

202.5 93RA

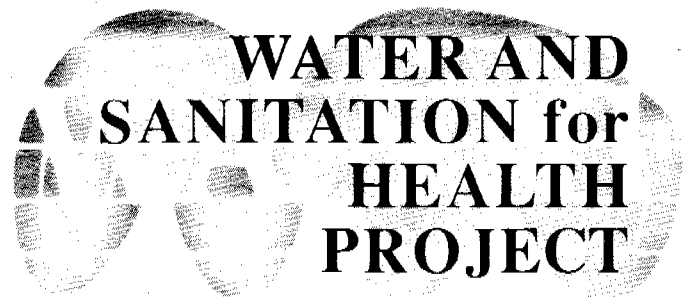
FIELD REPORT

RAPID ASSESSMENT FOR DECISION MAKING:

EFFICIENT METHODS FOR
DATA COLLECTION AND ANALYSIS

LIBRARY
INTERNATIONAL REFERENCE CENTRE
FOR COMMUNITY WATER SUPPLY AND
SANITATION (IRC)

WASH Field Report No. 391
January 1993



Sponsored by the U.S. Agency for International Development
Operated by CDM and Associates

202.5-93RA-
10458

WASH Field Report No. 391

RAPID ASSESSMENT FOR DECISION MAKING: EFFICIENT METHODS FOR DATA COLLECTION AND ANALYSIS

Prepared for the Bureau of Research and Development,
Office of Health,
U.S. Agency for International Development
under WASH Task No. 153

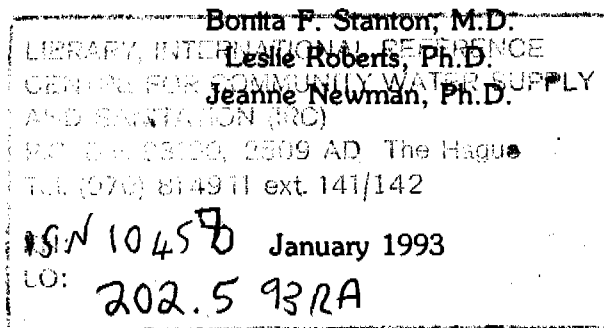
by

William A. Reinke, Ph.D.

~~Bonita F. Stanton, M.D.~~

~~Leslie Roberts, Ph.D.~~

~~Jeanne Newman, Ph.D.~~



Water and Sanitation for Health Project
Contract No. 5973-Z-00-8081-00, Project No. 936-5973
is sponsored by the Office of Health, Bureau for Research and Development
U.S. Agency for International Development
Washington, DC 20523

Related WASH Reports

Cholera Prevention and Control: Guidelines for Assessing the Options in Water Supply, Sanitation, and Hygiene Education, by Sarah K. Fry. Field Report No. 380. April 1992.

Development of a Behavior-Based Monitoring System for the Health Education Component of the Rural Water and Health Project, CARE-Guatemala, by Lori DiPrete Brown and Elena Hurtado. Field Report No. 364. July 1992.

Evaluation Guidelines for Community-Based Water and Sanitation Projects, by Philip Roark. Technical Report No. 64. May 1990.

CONTENTS

ACKNOWLEDGMENTS	vii
ABOUT THE AUTHORS	viii
ACRONYMS	ix
FOREWORD	xi
EXECUTIVE SUMMARY	xv
1. INTRODUCTION	1
Background	1
Scope of the Report	1
What You Need to Know to Use This Manual	2
2. OVERVIEW OF RAPID ASSESSMENT METHODS	3
Cluster Sampling	3
Lot Quality Assurance Sampling	4
Double Sampling	4
Reduced and Tightened Inspection	5
Epidemiological Surveillance	6
Demographic Surveillance	7
Industrial Process Control Methods	7
Case-Control Analysis	8
Sociocultural Group Assessment Methods	8
3. STATISTICAL PRINCIPLES	11
Discrete Measurement vs. the Continuous Scale	11
Sample Size	12
Specifying Precision	13
4. SELECTING RAPID ASSESSMENT METHODS	17
Informational Issues	17
What kind of information is needed and how will it be used?	17
How will the information be useful for decisionmaking?	19
How much information is needed and at what level of quality?	19

Methodological Characteristics	20
Is the method easy to use (or misuse)?	20
Will the method yield timely information?	20
Cost Considerations	21
Is the method cost-effective?	21
What resources are available for the task?	22
5. CLUSTER SAMPLING	23
Simple Random Sampling vs. Cluster Sampling	23
Description of the Cluster Sampling Method	23
Representative Uses of Cluster Sampling	24
Operational and Analytical Difficulties	24
Balancing Variation between and within Subunits	24
Design Effect	25
Heterogeneity of Clusters	26
Peri-urban Issues	26
Stratification	26
Bimodal Distributions	26
Advantages and Disadvantages/Problems/Warnings	27
Conclusion	27
6. LOT QUALITY ASSURANCE SAMPLING	29
Origins in Industry	29
Representative Uses of LQAS	29
"Significance Testing" Character of LQAS	30
Errors of Commission and Omission	30
Decision Rules	31
Applying Decision Rules and Sampling Plans	31
Advantages	35
Disadvantages	35
Validity of Results Achieved from Small Samples	36
7. DOUBLE SAMPLING	37
Relationship to Lot Quality Assurance	37
Description of the Method	37
Representative Uses	37
Advantages	39

Disadvantages	39
Sequential Sampling	39
8. REDUCED AND TIGHTENED INSPECTION	41
Reduced and Tightened Inspection Compared with Double Sampling	41
Description of the Method	41
Representative Use of Reduced and Tightened Inspection	42
Evaluation of the Method	43
9. EPIDEMIOLOGICAL SURVEILLANCE	45
Description of the Method	45
Representative Uses of the Method	45
Surveillance Sites	45
How Epidemiological Surveillance Can Be Made Rapid	46
Active and Passive Systems	46
Range of Applications	47
Peri-urban Applications	47
10. DEMOGRAPHIC SURVEILLANCE	49
Difficulties of Demographic Surveillance in Developing Countries	49
Techniques for Rapid Demographic Assessment	49
Pregnancy Histories	49
Routine Encounters	50
Survival of Close Relatives	50
Verbal Autopsies	50
Limitations of Rapid Demographic Surveillance	50
11. INDUSTRIAL PROCESS CONTROL METHODS	51
Process Control Methods—Not Yet Used in the Health Field	51
Description of the Method	51
Types of Process Control Methods	51
Ranges and Averages	51
P-Charts for Proportions	54
C-Charts for Case Counts	54
Moving Averages	55

Exponential Smoothing	55
Average Outgoing Quality Limit	56
Assessment of the Method	56
12. CASE-CONTROL ANALYSIS	59
Locating Cases of Rare Conditions	59
Description of the Method	59
Representative Use of the Method	59
Finding Representative Cases and Controls	60
Avoiding Distortion	60
Effects of Confounding	60
Interaction Effects	60
Separating Cause and Effect	60
Avoiding Difficulties	61
Odds Ratios	61
Experience with the Case-Control Approach	62
Validation of the Findings of Case-Control Studies	63
Assessment of the Case-Control Approach	63
13. SELECTING THE MOST SUITABLE METHOD	65
BIBLIOGRAPHY	71
MATHEMATICAL APPENDIX	
Sample Size Determination	79
FIGURES	
1. Comparative Sample Sizes	15
2. Operating Characteristics Curves	32
3. Ability to Detect Unsatisfactory Conditions (When Fewer Than 80 Percent Know Therapy)	33
4. Probability of Correct Classification	34
5. Underestimation of Unsatisfactory Conditions	35
6. Sequential Sampling for BCG Immunization	40
7. Control Charts for Home Visiting	53

TABLE

1. Selecting Rapid Assessment Methodologies 67



ACKNOWLEDGMENTS

The authors wish to acknowledge the many colleagues who have contributed to this review of rapid assessment methods. Thanks are due to those who participated in an initial roundtable held at the WASH offices to consider methods available for rapid assessment of water and sanitation-related health problems in peri-urban areas, and to develop a framework for this review. The roundtable included participants from United States Agency for International Development (A.I.D.), the Centers for Disease Control (CDC), UNICEF, the Research Triangle Institute, the Harvard Institute of International Development, the University of Massachusetts, Johns Hopkins University, the University Research Corporation, and the A.I.D.-funded Vector Biology and WASH Projects.

