Sampling Techniques for Evaluating Health Parameters in Developing Countries: A Working Paper (1988)

Pages
43

Size
7 x 10

ISBN
0309319757

Lemeshow, Stanley; George Stroh, Jr.; Board on Science and Technology for International Development; Office of International Affairs; National Research Council

Sampling Techniques for Evaluating Health Parameters in Developing Countries

A Working Paper Prepared by

Stanley Lemeshow, Ph.D.
University of Massachusetts, Amherst
George Stroh, Jr., MPH
Centers for Disease Control

for

Board on Science and Technology for International Development
Office of International Affairs
National Research Council
United States of America

NATIONAL ACADEMY PRESS
Washington, D.C. 1988

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competence and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Frank Press is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Samuel O. Thier is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Frank Press and Dr. Robert M. White are chairman and vice chairman, respectively, of the National Research Council.

The Board on Science and Technology for International Development (BOSTID) of the Office of International Affairs examines ways to apply science and technology in developing countries to problems of economic and social development through an integrated program of overseas activities, published reports of studies, advisory committees, conferences and seminars, and outreach activities. The BOSTID Research Program offers grants to developing country institutions for research in selected areas of science and technology important for economic development.

Funding for this report was provided by the Office of the Science Advisor, Agency for International Development, under Grant No. DAN 5538-G-SS-1023-00.

Copies are available from:
Board on Science and Technology
for International Development (BOSTID)
National Research Council
2101 Constitution Avenue, N.W.
Washington, D.C. 20418 USA

PREFACE

Rapid epidemiologic assessment (REA) is a relatively new concept in international development. It addresses the problem, keenly felt in the health ministries of the Third World, that the cost, time, and infrastructure required to carry out even a modest longitudinal study of a health problem in a less-developed country are usually not available. The consequence is that basic information on disease burden, efficiency and coverage of health services, and health-related statistics are not reliable, and decisions often must be made on the basis of guesswork or quick estimates of questionable epidemiologic value.

REA techniques are designed to be low cost, simple, or rapid, and to be within the resource capabilities of the developing countries. REA often involves small sample size techniques, sentinal indicators, case control methodologies, risk factor analysis, and use of uneducated community workers to collect health data. Precision is sacrificed in the interests of lower costs and simplicity, and for that reason it is doubly important to understand the limitations of REA techniques and to confine their use to situations where they have been explicitly verified. The development of REA techniques and the definition of their limitations and measurement of validity are the goals of the REA research program of the Board on Science and Technology for International Development (BOSTID), which is a unit of the National Research Council.

The authors of this monograph have contributed to the REA program in various ways: as proposal reviewers, participants in technical meetings, and volunteer consultants to developing country research grantees. Dr. Stanley Lemeshow, Professor and Chairman, Biostatistics/Epidemiology Program, School of Public Health, University of Massachusetts, Amherst, has developed computer simulation models for evaluating various sampling strategies for developing countries, and has worked as a consultant for the World Health Organization (WHO). Dr. George Stroh, Jr., Epidemiology Program Office, Division of Field Services, Center for Disease Control, has evaluated health care delivery systems in developing countries. Most notably, they have played a primary role in developing the Lot Quality Assurance Sampling (LQAS) method, which is almost a prototypic REA technique, illustrating by its ingenuity and versatility what REA is all about. LQAS has been the subject of research projects in Peru and Costa Rica under the BOSTID Research Program, and its successful application to evaluation of health services and measurement of infant malnutrition prevalence has been demonstrated.

This monograph is not intended to be a primer in the method, but rather to put LQAS in the context of other small sample size techniques and to illustrate its use and usefulness. Comments, suggestions, and reports of experience in the field are welcome.

Sampling Techniques for Evaluating Health Parameters in Developing Countries

Stanley Lemeshow, Ph.D., University of Massachusetts, Amherst, MA
George Stroh, Jr., MPH, Centers for Disease Control, Atlanta, GA

A typical goal of health workers employing one of the currently popular sampling procedures is to ascertain whether or not a population meets certain standards, such as the proportion vaccinated against a certain disease. After identifying the population of interest, the researcher would like to take a "representative" sample of subjects and, depending upon how many were not vaccinated, decide whether the coverage is adequate or whether additional efforts must be initiated to improve coverage in the population.

Because populations tend to be large and resources and time available for studies limited, it is usually not possible to study each person or family comprising a population. For this reason there is little choice but to select a sample from the population and, from it, make estimates regarding the entire population. In order for such estimates to be made, it is necessary that some scientifically valid sampling methodology be employed.

Several sampling methods are currently in use. These include simple random sampling, stratified random sampling, cluster sampling, and a modified method of cluster sampling recommended by the Expanded Programme on Immunization (EPI) of the World Health Organization (WHO). Recently, lot quality assurance sampling (LQAS), a type of stratified random sampling, has been proposed as a potentially useful method. In this paper each of these methods will be briefly reviewed and compared for use in health surveys in developing nations. LQAS will be described more fully, and an example of how this type of sampling might be used will be provided.

For illustrative purposes the Rural Health System presently intact in Costa Rica, which is the first community based health system to be established in Latin America, will be considered. The foundation of this system is 294 Health Posts (HPs) that provide the first level of care in the Costa Rican primary health care system. These HPs are distributed among five geographic regions that are responsible for 3,050 rural communities totalling more than 750,000 people. Of interest is the estimation of the level of vaccination coverage for the Costa Rican Rural Health System as a whole. Recognizing that some of the HPs may be functioning far more effectively than others, there is also interest in evaluating the performance of each HP for the purpose of identifying and correcting specific deficiencies. Each of the possible sampling strategies will be considered in turn. Computational formulae will be presented to illustrate the process of estimating the population proportion, **P**, for each of these sampling strategies. A more detailed discussion of strategies for sampling from human populations and derivation of formulae may be found in a number of sampling textbooks[1,2,3,4].

I. Simple Random Sampling

Consider the 750,000 individuals in Costa Rica to be the population of interest. The population size will be denoted **N**. The proportion of these individuals possessing some characteristic is denoted **P**, the mean level of some characteristic, X, over all N enumeration units is denoted μ, and the variance of the N values of X is denoted σ^2.

Because N may be very large or the time or budget available to carry out the survey very limited, a sample of size n of the original N individuals in the population must be selected. From the n selected individuals in the sample, the population proportion, mean and variance may be estimated by $\hat{P}$, $\bar{x}$, and s^2, respectively.

If this sample is selected at random from the population, these estimates will be "unbiased." This means that if many (e.g., "k") random samples were selected from this population, and if $\hat{P}$, $\bar{x}$, and s^2 were computed for each of these samples, the average of the k sample proportions would equal the population proportion, P, the average of the k sample means would equal μ, and the average of the k sample variances would be σ^2. Unbiasedness is a desirable statistical property since it assures that the sample values will, on average, be correct. The concept of unbiasedness relates to repeated sampling and the corresponding averaging process.

It must be stressed that an estimate computed from any one particular sample may be quite different from the population parameter. However, one can compute "confidence limits" about an estimate obtained from an unbiased sample, and thus, the degree of error due to chance can be controlled by design--mainly by the size of the sample taken.

In order to select a simple random sample it is necessary to

(1) Construct a list (or "frame") of the N enumeration units.
(2) Use a random process (such as a random numbers table) to generate n numbers between 1 and N.
(3) Identify the n individuals in the sample corresponding to the n numbers generated.

Note that there are $\binom{N}{n}$ possible samples which can be selected from this population [where $\binom{N}{n} = N!/n!(N-n)!$ and $a! = a \times (a-1) \times (a-2) \times \ldots \times 2 \times 1$]. For example, if N=25 and a sample size n=5 is to be selected, there are $\binom{25}{5}$=53,130 possible samples. A *simple random sample* may then be formally defined as a sample in which each of the $\binom{N}{n}$ possible samples has the same chance of being selected (i.e., $1+\binom{N}{n}$). Estimating the population proportion with simple randon sampling is easily accomplished as follows. The point estimate of the proportion is:

$$\hat{P} = \left[\sum_{i=1}^{n} y_i \right] / n$$

where $y_i = 1$ if the i^{th} individual has the characteristic of interest and $y_i=0$ otherwise. The variance of P is estimated as:

$$\hat{Var}(\hat{P}) = \frac{[N-n]}{N} \frac{[\hat{P}(1-\hat{P})]}{(n-1)}$$

A 100[(1-α/2)]% confidence interval for P is contructed as follows:

$$\hat{P} \pm z_{1-\alpha/2}\sqrt{\hat{Var}(\hat{P})}$$

where $z_{1-\alpha/2}$ is the upper $(1-\alpha/2)^{th}$ percentile of the standard normal distribution.

The advantages of simple random sampling may be stated as follows:

- It is simple to conceptualize.
- It provides the probabilistic foundation of much of statistical theory.
- It provides a baseline to which other methods can be compared.

