%	37	7	30
81- 90%	37	21	16
91-100%	36	34	2
Total	294	63	231

Number of HP with Coverage ≤ 70% = 184.
Number Correctly Classified = 183 (99%).
Number of HP with Coverage > 70% = 110.
Number Correctly Classified = 62 (56%).

70%, 62 (56%) would be accepted (*i.e.*, they are classified correctly as having an acceptable level of coverage). Although a substantial portion of the HPs (48 of 110) that had coverage higher than 70% would be incorrectly classified as having "low" coverage, it should be noted that 63% (30 HPs) of them had coverage that was in the "marginal" range (*i.e.*, coverage levels in the 70-80% range).

Based on the initial samples of 10 children completed for each of the 294 HPs, a national estimate can be computed as with any stratified random sample. Using the same assumptions as were made for the "conventional" plan, the 95% CI for the national estimate of coverage from the QAS scheme would estimate P to within 1.8%, a level of precision that is adequate for the purpose of the EPI manager.

It should also be noted that the total number of children that would be surveyed in each HP area would vary between 10 and 24. In fact, with the particular distribution of coverage levels assumed in this example, the majority of HPs would be classified on the basis of the initial sample of 10 children (*i.e.*, of the 184 HP with $<70\%$ coverage, about 98% would be classified as unacceptable from the initial $n_1{:}d_1 = 10{:}0$ sample). Of the minority of HPs which were not classifiable on the basis of the initial sample, few would require surveying all 14 children in n_2. Thus, the "average" number of children sampled across all 294 HP would be substantially less than $n_1 + n_2$.

In conclusion, LQAS may have useful application in certain settings in which conventional stratified random sampling – requiring sufficient sized samples from each stratum to produce useful confidence intervals for the estimates obtained – is too costly and/or time consuming. LQAS is, in fact, nothing more than another way of interpreting data obtained with a stratified random sample with samples too small to provide meaningful confidence intervals. Because it may be possible to do such small sampling more frequently, the potential exists for establishing a system for continual monitoring of an activity, perhaps using staff that with minimal training could include monitoring activity with other field duties. One further advantage of the more frequent sampling could be that rather than concentrate on an age cohort that has passed through the full period of exposure to all immunizations, managers could instruct surveyors to collect information on children in the process of being immunized – *i.e.*, determine whether children have received the immunizations that are appropriate for their age. This would provide a means of obtaining information on more current activity, and afford an opportunity to intervene in a more timely manner to improve coverage.

Although confidence intervals will always provide much more information than a simple binary decision, the sample sizes required to obtain any useful level of precision on estimates for relatively small strata may be prohibitive. In such instances, an appropriate QAS scheme may be an alternative approach worthy of consideration.

REFERENCES

BROWNLEE, K.A. (1965). *Statistical Theory and Methodology in Science and Engineering*, (2nd ed.). New York: John Wiley and Sons.

COCHRAN, W.G. (1977). *Sampling Techniques*, (3rd. ed.). New York: John Wiley and Sons.

DODGE, H.F., and ROMIG, H.G. (1959). *Sampling Inspection Tables*, (2nd. ed.). New York: John Wiley and Sons.

HANSEN, M.H., HURWITZ, W.N., and MADOW, W.G. (1953). *Sample Survey Methods and Theory*, Vols. 1 and 2. New York: John Wiley and Sons.

HENDERSON, R.H., *et al.* (1973). 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.

HENDERSON, R.H., and SUNDARESAN, T. (1982). Cluster Sampling to Assess Immunization Coverage: A Review of Experience with a Simplified Sampling Method. *Bulletin of the World Health Organization*, 60: 253-260.

KISH, L. (1965). *Survey Sampling*. New York: John Wiley and Sons.

LEMESHOW, S., *et al.* (1985). A Computer Simulation of the EPI Survey Strategy. *International Journal of Epidemiology*, 14, 3: 473-481.

LEMESHOW, S., and ROBINSON, D. (1985). Surveys To Measure Programme Coverage and Impact: A Review of the Methodology Used by the Expanded Programme on Immunization. *World Health Statistics Quarterly*, 38, 1.

LEMESHOW, S., HOSMER, D., and KLAR, J. (1987). *Sample Size Determination*. To be published by World Health Organization.

LEVY, P.S., and LEMESHOW, S. (1980). *Sampling for Health Professionals*. Lifetime Learning Publications, New York: Van Nostrand Reinhold.

SERFLING, R.E., and SHERMAN, I.L. (1975). Attribute Sampling Methods. Washington, D.C., US Department of Health and Human Services, Public Health Service, Publication No. 1230.

WORLD HEALTH ORGANIZATION. (1979). *Training for Mid-Level Managers. Evaluate Vaccination Coverage*. Expanded Programme on Immunization in Cooperation with US Department of Health and Human Services, Public Health Service, Center for Disease Control. Geneva: WHO.