Thanks are also due Drs. Masee Bateman and Stewart Blumenfeld for valuable technical insight and critical review, and to Diane Bendahmane, Lynda Edwards, and Christine De Joy for editorial assistance.

ABOUT THE AUTHORS

William A. Reinke is Professor and Deputy Chair in the Department of International Health, Johns Hopkins School of Hygiene and Public Health, where he also holds appointments in Biostatistics and Health Policy and Management. He has a Ph.D. in Statistics from Case Western Reserve University and several years of business experience in the field of industrial quality control. Since joining the Johns Hopkins faculty in 1964 he has participated in numerous operations research and technical assistance projects in Chile, China, Colombia, Egypt, India, Indonesia, Kenya, Malawi, Nigeria, Thailand and Uganda, mainly in the fields of health sector planning, management, and evaluation.

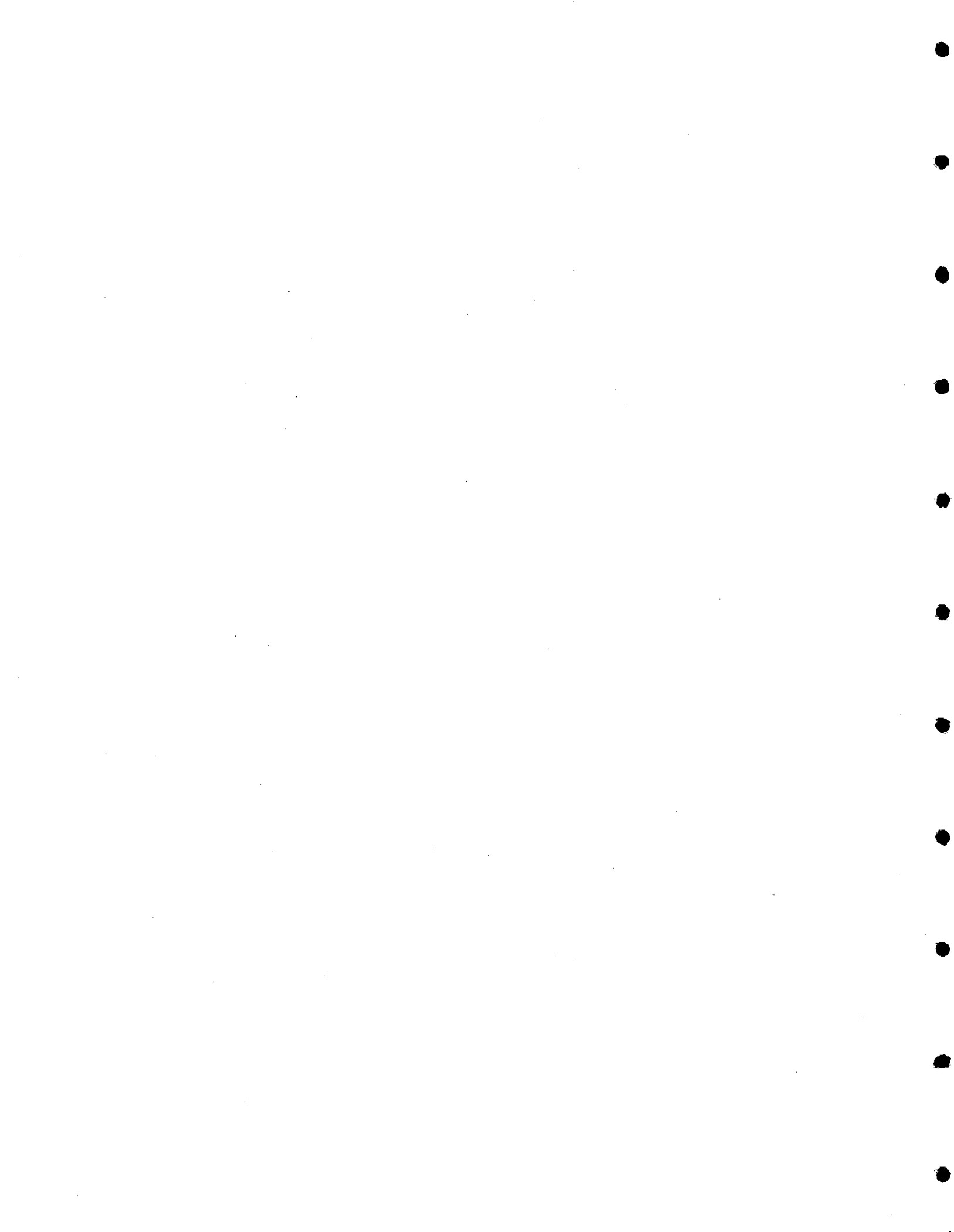
Bonita Stanton is Professor of Pediatrics at the University of Maryland. She received her M.D. from Yale Medical School where she also completed a Pediatric Infectious Disease fellowship. She lived and worked for 5 years in Bangladesh, and has worked for shorter time periods in many other Asian and African countries. Her research and service in these countries have focused on peri-urban health issues, and particularly on diarrheal diseases.

Leslie Roberts is currently an EIS Fellow in the International Health Program Office at the Centers for Disease Control. He has a Ph.D. in Environmental Engineering from the Johns Hopkins University and a Master's degree in Environmental Health from the Tulane School of Public Health and Tropical Medicine. He has conducted studies of water and sanitation in Peru and worked as a stormwater engineer for the State of Maryland. He is presently on assignment assessing sanitary interventions for refugees in Malawi.

Jeanne S. Newman has worked for more than 20 years in the planning, management, and evaluation of public health services in developing countries. She has a Ph.D. from Johns Hopkins University in Geography and Environmental Engineering, and an MBA in Health Management from the Sellinger School of Business Administration (Loyola College, Baltimore). Currently Senior Technical Advisor in the International Development Group of the University Research Corporation (Bethesda, MD), her research and technical assistance focus on strategies to incorporate quality assurance as a routine management function in primary health care.

ACRONYMS

A.I.D.	United States Agency for International Development
AOQL	Average Outgoing Quality Limit
AQL	Acceptable Quality Level
CDC	Centers for Disease Control
EPI	Expanded Program of Immunization
HP	Health Post
LQAS	Lot Quality Assurance Sampling
ORS	Oral Rehydration Salts
ORT	Oral Rehydration Therapy
QAS	Quality Assurance Sampling



FOREWORD

Programs in health, or any other field of development assistance, are only as good as the information on which they are based. In recent years, considerable attention has been focused on devising simple, inexpensive, and rapid methods of gathering information on disease rates, lifestyles, socioeconomic conditions, and factors contributing to both ill health and utilization of services. This volume assembles and synthesizes information on these rapid assessment methods to make them easily accessible to planners and policymakers. Although the methods are described here principally as they apply to the field of health—and in particular health conditions relating to water supply and sanitation—the methods may also be applied in other contexts and so should be of wide interest.