However, because of its numerous disadvantages, simple random sampling is rarely employed in actual surveys of human populations and would not be practical for the Costa Rican survey. The disadvantages of simple random sampling are:

- All N individuals (or enumeration units) in the population must be identified and labelled prior to sampling. This process is potentially so expensive and time consuming that it becomes unrealistic to implement in practice.
- Individuals selected in the sample may be highly dispersed--visiting each of the sampled individuals may be a very time consuming and expensive process.
- Individuals representing certain subgroups in the population may, by chance, be totally overlooked in the sample.

Fortunately, other methods are available that may provide more precise estimates (i.e., narrower confidence intervals) for the same cost.

II. Stratified Random Sampling

If a simple random sample were selected from the N individuals comprising the population, the possibility exists that, by chance alone, individuals in certain communities or in the service region of certain HPs would be either totally missed, oversampled, or undersampled.

Stratified random sampling is the process of creating mutually exclusive and exhaustive strata, selecting random samples from each of the strata, and finally combining these into a single sample to estimate the population parameters. In stratified random sampling, each individual in the population would be associated with exactly one of the established strata. In order to obtain the highest precision, individuals within the strata should be as homogeneous as possible, while stratum-to-stratum variation should be relatively large.

A *stratum* may be defined as a nonoverlapping subpopulation of the original population. Strata are defined on the basis of some known characteristic about the population that is believed to be related to the variable of interest. For example, in the Costa Rican project, HPs provide natural strata of individuals who may be similar with respect to available health services.

Notation used for this sampling design may be expressed as follows. Each individual in the population is categorized into one of L mutually exclusive strata. Let N_h denote the population size in stratum h. In stratified sampling a random sample is selected from each stratum. Let n_h denote the size of the sample drawn from stratum h. Using the

same computational process followed with simple random sampling, $\hat{P}_h$ and $\hat{Var}(\hat{P}_h)$ are calculated for the h^{th} stratum. An estimate of the proportion with the characteristic in the entire population is obtained by computing weighted averages of these statistics. That is

$$\hat{P} = \sum_{h=1}^{L} \frac{N_h}{N} \hat{P}_h$$

and

$$\hat{Var}(\hat{P}) = \sum_{h=1}^{L} \frac{N_h^2}{N^2} \left[\frac{N_h - n_h}{N_h} \frac{[\hat{P}_h (1-\hat{P}_h)]}{n_h - 1} \right]$$

A confidence interval for P is established using this $\hat{P}$ and $\hat{Var}(\hat{P})$ as was described with simple random sampling.

Once it is decided to use stratified random sampling, a decision must be reached as to how many elements are to be selected from each stratum. This is known as **allocation** of the sample. The simplest allocation scheme is known as **equal allocation** and involves selecting an equal number of observations from each stratum, irrespective of the sizes of the strata. That is, $n_h=n/L$, where L is the total number of strata, and n_h is the number of elements selected from stratum h. However, assuming that in advance of sampling, reliable estimates can be made of the sizes of the strata (N_h), a preferable allocation scheme is **proportional allocation.** In this scheme, the sampling fraction, n_h/N_h, is specified to be the same for each stratum. That is, the number of elements taken from the h^{th} stratum is given by

$$n_h = (N_h)(n/N).$$

When proportional allocation is used, estimates of the population mean and proportion are "self-weighting." This means that when estimating the population mean, proportion, or total, each sample element is multiplied by the same constant, 1/n, irrespective of the stratum to which the element belongs. Other allocation schemes are available but will not be discussed here.

The advantages of stratified random sampling may be summarized as follows:

- A stratified random sample of $n=n_1+n_2+...+n_L$ observations may provide increased precision (i.e., narrower confidence intervals) over that which is possible with a simple random sample of size n.
- Information concerning estimates within each of the "L" individual strata is easily obtainable since random samples have been selected in each stratum.
- For either administrative or logistical reasons, it may be easier to select a stratified sample than a simple random sample. This would probably be the case in the Costa Rican study since each of the HPs may have a reasonably up-to-date listing of the individuals for which they are responsible. It would be far more economical to take advantage of these lists than to create a new list from which a simple random sample would be selected. Furthermore, it may be of interest to obtain separate estimates of vaccination for each of the HPs so that successful ones as well as unsuccessful ones may be identified.

The major disadvantage of stratified sampling is that, since simple random samples are selected from each stratum, if lists do not exist, it is no less expensive than simple

random sampling since detailed frames must be constructed for each of the L strata prior to sampling.

III. Cluster Sampling

One solution to the problems associated with simple, or stratified, random sampling is to use a cluster sampling strategy. Sampling techniques such as simple random sampling and stratified sampling require that sampling frames be constructed that list the individual enumeration units (e.g., households). Sometimes, especially in surveys of human populations, it is not feasible to compile sampling frames of all enumeration units for the entire population. However, sampling frames can often be constructed that identify groups or clusters of enumeration units (e.g., villages, wards, etc.) without listing all the individual enumeration units. The term "cluster," when used in sample survey methodology, can be defined as any sampling unit with which one or more enumeration unit can be associated.

Cluster sampling can be performed by:

- listing the clusters
- taking a sample of clusters
- obtaining a list of enumeration units only for those clusters that have been selected in the sample
- selecting a sample of enumeration units within each of the selected clusters.

Cluster sampling is a hierarchical type of sampling in which the elementary units (e.g., children) are often at least two steps removed from the original sampling of clusters. *Cluster sampling* can be defined as any sampling plan that uses a frame consisting of clusters of listing units. Typically, the population is divided into "M" mutually exclusive and exhaustive clusters based usually upon geographic, operational, or political criteria. Unlike strata, these clusters should each be as heterogeneous as possible.

The process by which a sample of listing units is selected is typically stepwise. For example, if city blocks are clusters and households (HH) are enumeration units, there might be two steps involved in selecting the sample of HH:

Step 1: select a sample of blocks

Step 2: select a sample of HH within each block selected at Step one.

In sampling terminology, these steps are called "stages," and sampling plans are often categorized in terms of the number of stages involved. For example, a "single-stage cluster sample" is one in which the sampling is done in only one step--i.e., once the sample of clusters is selected, every enumeration unit within each of the selected clusters is included in the sample. At the first stage, "m" clusters are selected from the M available clusters. At the second stage, all N_j enumeration units are studied in the j^{th} selected cluster.

A "two-stage" sample is one in which "m" clusters are selected from the M available clusters at the first stage. At the second stage, n_j observations are selected, using simple random or systematic sampling techniques, from the j^{th} cluster, j=1,..,m. Hence samples of size $n_1,n_2,...,n_m$ are selected from the $N_1,N_2,...,N_m$ elementary units comprising the frames of each of the selected clusters.

Note that $n=n_1+n_2+...+n_m$. If $n_i=N_i$, i=1,...,m, we have a "simple one stage cluster sample". On the other hand, if $n_i<N_i$ for some i, we have a two-stage cluster sample.

A "multistage cluster sample" is performed in two or more steps. For example, to carry out the survey of vaccination status of school-aged children in rural Costa Rica, the 3,050 rural communities might be considered as the clusters. In order to select a cluster sample the following steps might be followed:

Step 1: Select m communities from the M mutually exclusive and exhaustive communities composing the nation.

Step 2: Select a sample of townships or other minor civil divisions within each of the communities selected at Step 1.

Step 3: Select a sample of school districts within each of the townships selected at Step 2.

Step 4: Select a sample of schools within each of the school districts selected at Step 3.

Step 5: Select a sample of classrooms within each of the schools selected at Step 4.

Step 6: Take every child within the classrooms selected at Step 5.

In this example, the children are the "elementary units" and the classrooms are the "enumeration units." In sampling involving more than two stages, the clusters used at the first stage of sampling are generally referred to as "primary sampling units" or "PSUs."

With these multistaged designs, writing down precise expressions for parameter estimates and associated standard errors can be difficult since each level of sampling must be accounted for. Formulae for these estimates in the simplest cluster survey designs can be found in sampling textbooks.[1,2,3,4] (Variance estimation techniques such as jackknife[5], bootstrap[6], balanced repeated replication[7], and linearization[8] are invaluable for more complex, multistage designs.)

It should be noted that in the cluster sampling schemes described thus far, the m clusters were selected at random from the M available clusters--without loss of generalizability, this selection may be achieved using systematic sampling as well. When clusters are selected with "**probability proportionate to size**," denoted "**PPS**," clusters are not selected at random. This method has considerable operational advantages and may be described as follows:

As with any cluster sampling method, the population under consideration is divided into M groups or clusters usually on the basis of geographic location. The population of each cluster must be known or estimable from a government census or other reliable source even though a sampling frame is not available for each.