A sense of urgency has attended the preparation of this compendium of methods, as national and international health planners increasingly realize that the world in which they work is becoming highly urbanized, while their own experience tends to be mainly rural. There is a need for reliable information on which to base urban health planning. For that reason these methods are geared not only toward health applications but also toward use in urban and peri-urban settings.

Just how urban is the world becoming? In 1980, the world's urban population was estimated at 1.8 billion; in the fast-approaching hallmark year 2000, it will be 3.2 billion. Overall the world's population will have increased 41 percent, while the urban population will have increased a whopping 78 percent (World Health Organization, 1985). By 2000, Latin America will be 77 percent urban, Africa 42 percent, East Asia 79 percent, South Asia 35 percent, and China 40 percent (Rietveld, 1988). From 1950 to 1980, Nairobi, Dar es Salaam, Abidjan, Lusaka, Lagos, and Kinshasa grew sevenfold or more; in the same period, Baghdad, Bombay, Dhaka, Jakarta, and Seoul tripled or quadrupled their populations. Mexico City is home to 18 million, Sao Paulo to 14 million. In 1981, the urban population in India was estimated at 167 million; by the year 2000 it will be 367 million (Cutting, 1988). The following comparison gives some idea of the speed at which this urbanization is proceeding.

Between 1800 and 1910, Greater London's population grew almost sevenfold, from 1.1 million to 7.3 million, an increase now achieved within a generation in many Third World cities. Similarly, Paris took more than a century to expand from 547,000 to about 3 million, a growth matched by some Third World cities just since World War II (Rietveld, 1988).

Such remarkable rates of urban population growth would be difficult to support in any country, regardless of developmental status. In developing countries the problem is exacerbated because this growth typically takes place in locations with little or no existing health and sanitation infrastructure. Settlements are literally formed overnight, with no planning or official recognition.

The situation in Bangladesh is illustrative. From 1961 to 1974, the urban population increased 140 percent; during these years the annual urban growth rate was 6.7 percent, compared with a general population growth rate of 3.7 percent. In Bangladesh, as in many countries, municipal areas are not included in the national health system and therefore are often excluded from national health programming. Even when all national constituencies are recognized in planning, long-standing needs to alleviate poverty in the rural majority must compete with the emerging urban demands in the allocation of extremely scarce resources.

The lack of health services in peri-urban areas often is not apparent to the uninitiated, or it may be disguised by the existence in all countries, rich or poor, of at least some urban areas with relatively satisfactory population-bed or population-doctor ratios. People tend to leap to the conclusion that these satisfactory ratios are the norm rather than the exception.

It is true that sophisticated urban medical facilities exist in many developing-country cities, but this does not mean they are available to all urban dwellers. Normally, they serve only the nation's wealthy families. High-technology facilities constructed in response to demands from the powerful elite in no way ensure access by marginal dwellers in the area and may actually decrease the general availability of primary health care (Fosu, 1989).

In many instances medical facilities exist close to both urban rich and poor. But this does not mean that the services are truly accessible to the poor. In rural areas, distance to a clinic or health center is often a factor limiting access of the poor to health services. In urban areas, it may not be distance but political control factors, restricted hours of service, and long waits that limit access (Okun, 1988).

Because squatter settlements are often illegal, or at least unsanctioned, certain segments of the population may not be included in government census enumerations. This results in inflated figures of resources expended per capita and also often makes squatters ineligible for the resources.

A number of health conditions are peculiar to low-income urban or peri-urban settings. Traditionally, rural mortality rates have exceeded urban rates in developing countries, but in several countries this relationship is being reversed, at least with respect to the rapidly growing urban slums. In Thailand, the 1970 mortality rate in the Bangkok metropolitan area was found to be one-third higher than the national average (Khanjanasthiti, 1974). In Haiti, death rates in the slums of Port-au-Prince were triple those in rural areas (Rohde, 1983).

Patterns of morbidity are also complex, but evidence suggests that the rate of water and sanitation-related disease is disproportionately high in peri-urban areas. While available data sets are not strictly comparable, they indicate that urban children in Bangladesh may experience up to one-third more diarrhea episodes than their rural counterparts (Black et al., 1982; Stanton, Clemens, and Khair, 1988). Recent data from a peri-urban area in Peru showed an average of 10 episodes of diarrhea in infancy, a higher level than any reported previously (Lopez de Romana et al., 1989). Infestations with parasites such as ascaris have been noted to be substantially higher in urban than in rural areas (Benyoussef et al., 1973).

Several developing countries reveal an incidence of low birth weight among the urban poor that is well above the national average. For example, according to two slum-based studies (Singh and Paul, 1988; Patel, 1989), low birth weight in Delhi and Calcutta occurred in more than 50 percent of births, compared with rates about half as high in India overall. Likewise, some national data sets from Bangladesh show a higher prevalence of malnutrition in the slums of Dhaka than in many rural areas (HKI, 1985). The picture is far from uniform, however, and care must be taken not to draw oversimplified conclusions. Some studies, for example, have revealed a nutritional status among urban dwellers that is superior to rural averages (CUS, 1989).

Certain characteristics of urban life itself may markedly affect health status or health-seeking behavior. For example, one study showed evidence of rural-urban differences in the impact of female wage labor on child-rearing practices (Engle, 1989). In rural areas, women appear to have developed support systems enabling them to engage in wage labor for up to eight hours per day before they perceive a decrease in time spent on child care; whereas, urban women report a decrease after just one hour of wage labor. Thus, a degree of access to health care systems that is adequate for rural women might be inadequate for urban women in the labor force.

Until planners and policymakers understand more about mortality and morbidity and sociocultural factors in peri-urban areas, extreme care must be exercised in extrapolating findings from rural studies. Even in countries where much is known about rural water, sanitation, and health practices, parallel studies in peri-urban areas will be necessary to test for possible differences.

As noted, rapid assessment methods are not limited in their applicability to peri-urban areas or to the field of health. But the authors have given this report a distinct peri-urban water-sanitation-and-health twist. Thus, the problems and solutions referred to relate to water and sanitation, and the reader is reminded to be aware of the unique features of peri-urban settings that complicate the acquisition of reliable data. Several of these features are worth mentioning at the outset.

- Because peri-urban areas frequently fail to gain official recognition, prior knowledge about them is usually inadequate. Census data may be unavailable, and other baseline information is likewise either unavailable or sketchy. These deficiencies make it difficult, for example, to select a representative sample to study and thus call for innovative methods of data gathering.
- Even when census or other baseline data are available, they become outdated quickly because of rapid migration in peri-urban areas. Therefore, peri-urban population dynamics must be recognized as a major concern. High rates of migration into urban areas, often coupled with frequent movement within the area itself, make it difficult not only to select a representative sample but also to follow individuals or families over time.

- Many peri-urban communities are extremely heterogeneous, especially in comparison with their rural counterparts. For example, in the peri-urban settlement of Pikine in Dakar, Senegal, eight major ethnic origins are represented (Fassin and Jeanne, 1989). This means that within a small geographic area, a number of variables are likely to contribute significantly to differences in health beliefs and practices. Such heterogeneity must be recognized in any data-gathering plan. The presence of diverse subgroups may suggest that each be treated as a separate entity. If an overall population assessment is desired, care must be taken to ensure that each of the groups is properly represented in the analysis.

Regardless of the venue or sector, however, there is a need for reliable information on which to base policy and planning. This document addresses these information needs by presenting a review of assessment methods that may be used to gather information in a rapid and efficient fashion. While the focus and examples employed here concern the water-and-sanitation sector and urban populations, the principles and methods described also are useful in other sectors and in rural areas.