Although detailed lists of enumeration units (e.g., HH) or elementary units (e.g., children) may not exist for each of the clusters on the list, all that is really needed at this point is a reasonably accurate population estimate for each cluster (e.g., community) so that PPS selection can take place. In practice, this may present some difficulties since it may be several years since the last census update. The most important assumption that must be made in this case is that any changes that have occurred in the population since the last figures were compiled affect equally all clusters in the population being studied. For example, if cluster A had 3 times as large a population as did cluster B at the time of the last census, the same 3:1 ratio is assumed to hold at the time of the survey. If census data are

not available, intelligent estimates must be made of the population sizes in the various clusters. With PPS sampling, the relative sizes of the clusters are more important than the actual sizes.

At the first stage of sampling, a subset of m clusters is selected from the complete list of M clusters. This selection is done in such a way that the probability of a cluster being selected is directly proportional to the number of individuals in the cluster. Thus, very large clusters have a much higher probability of being selected than do very small clusters. In fact, with this method it is possible for certain large clusters to be selected more than once in making up the subset of m clusters. Actual selection of clusters is carried out using random numbers.

Once the list of M clusters and associated population sizes has been compiled, "m" of the clusters can be selected with probability proportionate to size. This is accomplished systematically by computing the cumulative population, cluster by cluster, with the cumulative total equal to the total population size, N. By dividing the total population size by the number of clusters sought (30, for example), 30 "zones" are identified, each containing N/30 = K individuals. K is referred to as the "sampling interval." By selecting a single random number between 1 and K (call this "i"), the first cluster to be sampled is identified as the one that includes the i^{th} individual on the cumulative list. Starting from this position on the cumulative list, 29 further clusters are identified by successively adding the sampling interval, K. Thus, i, i+K, i+2K, ... , i+29K define the 30 clusters. This type of sampling is called a systematic sample of clusters and the probability that a particular cluster will be include in the sample is directly related to the size of the cluster.

At the second stage of sampling, $\bar{n}$ enumeration units are *randomly* selected from the total population (N_i) of enumeration units in each of the selected m clusters. In the formulae that will follow, n represents the common sample size from each selected cluster. When carried out in this manner there are a number of distinct computational advantages to the PPS strategy. Firstly, resulting estimates are "self-weighting." In other words, the sizes of the clusters do not enter into computations of proportions or associated standard errors. Secondly, as with any cluster sampling method, PPS cluster sampling has the advantage over other methods because detailed frames need be constructed only for the m clusters selected.

To estimate the proportion of the population vaccinated, the following formula is used:

$$\hat{P} = \sum_{i=1}^{m} \sum_{j=1}^{\bar{n}} y_{ij} \Big/ m\bar{n}$$

where $y_{ij}=1$ if the j^{th} child in the i^{th} cluster has been vaccinated and $y_{ij}=0$ otherwise. This can be recognized as simply the total number of children vaccinated over the total number of children studied. This estimate is self-weighting since with PPS cluster sampling there is no need to incorporate the N_i into the formula. Estimating the variance of the estimated proportion (necessary for construction of confidence interval estimates) is also a relatively straightforward procedure with PPS cluster sampling, with computation as follows:

$$\hat{Var}(\hat{P}) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (\hat{P}_i - \hat{P})^2$$

where P_i is the proportion vaccinated in the i^{th} cluster and P is the proportion vaccinated over all sample clusters as given above. This expression is considerably easier to calculate than typical cluster sampling formulae, since only the variability between the estimated proportions in the sampled clusters is needed.

In general, cluster sampling generally will not produce as precise an estimate as will simple random sampling or stratified sampling if each method were to use the same total sample size, n. However, due to the greatly reduced cost and administrative ease, a larger cluster sample may be selected, for the same cost, than that which is possible using the other sampling schemes discussed thus far. As a result of the larger sample size, a relatively high level of precision will result.

The two most important reasons cluster sampling is so widely used in practice--especially in sample surveys of human populations and in sample surveys covering large geographic areas--are feasibility and economy. Cluster sampling may be the only feasible method since the only frames readily available for the target population may be lists of clusters. If that is the case, it is almost never feasible, in terms of time and resources, to compile a list of individuals (or even households) for the sole purpose of conducting a survey. However, lists of blocks or other geographic units can be compiled relatively easily, and these can serve as the sampling frame of clusters. In addition, cluster sampling is often the most economical form of sampling since listing costs and travel costs are lowest of any potential method.

In the special case where $n_1=n_2=...=n_m=\bar{n}$, standard errors obtained by cluster sampling are approximately $\sqrt{1+\delta_x(\bar{n}-1)}$ times as large as those obtained from a simple random sample of the same total number of listing units, where δ_x is the intraclass correlation coefficient and $\bar{n}$ is the number of listing units selected in each cluster. This coefficient δ_x can range from very small negative values, when the elements within each cluster tend to be very diverse, or representative of the population of elements (this is termed "heterogeneity"), to a maximum of one, when the elements within each cluster are similar but differ from cluster to cluster (this is termed "homogeneity"). It is clear that standard errors with cluster sampling will equal those with simple random sampling when $\delta_x=0$ (i.e., heterogeneous clusters), but can be much larger when the clusters are homogeneous. The ratio of the variance with cluster sampling to the variance with simple random sampling is termed the **design effect**. This ratio of variances also applies to the total number of sampled elementary units--i.e., when the variance obtained with a simple random sample and another design (e.g., cluster sampling) are equal,

$$\sum_{i=1}^{m} n_i = (\text{design effect}) \times n_{s.r.s.}$$

IV. EPI Cluster Sampling

Since 1978, EPI has been advocating a modification of a PPS cluster sampling procedure for surveys of immunization coverage[9,10]. The method adopted is a modification of a survey technique originally used for immunization coverage in the United States[11] and

later updated for use in the Smallpox Eradication Programme in West Africa[12]. By the end of 1982 at least 441 surveys of this type had been carried out worldwide[13].

The EPI survey, for determining immunization coverage, involves the detailed review of immunization status of approximately 210 children by trained reviewers. The current convention is to identify and visit 30 clusters which may be cities, towns, or villages, and to visit as many HH as necessary in each cluster until 7 children are selected in each. As a result, the EPI surveys are commonly referred to as "30 x 7" surveys.

The rationale for using 210 individuals is as follows: Firstly, it was decided that it was necessary to be able to estimate immunization coverage to within 10 percentage points of the true population proportion. Using a population proportion of 50% coverage (the proportion with which maximal sample variance is obtained) as the basis, and desiring to be 95% confident that the resulting estimate would be in the interval 40%-60%, a simple random sample of size 96 would be required. To select a simple random sample of this size from the population was not operationally feasible and, as a result, a cluster sampling strategy was deemed necessary.

In order to achieve the same precision with cluster sampling as would be possible with simple random sampling, experience suggested that a cluster sample of approximately twice the size ("design effect") of the simple random sample would be needed. Because of the economy afforded by the cluster sampling strategy, this larger sample can be studied both more conveniently and less expensively. As a result, the necessary sample size with cluster sampling was estimated to be 192.

Based on procedures adopted for use in the United States at that time[3], and taking into account practical as well as logistic factors, it was decided that 30 clusters should be used. This meant that seven children per cluster must be studied in order to attain the specified sample size. There is no particular statistical advantage in using 30 clusters and it is perfectly reasonable that operational considerations should dictate the number to be selected. (However, it should be noted that if one is satisfied to have an estimate that will be within 10 percentage points of the true P with 95% confidence, a different combination of m and n might result in significant savings in time and cost. On the other hand, for the same time and cost, alternative combinations of m and n could yield increased precision. Decisions as to the exact value of m and n to use in a particular population would have to be tailored to the specific characteristics of that population and its environs. These decisions would involve numerous assumptions regarding costs likely to be incurred at each stage of sampling, as well as estimates of intracluster correlation coefficients which, in their own right, might be of questionable accuracy.)

The classic PPS cluster sampling scheme previously described has certain features that may make implementation in field conditions difficult. In particular, the random selection at the second stage may not always be possible--particularly in rural areas with scattered populations. The classic methodology was modified by the EPI, and the 30 x 7 survey currently being used may be characterized as a PPS cluster sample without random selection at the second stage.

Operationally, once the sample cluster is identified, it is then necessary to determine which individuals to study within the cluster. The method advocated by EPI is as follows: A household is selected "at random" from all households in the cluster. In practice, it may not be possible to make this selection truly at random, since the method used will depend upon the density of the population as well as other factors such as the availability of lists of

households. When household lists are available, the households are numbered and one random number is selected to represent the first sampled household. If, on the other hand, the cluster is a small village and household lists do not exist, a quick census of households should be taken. Households should be numbered and a random number selected to represent the first sampled household. In fact, a census is preferred because it reasonably assures a complete list.