EXECUTIVE SUMMARY

Considerable attention has been focused recently on devising simple, inexpensive, rapid methods for collecting and analyzing data on health conditions. Sector personnel charged with developing water and sanitation programs for peri-urban areas have a particular need for rapid assessment methods because little reliable health information on these areas is available. Until planners and policymakers understand more about mortality and morbidity and sociocultural factors in peri-urban areas, extreme care must be exercised in extrapolating findings from rural studies. Even in countries where much is known about rural water, sanitation, and health practices, parallel studies in peri-urban areas will be necessary to test for possible differences.

Rapid methods that have been devised include techniques from epidemiology, biostatistics, industrial quality control, and a variety of other disciplines. Nine such rapid assessment methods are described in this report, eight in some detail. The descriptions provide basic information for project managers and policymakers who must collect the information necessary to make decisions and manage programs in developing countries. Each method is explained in terms easily understandable by nonspecialists, and its strengths and weaknesses are enumerated. The report emphasizes not *how to*, but *when to* use the methods, seeking to answer the question: Under what circumstances are these methods both efficient and cost-effective?

In addition to descriptions of the methods, the report contains chapters on basic statistical principles for the uninitiated and on criteria for selecting the most appropriate method in a given circumstance.

The rapid methods covered are listed below with a few words of explanation for each one.

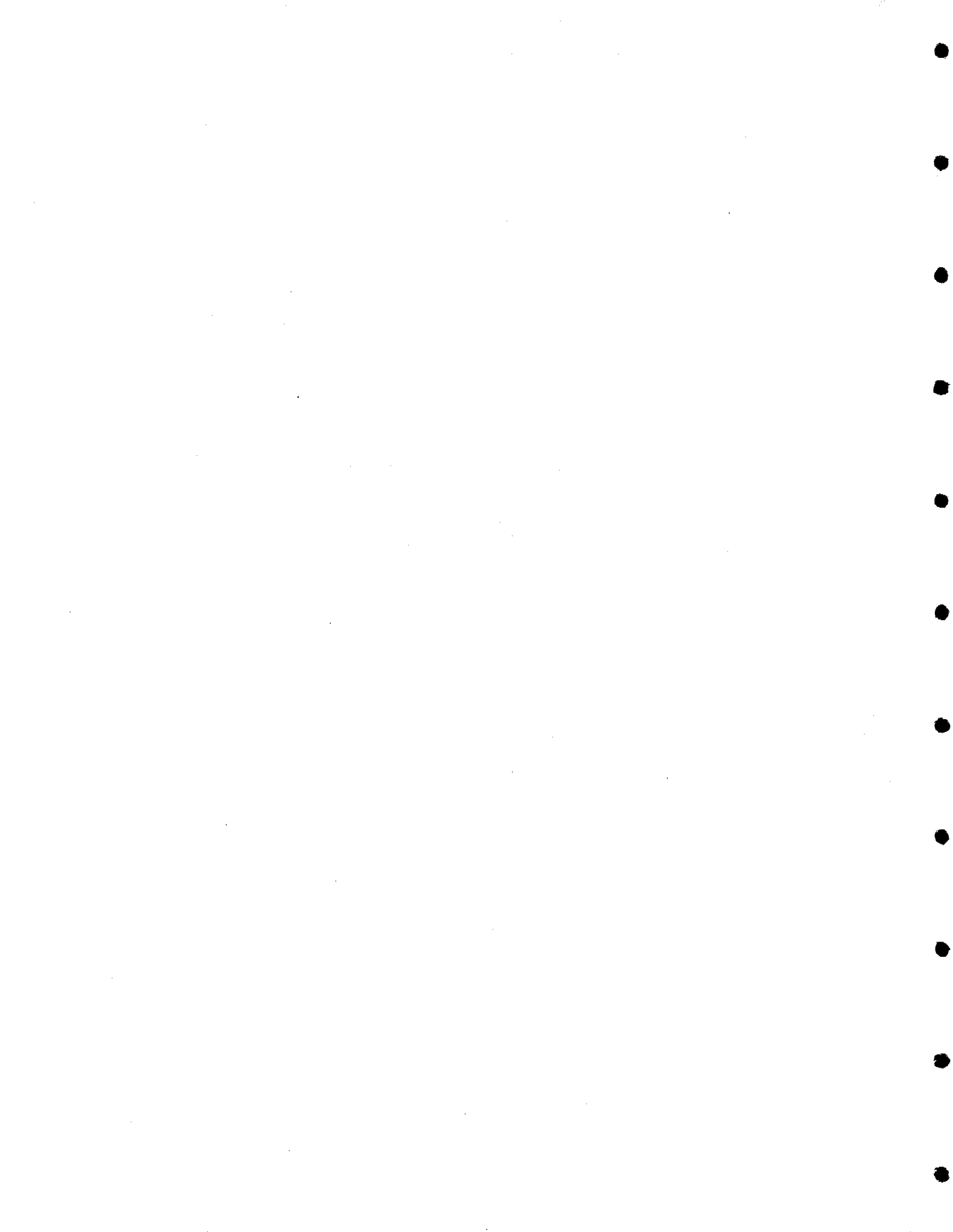
- *Cluster sampling* is a practical alternative to the more costly and time-consuming simple random sampling. Cluster sampling draws upon a sample selected from a list of groups, or clusters, such as city blocks, neighborhoods, or squatter settlements, rather than a list of single individuals or small entities. The cluster method lends itself to assessments of health behaviors in large peri-urban areas that encompass many neighborhoods. Respondents would be drawn from only a specified number of these neighborhoods, which would be selected randomly.
- *Lot quality assurance sampling (LQAS)*, a method borrowed from industrial quality control, is based on the assumption that a small sample from a batch (or lot) is representative of the whole batch. With this method, large samples are unnecessary because only small amounts of information are obtained. It is appropriate for situations in which a "yes" or "no" will yield enough information upon which to act—for example, do at least 70 percent of households have access to a toilet within the building?
- *Double sampling* is a way to gather more precise information than can be obtained with lot quality assurance sampling. Instead of increasing sample sizes across the

board, a follow-up assessment is carried out only for those lots or batches where the results are ambiguous, neither clearly "yes" nor "no."

- *Reduced and tightened inspection*, like double sampling, keeps the sample size small but, instead of re-examining lots or batches that fall below a certain standard, this method either tightens or loosens the level of scrutiny as needed, while the assessment is taking place. As the survey proceeds from subunit to subunit, the results are plotted. If any subunit falls below the acceptability level or if the average drops below the acceptability level, tightened inspection begins by using the same sample size but allowing fewer defectives. If fewer than the new target number meet the standard, the assessment is discontinued and it is concluded that the standard has not been met in the universe being assessed. If the overall results again reach the standard, then looser inspection is resumed.
- *Epidemiological surveillance* involves the systematic collection, consolidation, and evaluation of data to assess conditions at a particular time or to monitor changing conditions. Such a method might be used to study seasonal patterns of diarrhea, for example, or trends in the use of oral rehydration therapy (ORT). This information can usually be collected by individuals routinely employed for other purposes.
- *Demographic surveillance* is the monitoring of births and deaths, a difficult process in developing countries and not usually "rapid" or inexpensive. However, demographers have devised a number of techniques useful for rapid assessment: obtaining pregnancy histories, regularly gathering information at the time of delivery or vaccination, and monitoring the survival of the close relatives of a respondent.
- *Industrial process control methods* have rarely been used in the health field to date. Process control consists of charting a particular condition or circumstance, thereby providing a visual representation of the results. By means of pre-established control limits placed upon a control chart, observers can determine whether a problem exists and how serious it is. A control chart could be used, for example, to monitor the incidence of chronic diarrhea. The number of cases plotted inside and outside the control limits would delineate the nature of each outbreak.
- *Case-control analysis* begins with the selection of readily available "cases" and "controls" for testing a hypothesis about factors associated with a particular condition or circumstance. For example, the cases might be small children with persistent diarrhea whose mothers brought them to a particular hospital for treatment; the controls could be children within the same age group who were being treated at the facility for another condition, say injuries. The two groups could be compared regarding their usual source of drinking water to determine whether it played a role in the diarrhea of the cases.
- *Sociocultural group assessment methods* are only mentioned briefly in the report. Such methods, which might take the form of surveys of representative individuals or intensive interrogation of a few key persons or focus groups, come into play when the

aim is not to collect and analyze objective information but to define basic attitudes, beliefs, or desires of a particular group or even a culture as a whole, or to develop hypotheses for further investigation.