Enumerating all households is often impossible in moderately large towns or widely scattered rural populations. In those cases, EPI suggests that the interviewer go to a centrally located landmark (such as a church, school or market), randomly select a direction in which to walk (e.g., north, south, east, or west), and count the number of households (L) found in that direction from the central point to the town boundary. Finally, select a random number between 1 and L, which will identify the randomly selected starting household.

In urban areas, the process of identifying a random starting household may be more difficult. One procedure that has been used is a two-step process in which the city is subdivided into geographically contiguous zones, a zone is selected at random, and a starting household is then identified within the zone. If household lists do not exist and if the zones are made small enough, it may be possible to carry out a census in the selected zone before selecting a household. Clearly, there is no single strategy that can be applied to all situations, and solutions are often devised on the spot to deal with living arrangements such as multiple dwelling units and apartment buildings.

It should be noted that there are certain situations where valid lists of target populations are not available and cannot be reliably constructed. For instance, in certain societies there are large numbers of "street dwellers"--i.e., individuals who do not live in permanent dwellings--who do not figure in local census data. In fact, the very concept of household may not be clear to such individuals, and they are rarely included in household lists.

Upon entering the first household, the interviewer must determine whether there are any occupants who are in the target age group. If there are, the required information is collected for each such individual. As presently recommended by EPI, if no one is at home in the selected household, the interviewer moves on to the next household. There is no provision to revisit households. (The usual procedure in EPI surveys is to schedule the times for field work to those during which the dwellings are most likely to be occupied if absences in households are common.)

After the first household is visited, whether there is an individual in the target group or not, the interviewer proceeds to the "next" household. This is defined as that residence whose front door is physically closest to the one just visited. The process of visiting households is repeated until a total of seven children of the appropriate age have been studied in the sampled cluster. As an operating rule, all eligible children in the household contributing the seventh child to the sample are studied, even if that results in 8-10 children in the cluster rather than the target number of 7.

Potential Problems with the EPI Methodology

The EPI survey has proved to be a useful tool for providing health managers with essential information for planning health programs. Normally a survey will require about 5 days of work for four to six interviewers, and instruction on the survey method is routinely included as part of EPI management training. Results from these surveys have provided the incentive for resource allocation, which has allowed immunization programs to expand and increase their impact. Without these surveys, many national programs would have no means for assessing their progress.

It should be evident from the preceding sections that the procedure followed by EPI has diverged from standard PPS cluster sampling methodolgy. This divergence occurs at the second stage of sampling (i.e., selection of households). From a statistical point of view this is a cause for concern since, in order for the formulae presented earlier in this article to hold, it is assumed that households studied at the penultimate stage are the result of random selection, and the procedure advocated by EPI does not achieve this.

Theoretically, the households should be selected independently of each other and should be representative of the totality of households in the cluster. The EPI method, by selecting a starting household and then visiting proximal households, ensures that this will not be the case. The effect of this is impossible to quantify but, intuitively, households that are spatially related may have other factors in common, including access to immunization facilities, water supply, disease exposure, etc. A particular example of this might be the pocketing of unimmunized children in slum areas of cities. There is a risk that surveys of adjacent households could either over- or under-estimate the true population coverage depending upon where the starting household happens to fall. In practice, where such pocketing is suspected, special arrangements are usually made by EPI managers.

In studies of children's attributes, PPS cluster sampling estimates will be self-weighting only if selection at each stage is based on the number of children (rather than the total number of individuals) in each sampling unit. Since lists of children are typically not available, selecting an equal number of households in each cluster will provide a set of n_i of children that can be used to provide self-weighting estimates of children's attributes. Because the EPI survey methodology is to continue to visit households until a "quota" of seven children is identified from each cluster, estimates are not self-weighting. Fortunately, ignoring this technicality typically has relatively little effect upon resulting estimates.

In some situations, the distance between the central point and the edge of the community may be too large for the prescribed EPI procedure to be practical. Although it is not advocated by EPI, the interviewer may find it more realistic to simply select the direction to be taken from a fixed starting point and to pick out and visit a house at random in that direction without first counting the number of households in that direction. Strictly speaking, this method of selecting a starting household is not random, and although it does assure that the interviewer does not exercise personal judgement in the selection process, it still introduces statistical bias.

Leaving selection of successive households to the interviewer presents another opportunity for bias. This may occur when an interviewer must decide which household is closest to the one just visited. If, for example, this choice is between one household in a slum area and another not in a slum area, there is a possibility that the interviewer's

preferences may result in one or the other not being adequately represented. Areas that are not easily accessible may be underrepresented if left to the discretion of the interviewer.

A further potential for bias is inherent in the practice of not revisiting households that were unoccupied at the time of the interviewer's visit since certain subgroups may not be adequately represented.

The selection of the starting household is also a cause for concern because the exigencies of field operations may rule out the possibility of a truly random selection. This may result in households inadvertently being selected on grounds of convenience. In a scattered rural community, for example, the tendency may be to select the starting household in the area of densest population. Such selection is subject to bias because those centers may also be the focus for outreach services and other health care facilities. In these circumstances, even if the direction in which the interviewer is to walk is selected at random and the household is selected at random from all houses lying in that direction, considerable bias may still result.

Finally, there is a potential problem in having to use nondocumented evidence of immunization status. This problem cannot be fully overcome but is often taken into account when analyzing the data.

Considering the potential sources of bias with the EPI survey strategy, there has been relatively little published research evaluating the performance of this methodology. In one such study[16] a computer simulation model was developed to evaluate the EPI survey strategy as compared to a more traditional PPS cluster sample in artificially created populations having specific characteristics. In addition to comparative measures such as bias and variability of resulting estimates, it was also of interest to investigate whether the resulting estimates were accurate to within 10 percentage points of the actual levels. It was found that within particular clusters the EPI method performed poorly when there was pocketing of vaccinated individuals; the more traditional PPS cluster sampling technique gave more accurate and less variable results under a variety of situations. However, the stated goal of the EPI, that is, being able to produce population estimates accurate to within 10 percentage points of the true levels in the population, was satisfied in the artificially created populations studied. This would appear to provide reassuring evidence to users of EPI surveys that it is possible to provide estimates of population parameters accurate to within 10 percentage points. It also sets forth a warning to those users who might make inferences about individual clusters or groups of clusters. Such disaggregation of cluster survey results is inadvisable.

Other Applications of the EPI Methodology

The EPI survey was designed for the express purpose of measuring immunization coverage, either in the absence of data, or when data of doubtful validity exist. In recent years, the sampling methodology developed for these surveys has been applied not only to assessments of immunization coverage but also to assessments of changes in immunization coverage over time. With some modifications, the methodology has also been applied to studies of the incidence of poliomyelitis, neonatal tetanus[14], and diarrhea, as well as to studies of mortality due to measles. Recently, the same procedure has been used in surveys to assess various factors relating to the availability and use of health services.

Based on the above discussion, it should be clear that the particular methodology developed by EPI specifically for coverage surveys might not be appropriate for other surveys having different objectives. For example, for surveys designed to document the expansion of coverage of an immunization program over time, sample size computation should incorporate factors such as the estimated coverage rate before expansion, the anticipated coverage rate after the expansion, as well as requirements for type I and type II error rates. The rationale underlying the number 210 does not take these factors into account and if this number were used, the study could involve an extremely high type II error rate (i.e., the probability of failing to detect a difference or change that actually occurred).

However, much of modern knowledge concerning neonatal tetanus and poliomyelitis in developing countries arises from information from EPI type cluster sample surveys[15]. Researchers studying the incidence of these diseases recognized that the procedures advocated for coverage surveys needed to be modified and, in particular, that larger sample sizes were required. A common practice is that in each of 30 clusters (the continued use of 30 clusters is based more on tradition and intuition than on statistical theory), 70 live births that occurred within a stated recall period (usually 4 to 6 months) are sought. The deaths among these live births are investigated and the proportion due to neonatal tetanus is estimated. All children aged 5-9 years in the households visited during the survey are examined for evidence of lameness due to poliomyelitis. This usually results in at least 500 children 5-9 years of age studied in each of the 30 clusters.

The need to consider carefully the choice of sample size can be illustrated by another example. If a study was planned to estimate the prevalence of a comparatively uncommon disease such as leprosy, the total required sample size would be much greater than that required in either of the previously described surveys in order to make estimates with acceptable precision. Specifically, if the rate of leprosy in a country is 1 per 1,000 (i.e., P=.001), then a simple random sample of size 15,350 would be required to be 95% confident that the sample estimate would be within 50% of the true value (i.e., between .0005 and .0015)[4]. An even larger sample would be required if a more complex sampling scheme were used or if greater precision is required.

For surveys of health service utilization and health status, there are often numerous and unrelated parameters being estimated in a wide age range. This difficulty is compounded by the fact that some of the parameters being studied are unlikely to be distributed homogenously within communities and are associated with different units. For example, households or communities are associated with available water supply or access to services, while individuals are associated with immunization status, specific types of health services, etc. Planning sample size requirements under these conditions is extremely difficult, but it is an issue that must be confronted by those needing the information resulting from such studies. Acceptance of a survey design involving 210 individuals intended to obtain information on immunization coverage may seriously compromise the results of a survey meant to measure characteristics associated with units other than individuals.

V. Quality Assurance Sampling (QAS)

The origin of QAS methods is in sampling and inspecting a manufactured product where it was necessary to keep labor and other sampling costs to minimal levels. One type of QAS 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"). 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 from it 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 average quality of a product is often continually monitored by the manufacturer to (1) identify where improvement can be made in the manufacturing process and (2) adjust the sample design as the average quality of the product changes.

The strategy and goals of QAS in the health field are similar to those in the manufacturing field. The purchaser of the 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, as supervision can then be focused on causes of defective production. It is not unusual for P_1 and P_2 to be unequal.

Generally, a lot is an "operationally useful" unit. For example, in an industrial application, if there were several machines producing the same part, and if there were 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. The manufacturer's sampling interval should be short enough to identify any drift in measurements before tolerances were exceeded. For this type of application, it would also be worthwhile to monitor the sequences of measurements for early identification of tendency to drift.

In order to provide a more familiar framework for health workers, QAS will be discussed and illustrated via the vaccination example, using LQAS. For public health work, a national manager might define lots as recipients of services from a single operational unit--such as a HP immunization team--over a specified period of time. The amount of time between sampling might be related to intervals between "high incidence" seasons for immunizable diseases, but would probably be related as much to the amount of time and cost associated with the sampling than any other single consideration. However, a HP supervisor who, for example, had three different immunization teams, might define the lots as the populations immunized in a specified period of time (e.g., the time required for a cycle of visits to all the villages/wards by each team).

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, the procedure is set up as a one-sided test. Let **d** denote the number of persons not vaccinated out of our sample of **n** subjects. Let **P** denote the true proportion of individuals not vaccinated in the population of size **N**. It is assumed that **N** is very large relative to **n**. (If it happens that **N** is not large relative to **n** then the reader should consult a text such as Brownlee[17] (Sec. 3.15) that demonstrates how the hypergeometric distribution is used to evaluate the LQAS procedure.)

The null hypothesis is

H_o: $P \geq P_o$ (i.e., proportion of unvaccinated children $\geq .50^*$)

versus

H_a: $P < P_o$ (i.e., proportion of unvaccinated children $< .50$).

The four-celled table presented in Table 1 describes the consequences of the testing procedure.

Table 1: Consequences of Hypothesis Testing in LQAS Procedure

Decision	Actual Population: Not adequately vaccinated	Actual Population: Adequately vaccinated	
Fail to reject Ho "not adequate coverage"	test recognizes or is sensitive to lack of adequate coverage $1 - \alpha$ **sensitivity**	*"Provider Risk"* β **false positive rate**	←"reject" the lot
Reject Ho "adequate coverage"	*"Consumer Risk"* α **false negative rate**	test recognizes adequate coverage $1 - \beta$ **specificity**	←"accept" the lot

Note that in this table, 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), whose probability we can control, 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 when introduced into the lot. Hence, we consider the "cost" of declaring that the population is adequately vaccinated, when in fact it is not, to be high. On the other hand, the type II error, rejection of an acceptable lot, is judged not to be as serious since the result of a false-positive decision would be to concentrate program resources on an already adequately vaccinated population.

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

---

* The level 50% is chosen here as one example. Actually, any level could have been chosen.

In practice, the critical level of the activity would be chosen first. That is, initially a "standard" 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 may be considered relative to the numbers of lots that would be misclassified with stated type I and type II errors. If the sample size were too large to be practical to use, there would be several options including:

- Retaining the sampling scheme, but lengthening the time interval between sampling.
- Choosing another "standard" (critical region) that will allow use of a smaller sample size.
- Choosing another sampling scheme, such as double sampling (perhaps even sequential sampling) that will meet the objectives of classifying the lots and still be operationally feasible.
- Abandoning a QAS scheme.

One means of computing probabilities and determining necessary sample sizes can be accomplished using the **binomial distribution.** (As previously noted, with small N, the hypergeometric distribution may be applicable; with large N, the Poisson can be practically substituted for the binomial.) The binomial distribution is the statistical distribution that describes the probability of a particular configuration of dichotomous outcomes when the total number of trials is finite (e.g., the number of times a "head" appears in seven tosses of a coin). If P denotes the probability of observing the characteristic, then the chance that there will be exactly d individuals with the event in a sample of size n is given by the expression

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

where $\binom{n}{d} = n!/[d!(n-d)!]$; ($a! = a \times (a-1) \times (a-2) \times \ldots \times 2 \times 1$ and, by definition, $0! = 1$). Thus, if 50% of the population is not vaccinated, the chance that there will be only one person who is not vaccinated in a sample of seven subjects is

$$p(1) = \binom{7}{1} (0.5)^1 (1-0.5)^6 = 0.0547.$$

Similarly, the chance of obtaining exactly one unvaccinated subject, if 70% of the population is not vaccinated, is

$$p(1) = \binom{7}{1} (0.7)^1 (1-0.7)^6 = 0.0036.$$

Suppose that the sample size is seven. The rejection region for the test states that H_o should be rejected (and "accept the lot" as adequately vaccinated) if $d \leq d^*$ (i.e., if the number of subjects in the sample found to be unvaccinated is less than or equal to the critical value, d*). First consider whether there is a value of d* such that the probability that $d \leq d^*$ when H_o is true is exactly equal to $\alpha = .05$. The probability of $d \leq d^*$, for a specified sample size n, probability P_o, and number d* is given by the expression

$$Pr\{d \leq d^*\} = \sum_{d=0}^{d^*} p(d) = \sum_{d=0}^{d^*} \binom{n}{d} (P_o)^d (1-P_o)^{n-d}$$

To establish the existence of a d* such that $Pr(d \leq d^*) = \alpha$, $Pr(d \leq d^*)$ must first be computed for a number of values of d*. In the example where n=7 and P=0.5, these values are presented in Table 2 as follows:

Table 2: Actual Probability of a Type I Error for Possible Values of d*, n=7, P=0.5

	d*							
	0	1	2	3	4	5	6	7
$Pr(d \leq d^*)$	.0078	.0625	.2266	.5000	.7734	.9375	.9922	1.0000

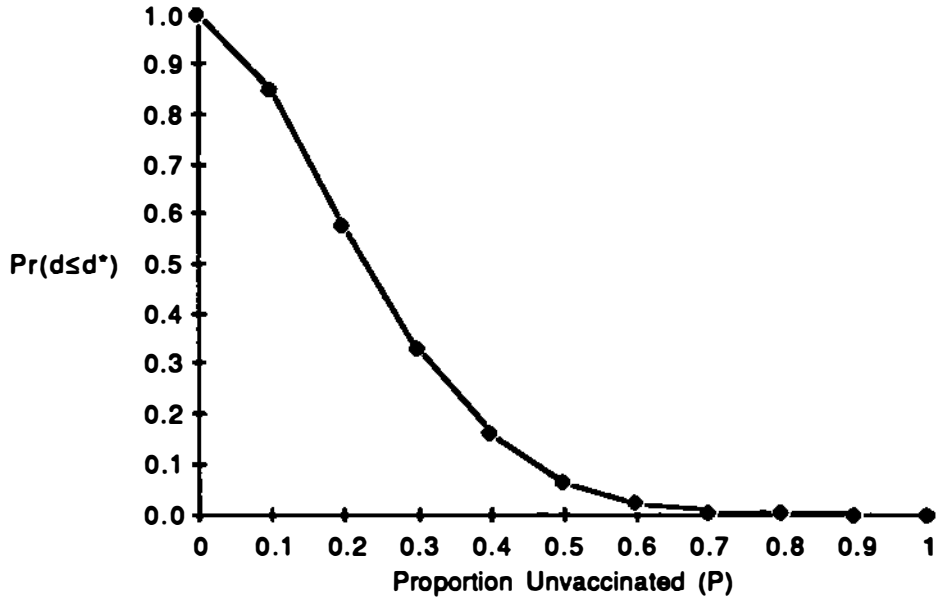
From Table 2 we see that choosing d*=0 would yield an $\alpha = .0078$ level test and choosing d*=1 would yield an $\alpha = .0625$ level test. If it was decided to take n=7 subjects, then choosing d*=1 would probably be safe; but only d*=0 results in a value of α less than or equal to .05.