In designing a project, information is a tool for identifying the prevalence of problems, determining their causes, and designing solutions. After a project begins, information becomes essential as a means for monitoring and evaluation. Projects require different information at different stages: it may be epidemiological information on the prevalence of disease and associated risk factors; information on environmental conditions contributing to those health problems; information on behavioral practices; or information on institutional capabilities and constraints. Rapid assessment requires that managers or other decision makers first identify the specific data needed to make a particular decision and then select the most efficient methodology to collect and analyze that information.



Chapter 1

INTRODUCTION

Background

Decision makers need information: It must be reliable, timely, and relevant to the decision being made, and available in sufficient quantity to permit analysis. Moreover, it must be obtained in a cost-effective way. Increasingly, the term *rapid assessment* is used to define an efficient method for obtaining and analyzing data needed for decisionmaking.

The decision to be made determines the quantity and type of information and the levels of precision required. In designing a project, information is required to identify the prevalence of problems, determine their causes, design solutions, and plan for implementation of solutions. After a project begins, information becomes essential as a means to monitor implementation and evaluate and modify project efforts. Projects require different information at different stages: it may be epidemiological information describing the prevalence of disease and associated risk factors; information on environmental conditions contributing to those health problems; information on behavioral practices; or perhaps information on institutional capabilities, processes, and constraints. A rapid assessment approach requires first, that managers or other decision makers identify the specific data needed to make a particular decision (to avoid haphazardly collecting too much data, extraneous data, or the wrong data altogether); and second, that they then select the most efficient methodology to collect and analyze that information.

Scope of the Report

This presentation has two objectives: (1) to provide a systematic, critical review of the rapid assessment methods currently in use or being proposed, and (2) to examine under what circumstances specific methods should be adopted. Focusing on improved management decisions and program actions, this broadly based critique covers ease of data collection, simplicity of analysis, and interpretation of findings. The reader should gain a clear idea of how each of the methods functions, discover its strengths and limitations, and develop a sense of the informational needs best served by each.

The review places primary emphasis, however, on the circumstances that call for the application of specific methods and, therefore, may be considered a *when-to* rather than a *how-to* document. Field-oriented examples provide a practical focus for project managers and policymakers who must collect the information necessary to make decisions and manage programs in developing countries. Although this volume was developed in response to the specific data needs of urban and peri-urban water supply and sanitation projects, the information provided herein applies equally well to a variety of health-related assessments in both urban and rural settings.

The information is presented in two parts: Chapters 1 through 4 provide a brief overview of several assessment methods; review statistical issues, particularly those relating to sample size; and discuss the issues that must be considered when selecting a methodology. Chapters 5 through 12 describe eight methods in some detail. The final chapter organizes the nine rapid assessment methodologies into a table that will help managers select the most suitable method for obtaining health information. Table 1 presents a practical guide for using the nine assessment methodologies. If managers need information about a particular health condition, they can use the table to select the most appropriate method for obtaining that information.

What You Need to Know to Use This Manual

This presentation is written for the nonexpert in statistics and assessment methodologies. In fact, its purpose is to make rapid assessment methodologies and the principles used to select among them accessible to those who need information, primarily program managers and decision makers. The material should be understandable to anyone generally familiar with statistical and sampling principles. Those with more experience and training may wish to refer directly to the descriptions of individual methods in the second part of the report. Nonetheless, a review of Chapters 3 and 4 is recommended to provide the context and background for judging the utility and appropriateness of the methods described here and other assessment methods that may be under consideration for a specific application.

Chapter 2

OVERVIEW OF RAPID ASSESSMENT METHODS

Rapid assessment methodologies include techniques from a variety of disciplines, such as epidemiology, biostatistics, industrial quality control, and the social and anthropological sciences. Several methodologies are currently available or have been proposed, each with its own strengths and each requiring careful consideration before being applied. This chapter introduces nine assessment methods.

Cluster Sampling

This method, quicker and cheaper than simple random sampling, selects samples from a list of groups (*clusters*) rather than a list of single individuals or small entities. As with a simple random sample, heads of households, mothers, families, and so on might still serve as the *final sampling units*; however, these respondents would be found only in a sample of "clusters," such as city blocks, neighborhoods, or squatter settlements.

With this method, the area of interest is divided geographically into primary sampling units—districts or counties, for example. From these units, a given number are randomly selected and further divided into *subunits*, with the process continuing until the lowest level of subunit has been identified. From randomly selected subunits at the lowest level, an equal number of respondents are selected for interview. A widely used formula calls for 7 respondents from 30 clusters; however, the number ultimately chosen should serve as a reasonable compromise between the ideal single observation per cluster and the logistically simpler but statistically less efficient multiple observations per cluster.

The cluster method lends itself to assessments of health behaviors in large peri-urban areas that encompass many neighborhoods. Respondents would be drawn from only a specified number of these neighborhoods, which would be selected randomly.

Advantages: This method provides a fast, logistically simple, and inexpensive way to determine knowledge and practices regarding a defined activity (e.g., latrine use) from a large, possibly dispersed population.

Disadvantages: Cluster samples provide reliable information only about the population as a whole; sample sizes are generally too small to draw conclusions about a single cluster. Moreover, when communities encompass both well-served and underserved groups or, in this case, high-use and low-use groups, overall averages obtained through cluster sampling can be misleading.

Lot Quality Assurance Sampling

Another quick, inexpensive, and easy assessment method—lot quality assurance sampling (LQAS)—has recently gained wide recognition in the health field. With this method, large samples are unnecessary because only small amounts of information are obtained: in industry, where LQAS was developed, it is assumed that when a small sample from a batch (or lot) passes inspection, the entire lot under examination is sound. In epidemiological surveys, the lot would be the target group.

The method is appropriate for situations in which a “yes” or “no” will yield enough information upon which to act: does the respondent have a latrine, for example. Responses would allow policymakers to discover which health districts needed help with their latrine-promotion programs. It is very important, however, to set the pass/fail points appropriately. Chapter 6 discusses how this may be done and also touches upon determination of sample size and the pass/fail number of defectives.

The main thing to bear in mind is that lot quality assurance sampling works best when managers or policymakers need to identify exceptional needs or inadequate services. While not a method to uncover community attitudes or hidden practices, LQAS can do the following well:

- Give feedback on existing conditions
- Identify any departures from predetermined goals or standards
- Associate those departures with individual communities or service units

Advantages: Lot quality assurance sampling can detect gross departures from prevailing standards. Quick, inexpensive, and easy to administer, it is a useful method when policymakers can get by with rough information. Small sample sizes make frequent assessment feasible.

Disadvantage: Its small sample sizes make the LQAS method reliable only for detecting large departures from prescribed standards.

Double Sampling

Whereas small samples may be adequate to detect clearly acceptable or clearly unacceptable conditions, ambiguous intermediate circumstances require more information. With double sampling, an initial lot quality assurance sample is followed by a second assessment of those lots or batches where the results fall into the gray area.