Now, if we decide to use d*=1 ($\alpha = .0625$), what is the power of the test if 70% of the population is actually unvaccinated? The probability of rejecting H_0 (i.e., accepting the lot or declaring it to have an acceptable vaccination level) is the chance that $d \leq d^* = 1$, given P=0.7 and is computed as follows:

$$Pr\{d \leq 1\} = \sum_{d=0}^{1} \binom{7}{d} (0.7)^d (1-0.7)^{7-d} = .0038.$$

We may graph the results of a particular choice of n and d* into what is called 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_0: $P=P_0$ and concluding that the vaccination coverage in the population is adequate. Each combination of n and d* will generate a unique curve. We know that if no one in the population is vaccinated then P=1 and there will be no chance of rejecting H_0. On the other hand, if everyone in the population is vaccinated then P=0 and we would always reject H_0. We look for rules that give us a very high probability of rejecting H_0 when there is adequate coverage (i.e., P small). Figure 1 presents a typical OC curve for n=7, d*=1.

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



The power of the test is reflected in how steeply the curve rises to 1.0 in the region $0 \leq P \leq P_o$. For each value of n, there will be only one value of d* at or about the chosen value of α. It is usually not possible to attain the level α exactly. Thus one choice for d* will have the type I error less than α and d*+1 will have type I error greater than α. The investigator will usually choose the value of d* yielding the type I error less than α. Sometimes this strategy results in an extremely conservative test such as the one illustrated in the example above where, with n=7, d*=0 and P_o=0.5, α equalled .0078. Here the use of d*=1 with α = .0625 might be justified. Table 5 (see appendix) presents values of d* for small n ($\leq$20) such that α will not exceed the stated type I error probability (.01, .05, or .10) for various combinations of n and P_o. Details for the construction of this table are presented elsewhere[18].

The choice of the sampling scheme comes down to one of combining the desired power, 1-β, with the desired α level. Rather than providing curves, which are difficult to read precisely, for this discussion we developed Tables 6a-i (see appendix) which present values of (n,d*) pairs for chosen values of α, β, P_o and P_a. In these tables, (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[18].)

The LQAS survey problem is a one-sided test of H_o:P=P_o versus H_a:P=P_a, where $P_a<P_o$ (i.e., it is a test of the hypothesis that the proportion unvaccinated is a specified level, versus the alternative that the proportion unvaccinated is less than the specified level). A sample size is chosen that will yield a test with stated α and β errors for the particular null and alternative hypotheses specified using the standard sample size formula. Use of this formula is based on the assumption that the normal approximation to the binomial is valid. The value of d* for the necessary n is determined by using the formula

$$d^* = nP_o - z_{1-\alpha}\sqrt{nP_o(1-P_o)}$$

where values of d* are always rounded down (e.g., [5.3] = 5; [6.8] = 6). When $n \leq 20$, d* is determined by exact computations with the binomial distribution. The (n,d*) pairs are presented in Tables 6a-i.

EXAMPLE

For $\alpha=.05$ and $\beta=.20$, if $P_o=0.5$ and $P_a=0.4$, Table 6e shows that d*= 66 when n=153. In other words, if in a sample of 153 children 66 or fewer are unvaccinated, H_o would be rejected and the population would be accepted as being adequately vaccinated. On the other hand, if 67 or more children were not vaccinated, the hypothesis would not be rejected, and the population would be declared "inadequately vaccinated."

This table clearly shows the trade-off one must make between power and sample size in LQAS surveys. It is essentially impossible to have $\alpha=.05$, $\beta=.2$ 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 more serious error of concluding that an inadequately vaccinated population has adequate coverage is being guarded against by the value of α, which can always be controlled.

The method of quality assurance sampling described to this point is known as "single sampling" because only one sample is taken before a decision is reached regarding the disposition of the lot. A modification of this LQAS procedure incorporates a "double sampling" strategy and may be useful under certain field conditions. 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 advantage of the double sampling scheme is that, if the defective rate is relatively low, it may be possible to study fewer subjects than with single sampling because 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 because sampling stops as soon as the critical value d_2 is exceeded. Details for this procedure are presented elesewhere[19] and an example will be presented in Section VI.

Estimating the Overall Population Proportion with LQAS Sampling

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

For example, suppose all 294 HP's were sampled selecting seven children from each and rejecting H_o: $P \geq .5$ (and accepting the lot) if $d \leq 1$ (then $\alpha=.0625$).

Table 3: Example of Decisions in Four Health Posts

HP	N_i	* Unvaccinated	n_i	P_i	Decision
1	2501	1	7	0.14	accept the lot
2	3366	5	7	0.71	reject the lot
3	1498	0	7	0.00	accept the lot
:	:	:	:	:	:
294	2703	4	7	0.57	reject the lot
Total	750000				

The results can be combined over all 294 strata using the standard stratified sampling formulae as presented on page 3. Confidence intervals may be established as:

$$\hat{P} - z_{1-\alpha/2}\sqrt{\hat{Var}(\hat{P})} \leq P \leq \hat{P} + z_{1-\alpha/2}\sqrt{\hat{Var}(\hat{P})}$$

Hence, LQAS is really a stratified sample in disguise. Does it provide more information than conventional stratified random sampling? Of course not, because 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).

For example, if a lot is rejected when P is ≥.25 and if, in HP 3,

$$.30 \leq P_3 \leq .33$$

then the lot would be rejected.

The question then reduces to: how large should n be in order to achieve the desired precision in each stratum. It is precisely the question "...how large should n be...?" that, when answered by calculating the n required for strata in a conventional stratified random sample scheme, is a reason for considering the use of an LQAS scheme. Although the n for each stratum with an LQAS scheme 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. Because LQAS sample sizes are relatively small, there is greater likelihood that sampling can be done more frequently. Perhaps samples can be drawn concurrently with other duties that take staff to the field. Because the rules are simple to follow, the surveyor/classifier needs minimal training. And, because 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 the nation as a whole.

The potential benefits of using 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.

VI. AN EXAMPLE OF THE APPLICATION OF QUALITY ASSURANCE SAMPLING (QAS)

The example is set in circumstances similar to those in Costa Rica, and is applied to immunization coverage of children. The manager of the EPI in the country would like to know the percentage of children 12-23 months of age who have 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 HPs varies from 20% to 100%. After plotting the estimates of coverage for the individual HPs on a graph, the distribution of coverage rates is considered 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 to obtain estimates of coverage for each of the 294 areas, because 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 α=0.05. Since the average HP population is approximately 2,500, and because 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 2,500x.035 = 88. The formula for sample size determination that incorporates a finite population correction is as follows:

$$n = \frac{Nz^2_{1-\alpha/2}[P(1-P)]}{z^2_{1-\alpha/2}[P(1-P)] + (N-1)d^2}$$

Solving for n:

$$n = \frac{88(1.96)^2[.5(.5)]}{(1.96)^2[.5(.5)] + (87)(.1)^2} = 46.17.$$

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, 53% of all children in this age group will be surveyed, producing a combined sample size of 13,818. For the national estimate of coverage, the 95% confidence interval for P, based on the formula on page 17, will be

$$\hat{P} - 1.96(.0029) \leq P \leq \hat{P} + 1.96(.0029)$$

$$\hat{P} - .005 \leq P \leq \hat{P} + .005.$$

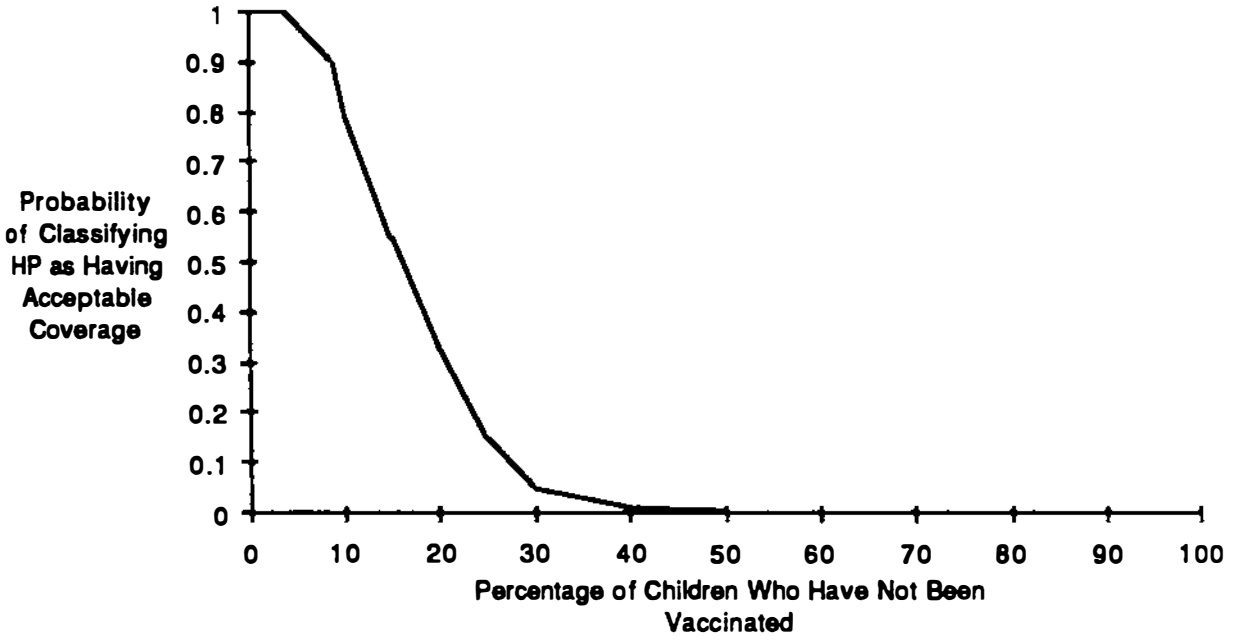
Hence, P can be estimated to within .5% (assuming the worst level of coverage for precision (50%) and little variation in HP populations).

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 scheme recommended is a double sampling scheme in which $n_1:d_1=10:0$ 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 in the 10 children, 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 in an HP area, none were found unimmunized, the HP will be categorized as having "acceptable" coverage. If four or more children were found unimmunized, the HP will be classified as having "unacceptable" coverage. In either of the above two circumstances, no further sampling is required in the HP area. However, if upon completing the survey of the initial 10 children, 1, 2, or 3 children were found unimmunized, a second sample of 14 additional children is drawn. During the survey of the second sample of children, whenever a total of four unimmunized children is reached (including those from the first sample of 10) the survey in that HP area is stopped, and the HP area is classified as having "unacceptable" coverage. If, however, upon completion of the second sample of 14 children, a total of only 1, 2, or 3 unimunized children have been found (including those unimmunized in the first sample of 10) the HP area is classified as "acceptable."

Figure 2 shows the operating characteristic curve for this particular sampling scheme.

Figure 2: Operating Characteristic Curve for Double Sampling Scheme with $n_1:d_1=10:0$ and $n_2:d_2=14:3$



Adapted from [19] Appendix 2: OC Curves For All Double Sampling Plans - (N=51-100)

An OC curve allows one to predict what the probabilities are for correctly classifying HP areas on the basis of the level of coverage. As assumed for purposes of this example, the distribution of the 294 HPs was uniform (i.e., there would be 36 or 37 HPs in each decile of the percent coverage range from 20-100 percent). Let us also assume that all HPs in each decile have a coverage that corresponds to the midpoint value for each decile: those in the 20-30% decile actually have a common value of 25%, those in the 30-40% decile have a common value of 35%, etc. If the probabilities of accepting a HP as having acceptable coverage are read from the OC curve and are applied against 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 4.

Table 4: 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%)

As computed from the results shown in the table, greater than 99% (183 of 184) of the HPs that had coverage less than 70% were "rejected" (i.e., they were classified as having an unacceptable level of coverage). Of the 110 HPs that had coverage above 70%, 62 (56%) were accepted (i.e., they were 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% were incorrectly classified as having "low" coverage, it should be noted that 63% (30 HPs) had coverage that was in the "marginal" range (i.e., 70-80% range).

Because the initial samples of 10 children were completed for each of the 294 HP, a national estimate can be computed from the combined sample size of 2,940 as with any stratified random sample. Using the assumptions that were made for the "conventional" plan, the 95% confidence interval on the national estimate of coverage from the QAS scheme results would be

$$\hat{P} - 1.96(.00915) \leq P \leq \hat{P} + 1.96(.00915)$$
$$\hat{P} - .0179 \leq P \leq \hat{P} + .0179$$

which is a level of precision that is adequate for 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 that 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 size 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 program for continual monitoring of an activity, perhaps using staff who, with minimal training, could include monitoring activity with other field duties.

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.

Glossary

Alpha (α)	The probability of rejecting a null hypothesis when it is true [Type I error].
Beta (β)	The probability of failing to reject a null hypothesis when it is false [Type II error].
d	Number of persons not vaccinated out of n subjects studied.
d^*	Critical value for number of defectives. If $d \leq d^*$, H_0 is rejected.
δ_x	Intraclass correlation coefficient.
Design Effect	Ratio of variance with cluster sampling to variance with simple random sampling.
False Negative Rate	Probability of accepting a lot that is not adequately covered.
False Positive Rate	Probability of rejecting a lot that is adequately covered.
H_a	Alternate hypothesis.
H_0	Null hypothesis.
HH	Household.
HP	Health Post.
K	Sampling interval for systematic sampling.
L	Number of strata into which the population is divided for stratified sampling.
Lot	Group of individuals receiving services from a common source.
LQAS	Lot Quality Assurance Sampling.
M	Number of clusters into which population is divided.
m	Number of clusters selected for inclusion in sample.
μ	Mean level of a random variable, X, in a population.
n	Size of sample selected from the population.
$\dot{n}$	Number of sampling units selected from each cluster in cluster sampling.
n_h	Size of sample drawn from stratum h.
N	Population Size.
N_h	Population size in stratum h.
OC Curve	Operating characteristic curve: plots probability of rejecting H_o vs various values of P.
P	Proportion of individuals in population possessing some characteristic.
P_o	Value of population proportion under null hypothesis.
P_a	Value of population proportion under alternate hypothesis.
$\hat{P}$	Proportion of individuals in sample possessing some characteristic.
$\hat{P}_h$	Proportion of individuals in sample selected from stratum h possessing some characteristic.
PPS	Probability proportionate to size cluster sampling.
PSU	Primary sampling unit.
QAS	Quality assurance sampling.

Sensitivity	Probability of rejecting a lot that is not adequately covered.
Specificity	Probability of accepting a lot that is adequately covered.
s^2	Variance of the n values of X in the sample.
σ^2	Variance of the N values of X in the population.
$\text{Var}(\hat{P})$	Variance of the estimated proportion.
$\hat{\text{Var}}(\hat{P})$	Estimate of the variance of the estimated proportion.
$\hat{\text{Var}}(\hat{P}_h)$	Estimate of the variance of the estimated proportion in stratum h.
X	Characteristic (or random variable) of individuals in the population.
$\bar{x}$	Mean level of a random variable, X, in the sample.
$z_{1-\alpha/2}$	Upper $100(1-\alpha/2)$ percentile of the standard normal distribution.

References

1. Cochran, W. G. Sampling Techniques, 3d ed. New York: Wiley, 1977.
2. Hansen, M. H.; Hurwitz, W. N.; and Madow, W. G. Sample Survey Methods and Theory. Vols. 1 and 2, New York: Wiley, 1953.
3. Kish, L. Survey Sampling. New York: Wiley, 1965.
4. Levy, P. S. and Lemeshow, S. Sampling for Health Professionals. Lifetime Learning Publications, Van Nostrand Reinhold Publishing Co., New York, N.Y. 1980.
5. Quenouille, M. H. (1956), "Notes on Bias in Estimation," Biometrika 43, 353-360.
6. Efron, B. and Gong, G. "A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation," The American Statistician, Vol. 37, No. 1, 36-48, 1983.
7. McCarthy, P. J., "Replication. An Approach to the Analysis of Data from Complex Surveys," Vital and Health Statistics, NCHS, Series 2, No. 14, 1966.
8. Tepping, B. J. "The Estimation of Variance in Complex Surveys." Proceedings of the Social Statistics Section of the American Statistical Association, 1968.
9. Henderson, R. H. and Sundaresan, T. "Cluster Sampling to Assess Immunization Coverage: A Review of Experience with a Simplified Sampling Method." Bulletin of the World Health Organization, 60: 253-260, 1982.
10. World Health Organization. "Training for Mid-Level Managers. Evaluate Vaccination Coverage". Geneva, WHO Expanded Programme on Immunization in cooperation with US Department of Health and Human Services, Public Health Service, Center for Disease Control, 1979.
11. Serfling, R. E. and Sherman, I. L. "Attribute Sampling Methods". Washington, DC., US Department of Health and Human Services, Public Health Service, Publication No. 1230, 1975.
12. 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.
13. "Expanded Programme on Immunization. Global Status Report." Weekly Epidemiological Record, 58 (23): 173-180, 1983.
14. World Health Organization Eastern Mediterranean Region / South-East Asian Region Meeting on the Prevention of Neonatal Tetanus. Lahore, 1982. (EMRO Technical Publication No. 7, SEARO Technical Publication No. 3).
15. Directorate General of Health Services, India. Guidelines. Combined Survey to Estimate the Incidence of Neonatal Tetanus and Poliomyelitis. New Delhi, Ministry of Health, 1981.
16. Lemeshow, S., et al. "A Computer Simulation of the EPI Survey Strategy" International Journal of Epidemiology Vol. 14, No. 3: 473-481, 1985.
17. Brownlee, K. A. Statistical Theory and Methodology in Science and Engineering, (2nd ed.). New York, John Wiley & Sons, 1965.
18. Lemeshow, S., Hosmer, D., and Klar, J. Sample Size Determination To be published by World Health Organization, 1987.
19. Dodge, H.F. and Romig, H.G.: Sampling Inspection Tables (2nd ed.), New York, John Wiley & Sons, 1959.