In the following example, a survey is designed to determine whether the goal of 80 percent of households having soap is being met. The double sampling scheme chosen establishes an initial sample size (35); the number of “defectives,” or responses not meeting the criterion, permitted (10, in this case); a size for the second sample (70); and the maximum cumulative

number of defectives permitted for acceptance (26). The full two-stage sample is statistically equivalent to a single LQAS sample of around 50.

In surveying the householders' use of soap, investigators would check the initial sample of 35, allowing for up to 10 "defectives." If the first sample reveals 10 or fewer households without soap, conditions are declared to be satisfactory and sampling is discontinued. If more than 26 households have no soap, sampling will also be discontinued because standards (80-percent usage) have clearly not been met. If, however, 11 to 26 households have no soap, the second sample of 70 is tested and acceptance will depend upon finding a cumulative total of 26 or fewer of the entire 105 study households without soap.

Advantage: Double sampling is a way to gather more precise information than can be obtained with lot quality assurance sampling, without having to increase sample sizes across the board. That is, sampling is repeated and the sample size doubled only when the initial results do not appear clearly satisfactory or unsatisfactory.

Disadvantage: A double or sequential sampling scheme is administratively more complex than a single-sample plan.

Reduced and Tightened Inspection

Like double sampling, reduced and tightened inspection keeps the sample size as small as possible. However, instead of re-examining units that fall below a certain standard, this method either tightens or loosens the level of scrutiny while the assessment is still taking place. As the results are received, the sample size and number of defectives allowed are being adjusted accordingly.

Because of its small samples, this method (again like double sampling) provides only a small amount of information. It would be useful when policymakers wanted to judge whether certain promotional programs—latrine building or oral rehydration salts (ORS) use, for example—have been successful in reaching their targets.

Based on a predetermined acceptability level (85 percent, perhaps), a sampling plan is designed that specifies sample size and number of defectives allowed for each subunit. As the survey moves from subunit to subunit, the number of defectives in each is determined, and a running average calculated concurrently. If any subunit falls below the acceptability level or the average level drops below this level, tightened inspection begins; the same sample size is maintained in each subunit, but fewer defectives are allowed for a subunit to be classified as "acceptable." If fewer than the new target number meet the standard, the survey is stopped because it is assumed that the program has not met the coverage goal. If none of the subunits falls below the new standard, tightened inspection continues until the overall average reaches the initial acceptability level, i.e., 85 percent in this example. At this point, normal inspection is resumed.

If, however, the overall average coverage were to rise above the initial acceptability level, reduced inspection begins. For each subunit, both the sample size and the number of

households that must meet the standard are reduced. Chapter 8 presents a thorough explanation of this method as it might be applied to a study of ORS usage. Reduced sampling continues as long as each subunit and the overall average meet the criteria of acceptability.

Advantage: Because of its continuous monitoring of results, this method presents an objective appraisal of informational needs and thus serves the cause of efficiency in data collection.

Disadvantage: Its principal drawback is that the prescribed adjustment in sample size will be valid only when recent past experience is relevant for current decisionmaking. Therefore, when conditions are erratic or unpredictable, this method offers no advantages.

Epidemiological Surveillance

With this method, observers systematically collect, consolidate, and evaluate data to assess conditions at a particular time or to monitor changing conditions. This type of surveillance allows observers to detect trends as they develop and to identify problems as they arise. Such a method might be used, for example, to study seasonal patterns of diarrhea, disease distribution throughout outbreaks, and trends in the use of ORT. Information about episodes and actions taken would come directly from women of child-bearing age in specified communities. These communities would be nonrandomly selected according to the conditions and circumstances of interest to the observers.

Surveillance methods are diverse, but the principal distinction is between active and passive systems. Active systems employ health workers to collect the information; passive systems rely upon the individuals involved to report events as they occur. (Not surprisingly, passive systems commonly suffer from dramatic underreporting.)

Although it has sometimes been adapted to suit broader management concerns, epidemiological surveillance lends itself particularly well to monitoring communicable diseases. (See Chapter 9 for a description of epidemiological surveillance as a tool for controlling guinea worm disease.) This method is particularly useful in peri-urban settings, where communicable diseases are a significant danger and rapid changes in incidence and distribution must be detected promptly.

Advantage: Because periodic reports are linked over time, the amount of information required for each report is minimal and can usually be collected by individuals routinely employed for other purposes. Moreover, the close control typical of sentinel sites will tend to yield high quality data.

Disadvantages: Passive surveillance often results in serious underreporting, yet active surveillance can be costly unless limited to a few sites. However, with repeated close attention over time, the active sentinel sites may not remain sufficiently representative of the population as a whole, limiting the extent to which findings can be generalized.

Demographic Surveillance

Demographic surveillance, the monitoring of vital events (births, deaths, marriages, divorces, immigration, and emigration) is a particularly difficult process in developing countries. Trends for these, although often monitored, are also often underreported because people rarely remember all of these events—particularly early infant deaths. Unfortunately, public health records may be of little help to investigators, as many of these events are unattended by health personnel.

Demographers have, however, devised a number of demographic techniques useful for rapid assessment of births and deaths: three of these (described in Chapter 10) are to obtain pregnancy histories, establish routine encounters at the time of delivery or vaccination, and monitor the survival of close relatives of the respondent.

Advantage: With this method, the number of vital events necessary for analytical purposes is obtained from a sample of manageable size.

Disadvantage: Because of its dependence upon recall of past events, this method suffers from underreporting (although some informants overreport, as well). Such underreporting is particularly severe in cases of early infant deaths, as both birth and death may be overlooked.

Industrial Process Control Methods

Although applicable to health surveillance, control charts—one technique for monitoring industrial processes—have rarely been used in the health field to date. Such charts present a visual representation of a particular condition or circumstance in a given area. By means of preestablished control limits placed upon the chart, observers can determine whether a problem exists and, if so, how serious it appears.

A control chart would be useful for such activities as monitoring the incidence of chronic diarrhea in a particular area because the number of cases plotted inside and outside the control limits would delineate the nature of each outbreak. Each week the chart could be updated to show new cases and provide visual evidence of developing trends.

Chapter 11 provides a comprehensive discussion of the various types of process control methods; each features frequent small samples to signal trouble promptly, calculate simple indicators from the findings, and visually display the results.

Advantages: Industrial process control methods are rapid and easy to administer. Because the amount of data needed at each inspection is minimal, indicators can be calculated quickly and problems and trends identified easily. Visual charts make it easy to compare communities or service units.

Disadvantage: Because flow charts are intended for use after a process has become relatively stable, their use is inadvisable when conditions are erratic—for example, in the early stages of a program.

Case-Control Analysis

This method, which begins with the selection of readily available “cases” and “controls,” is useful for testing a hypothesis about factors associated with a particular condition or circumstance. The cases need not refer to an illness but may be users of a certain type of service. Controls are people who are similar in certain respects—age and gender, perhaps—but who differ in their lack of the condition that distinguishes a case. In a health study that used this method, the cases might be small children with persistent diarrhea whose mothers brought them to a particular hospital for treatment; the controls could be children within the same age group who were being treated at the facility for injuries. The two groups could then be compared regarding their usual source of drinking water to determine whether it played a role in the diarrhea of the cases.

Its economy and ease of use make case-control analysis an attractive investigative method for certain purposes: for example, it was used in the Philippines and Malawi to assess the connection between diarrhea and water use; in India to determine the influence of severe diarrheal disease and/or heatstroke on the development of blinding cataracts; and in Colombia and Haiti to study the use of health services.