Appendix

Table 5: Values of d* for Combinations of Po and n to Achieve alpha ≤ .01, .05, or .10

	Po, alpha ≤ .01					Po, alpha ≤ .05					Po, alpha ≤ .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 6a: Sample Size and Decision Rule for LQAS, Alpha = .01, Beta = .10, One-sided Test

	Po									
Pa	0.50		0.60		0.70		0.80		0.90	
	n ,	d*	n ,	d*	n ,	d*	n ,	d*	n ,	d*
0.05	11 ,	1	7 ,	0	5 ,	0	*		*	
0.10	15 ,	2	10 ,	1	6 ,	0	*		*	
0.15	22 ,	5	13 ,	3	8 ,	1	5 ,	1	*	
0.20	32 ,	9	18 ,	5	10 ,	2	6 ,	1	*	
0.25	48 ,	15	24 ,	8	13 ,	4	8 ,	2	*	
0.30	77 ,	28	34 ,	13	18 ,	7	10 ,	4	5 ,	2
0.35	140 ,	56	50 ,	21	23 ,	11	12 ,	5	6 ,	2
0.40	321 ,	139	79 ,	37	32 ,	16	16 ,	8	8 ,	4
0.45	1298 ,	606	141 ,	71	47 ,	25	21 ,	12	9 ,	5
0.50			318 ,	170	73 ,	42	28 ,	17	12 ,	7
0.55			1264 ,	717	130 ,	78	40 ,	26	15 ,	9
0.60					287 ,	183	61 ,	41	20 ,	13
0.65					1126 ,	752	106 ,	75	28 ,	21
0.70							231 ,	170	42 ,	33
0.75							883 ,	678	70 ,	57
0.80									147 ,	123
0.85									535 ,	465

* Sample size less than 5

Table 6b: **Sample Size and Decision Rule for LQAS, Alpha=0.01, Beta=0.20, One-sided Test**

	Po									
Pa	0.50		0.60		0.70		0.80		0.90	
	n	d*	n	d*	n	d*	n	d*	n	d*
0.05	9	0	6	0	*		*		*	
0.10	13	1	8	1	5	0	*		*	
0.15	18	3	11	2	7	1	*		*	
0.20	25	6	14	3	8	1	5	1	*	
0.25	38	11	19	5	11	3	6	1	*	
0.30	60	21	26	9	14	5	7	2	*	
0.35	109	42	38	16	18	7	9	3	*	
0.40	249	106	61	27	25	12	12	5	5	2
0.45	1001	463	108	53	36	18	15	7	7	3
0.50			244	128	56	31	21	12	8	4
0.55			972	547	98	58	30	18	11	6
0.60					219	137	46	30	14	9
0.65					862	572	79	55	20	13
0.70							174	126	30	23
0.75							671	512	51	40
0.80									108	89
0.85									399	345

*Sample size less than 5

Table 6d: Sample Size and Decision Rule for LQAS, Alpha=0.05, Beta=0.10, One-sided Test

Pa	Po = 0.50		Po = 0.60		Po = 0.70		Po = 0.80		Po = 0.90	
	n	d*	n	d*	n	d*	n	d*	n	d*
0.05	6	0	*		*		*		*	
0.10	10	1	6	1	*		*		*	
0.15	14	3	8	2	5	1	*		*	
0.20	20	5	11	3	7	2	*		*	
0.25	31	10	16	5	9	3	5	1	*	
0.30	50	19	22	9	12	5	7	3	*	
0.35	92	38	33	14	16	7	8	3	5	2
0.40	211	93	52	25	22	11	11	5	6	3
0.45	853	402	93	48	31	17	14	9	7	4
0.50			210	114	49	29	19	11	9	5
0.55			834	477	87	53	27	18	11	7
0.60					191	123	42	29	14	10
0.65					746	501	72	52	20	15
0.70							156	116	30	24
0.75							589	455	49	40
0.80									102	86
0.85									362	316

* Sample size less than 5

Table 6e: Sample Size and Decision Rule for LQAS, Alpha = 0.05, Beta = 0.20, One-sided Test

Pa	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

Table 6f: Sample Size and Decision Rule for LQAS, Alpha=0.05, Beta=0.90, One-sided Test

Pa	Po = 0.50		0.60		0.70		0.80		0.90	
	n,	d*	n,	d*	n,	d*	n,	d*	n,	d*
0.05	*		*		*		*		*	
0.10	5,	0	*		*		*		*	
0.15	6,	0	*		*		*		*	
0.20	8,	1	5,	1	*		*		*	
0.25	11,	2	6,	1	*		*		*	
0.30	17,	3	8,	2	*		*		*	
0.35	31,	10	11,	3	5,	1	*		*	
0.40	68,	27	17,	6	7,	2	*		*	
0.45	271,	122	29,	13	10,	4	*		*	
0.50			65,	32	15,	7	5,	1	*	
0.55			260,	143	26,	14	7,	3	*	
0.60					57,	34	11,	5	*	
0.65					228,	148	20,	12	*	
0.70							44,	30	7,	4
0.75							174,	130	11,	7
0.80									25,	19
0.85									98,	83

* Sample size less than 5

Table 6g: Sample Size and Decision Rule for LQAS, Alpha=0.10, Beta=0.10, One-sided Test

	Po									
Pa	0.50		0.60		0.70		0.80		0.90	
	n,	d*	n,	d*	n,	d*	n,	d*	n,	d*
0.05	5,	0	3		*		*		*	
0.10	7,	1	5,	1	*		*		*	
0.15	10,	2	6,	1	*		*		*	
0.20	15,	4	9,	3	5,	1	*		*	
0.25	23,	8	12,	4	7,	2	*		*	
0.30	38,	15	17,	7	9,	4	5,	2	*	
0.35	70,	29	25,	11	12,	5	7,	3	*	
0.40	162,	72	40,	19	17,	8	9,	5	5,	3
0.45	655,	310	72,	37	25,	14	11,	6	6,	3
0.50			162,	88	38,	23	15,	9	7,	4
0.55			61,	368	67,	42	22,	14	9,	6
0.60					148,	96	33,	23	12,	8
0.65					576,	388	57,	41	16,	12
0.70							122,	91	24,	19
0.75							457,	354	40,	33
0.80									81,	69
0.85									284,	249

* Sample size less than 5

Table 6h: Sample Size and Decision Rule for LQAS, Alpha=0.10, Beta=0.20, One-sided Test

	Po									
Pa	0.50		0.60		0.70		0.80		0.90	
	n,	d*	n,	d*	n,	d*	n,	d*	n,	d*
0.05	*		*		*		*		*	
0.10	5,	0	*		*		*		*	
0.15	8,	1	5,	1	*		*		*	
0.20	11,	2	6,	1	*		*		*	
0.25	17,	5	9,	3	5,	1	*		*	
0.30	27,	10	12,	4	6,	2	*		*	
0.35	49,	19	17,	7	8,	3	5,	2	*	
0.40	111,	48	28,	13	12,	5	6,	3	*	
0.45	450,	211	49,	25	17,	8	8,	4	*	
0.50			111,	59	26,	15	10,	5	5,	3
0.55			439,	250	46,	27	14,	8	6,	3
0.60					100,	64	22,	15	8,	5
0.65					392,	262	38,	26	10,	6
0.70							81,	60	15,	11
0.75							308,	237	25,	20
0.80									53,	44
0.85									188,	164

* Sample size less than 5

Table 6i: Sample Size and Decision Rule for LQAS, Alpha=0.10, Beta=0.50, One-sided Test

	Po									
Pa	0.60		0.70		0.80		0.90			
	n,	d*	n,	d*	n,	d*	n,	d*	n,	d*
0.05	*		*		*		*		*	
0.10	*		*		*		*		*	
0.15	*		*		*		*		*	
0.20	5,	0	*		*		*		*	
0.25	7,	1	*		*		*		*	
0.30	11,	2	5,	1	*		*		*	
0.35	19,	6	7,	2	*		*		*	
0.40	42,	16	10,	3	*		*		*	
0.45	165,	74	18,	7	6,	2	*		*	
0.50			40,	19	9,	4	*		*	
0.55			158,	87	16,	8	5,	2	*	
0.60					35,	21	7,	3	*	
0.65					139,	90	12,	7	*	
0.70							27,	18	*	
0.75							106,	79	7,	4
0.80									15,	11
0.85									60,	50

* Sample size less than 5