Advantages: Because of the ready availability of cases and controls at health facilities, such studies tend to cost far less to conduct than do those that must go out into the community to find their cases and controls, particularly where the condition of interest, for example, chronic diarrhea, is relatively rare and therefore difficult and costly to identify. The case-control approach may represent the only feasible means of obtaining, quickly and cheaply, enough cases of an uncommon condition that takes a long time to manifest itself. Moreover, ethical considerations argue against a prospective study in which such a condition is allowed to develop unchecked within the general population to the point of discernment.

Disadvantages: Readily available cases, as well as controls, may be unrepresentative of those in the population at large. Odds ratios used to approximate relative risks tend to overestimate levels of risk, and sample sizes are generally too small to isolate important interactions and confounding factors that may be present. See Chapter 12 for a discussion of the various ways to avoid such distortions.

Sociocultural Group Assessment Methods

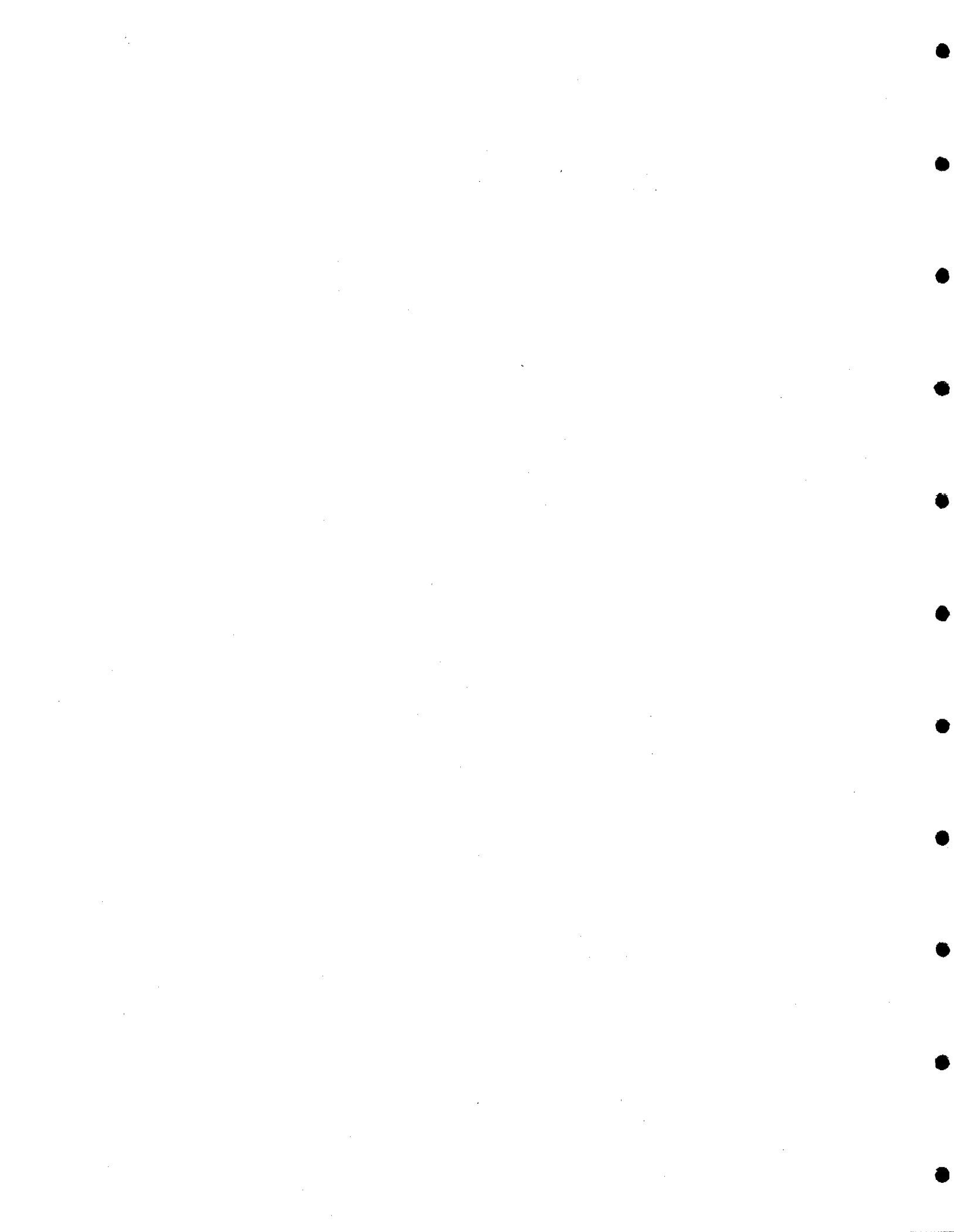
Although a comprehensive review of such methods is beyond the scope of this document, sociocultural group assessment methods are extremely useful at certain times and under certain circumstances and should not be overlooked as a methodological choice. Such methods come into play when the aim is not to collect and analyze objective information but to explore basic attitudes, beliefs, or desires of a particular group or even a culture as a whole. To illustrate, what are the factors behind the local practice of applying cow dung to the umbilical cord during childbirth? Or what are the cultural beliefs underlying certain practices that, although not recommended in the presence of diarrhea, people continue to carry out?

At other times a somewhat different perspective is taken in which individual views are combined to gauge group consensus, as opposed to simply averaging individual judgments. Such an application would be useful in determining community consensus on the issues of tubewell or latrine placement, for example.

Investigative techniques might take the form of a survey of representative individuals or perhaps the intensive interrogation of a few key persons who thoroughly understand the local culture and can offer insight into their neighbors' behavior. Sometimes the most useful information might come from a purposive group, such as a mothers' club, rather than from the same number of persons randomly contacted individually. The primary distinguishing feature of sociocultural methods is their emphasis on *why* as opposed to *how many*.

Sociocultural methods are often used to obtain information on beliefs and perceptions about community attitudes, for example, or uncovering hidden hygiene or sanitation practices and the reasons behind them. Such information can be vital to the success of water supply and sanitation projects.

All the methods mentioned in this chapter have useful applications; the challenge facing managers and other decision makers is to fit the appropriate method with the task at hand. Chapter 4 discusses issues to consider when defining informational needs and selecting a method to collect and analyze that information.



Chapter 3

STATISTICAL PRINCIPLES

Certain principles and limitations relating to statistical analyses in general are worth highlighting before moving ahead to selection issues. First, *results obtained from a representative sample are only estimates of conditions in the target population and are, therefore, necessarily imprecise*. The level of imprecision depends on two factors: (1) the extent to which the measure of interest (a physical characteristic, for example) varies among individuals in the population, and (2) the number of individuals sampled. If every individual in a community were exactly the same height, a sample that included only one person would provide an unerring estimate of average height in that community. Because individuals do vary in height, however, a small sample that may by chance include mostly short or tall folk is unlikely to produce an accurate picture of the total population. This deficiency can be overcome only by selecting a larger, and presumably more representative, sample.

Although variability is an inherent population characteristic over which investigators have little control, they may employ certain techniques to reduce variability within their samples. One of these is to define a more homogeneous population: for example, a group of families taking their water from a protected well will likely be more similar in diarrheal incidence than will a population that includes those who use a variety of sources. By separating the groups according to water source, interviewers can survey each homogeneous group separately and can likely produce a clearer estimate of incidence for each group than could be derived from a more diverse population.

Discrete Measurement vs. the Continuous Scale

Another way to manipulate variability is to distinguish between *discrete*, or categorical, variables and those measured on a *continuous scale*. Illustrating the former would be the classification of children according to whether they have persistent (lasting more than 14 days) diarrhea. The indicator of interest then would be the *number of children* having persistent diarrhea. By contrast, a measurement of the duration of *each episode* would form a continuous variable that directed attention to the average duration.

Often a categorical measure, such as the presence or absence of disease, is the only feasible alternative; at other times, however, a choice does exist between the two approaches. In looking at the duration of an illness episode, for example, we might record the actual number of days or simply indicate whether the duration exceeded 14 days. The latter approach may be simpler, but it is likely to demand a much larger sample to achieve a desired level of precision because the gross distinctions associated with discrete measures produce limited information from each case. Continuous measures, to the contrary, provide more information per individual, since shades of difference are recorded.

The principle that emerges here is this: *Sampling efficiency is served by the use of continuous measures whenever they are readily obtained, follow a regular pattern (e.g., a statistically normal distribution), and satisfy the specific analytical needs for the decision at hand.* Knowledge of average illness duration would be useful in assessing the total impact of disease on normal activity; concern for the magnitude of the problem of persistent diarrhea would require only that cases lasting more than 14 days be identified.

Sample Size

We have seen that investigators can increase precision by forming homogeneous groups and/or by using continuous variables. A third technique calls for the selection of larger samples, since assessment methods that rely on small samples restrict the ability to make decisions based on fine distinctions. For example, small samples distinguish between an 80-percent level of service coverage and one of 70 percent, but would rarely be able to differentiate between 80- and 75-percent levels. Stated more directly, *small samples yield no more than gross distinctions.*

The advantage of homogeneous groupings previously cited calls attention to the benefits of a stratification that divides the population into distinct subgroups, such as urban and rural or differing socioeconomic levels, in order to sample each group separately. Indeed, stratification is necessary if one is to assess the differential effects of literacy, poverty, or exposure to a potentially hazardous environment.

Subgroup analysis, however, forces recognition of another important statistical principle: *The size of the sample needed to estimate the prevalence of a specific condition in a target population has little to do with the population size; instead, it depends upon the level of precision required* (Reinke, 1988, pp. 17-18). We need not dwell upon the rather complex formula for sample-size determination, showing the relatively slight influence of population size. Of greater practical interest are the illustrative results achieved from the application of that formula. Consider, for example, a sample of 50 taken from a population of 5,000 for purposes of estimating the rate of latrine use in a given area. It is determined, by using the appropriate formula, that the estimate obtained could be in error by as much as 14 percent. If for some reason 10 groups of 500 were identified and estimates of similar precision were sought for each group separately, the formula indicates that samples of 46 from each group (460 in all) would be required. Although the new populations are one-tenth the size of the original, the new sample sizes required are 92 percent ($46/50$) as large as the original. Moreover, the total amount of sample has been increased more than ninefold to 460.

The relative unimportance of population size, as opposed to *number* of samples, leads to a corollary principle: *The most efficient way to estimate the difference between two groups is to obtain the same amount of information from each group, regardless of disparities in group size.*

Since sample size depends above all else on the level of precision desired, the precision factor deserves closer scrutiny in relation to the parameter being estimated. Specifically, if interest centers on a level of latrine use that is judged in advance to be about 50 percent—in a

shantytown, perhaps—a final estimate that errs by as much as 5 percent would probably be considered acceptable. However, an error of this magnitude in the estimate of a relatively rare event would be intolerable. For example, an investigation of schistosomiasis in Zimbabwe, where roughly one-tenth of the school children are affected, could yield a sample prevalence of between 5 and 15 percent—a threefold range of possibilities. The point here is that whereas estimates are most imprecise for prevalence rates of 50 percent, under the circumstances sample size matters less because the need for precision is less urgent.

In practice, the tolerable error in estimating the prevalence of a specified condition is likely to be a roughly constant fraction (say, 20 percent) of that prevalence. Thus, if an investigator expects to find that approximately half the community members use latrines, she is likely to be satisfied with an estimate that may be in error by up to 10 percent (20 percent of 50); the same analyst, however, would be unwilling to tolerate an error above 3 percent (20 percent of 15) if she were surveying the prevalence of schistosomiasis in Zimbabwe.

In the following discussion, 95 percent of the prevalence, or rarity (R), estimates made of specified conditions will maintain a relative error of less than 20 percent of the true prevalence. Following this rule, the requisite sample size (n) is determined to be $n = 100R$ (a benchmark formula).

Rarity is defined as the ratio of noncases to cases. Thus, if one child in five has diarrhea (so that four of five are free of the condition) the rarity factor is four. Assuming that six “well” children are likely to be found for every child with diarrhea, R becomes 6 and a sample of 600 is thus called for under the formula. In contrast, if only 1 child in 20 has persistent diarrhea (19 children free of the problem), a sample of 1,900 children is needed to identify enough cases to carry out meaningful analysis. Consistent application of a single set of decision criteria enables us to consider a number of situations of interest without having to give undue attention to the arithmetic involved.¹

Specifying Precision

Examination of the same formula reveals one feature of sample estimation that deserves underscoring: a *doubling* of the desired level of precision requires a *quadrupling* of the requisite sample size. In the preceding example, a sample of 600 was considered large enough to obtain an estimate of diarrhea prevalence that would fall within 20 percent of the true rate. In order to reduce the maximum error to 10 percent, a sample of 2,400 would be needed.

Examples given thus far have dealt with a single, relatively simple issue: how to estimate with specified precision the prevalence of a stated condition within a defined target population. As might be expected, answers to more searching questions demand that more information be

¹ The mathematically oriented reader seeking greater understanding of the formula used for sample-size determination and a full range of applications stemming from it should refer to the detailed description in the mathematical appendix accompanying this report.

gathered. Thus far, we have dealt only with single populations and with estimates of existing conditions in the absence of any preconceptions of what is expected or desired. Suppose, however, that concern were expressed over possible differences between groups or that preconceived norms were available against which to judge sample findings. From a statistical standpoint, decision makers are nearly always in one of these four situations. The associated possibilities produce four scenarios that vary in complexity and in the corresponding need for information.

For the simplest case, involving a single population and no prior hypothesis regarding conditions (e.g., usage rate of latrines) in that population, it has already been determined that a sample size of 100R will enable the decision maker to make adequately precise judgments about matters of interest within the population. But if similar precision were sought in comparing the status of *two* populations—their use of ORT, for example—the sample size would have to increase because now there would be two sources of imprecision. Specifically, the sample for each community must be doubled (to 200R). Furthermore, because the larger sample size applies to *each* of the two populations, 400R observations are required in all. This more demanding case is common, for investigators often seek to compare a population exposed to a health hazard or to a certain intervention with a second group that has had no such exposure.

In the preceding two scenarios, no prior standards had been set. The final pair of circumstances, however, involve preconceived notions of “satisfactory” and “unsatisfactory” states. As a result, enough information must be obtained to distinguish between them. Suppose, for example, the effect of an improved water supply on diarrheal incidence is to be examined. If a 20-percent reduction (to retain the same relative difference used in earlier examples) from an established prior level is considered satisfactory and no effect whatsoever is deemed unsatisfactory, the consequent sample size for classifying one population turns out to be 215R.²

Thus, sample size must increase when differences between two populations are compared and when benchmark standards of satisfactory and unsatisfactory performance have already been established. It is not surprising, therefore, that the most stringent demands for information occur when actual differences between two populations are to be compared with prior expectations. For example, a program of education may be expected to have a greater effect on ORT use in a community served by outreach than in one served only by fixed health facilities. Comparison of the two sets of results would require 430R observations in each community.

Comparative sample sizes for the four sets of circumstances (see Figure 1) demonstrate that, in the face of prior standards, definitive testing for differences between two populations

² This figure is based on a 10-percent risk of concluding erroneously that there is an effect and a 5-percent risk of failing to detect an effect. For the sake of comparability, these risk values will stand throughout the illustrations presented. Other risk specifications would alter the required sample size.

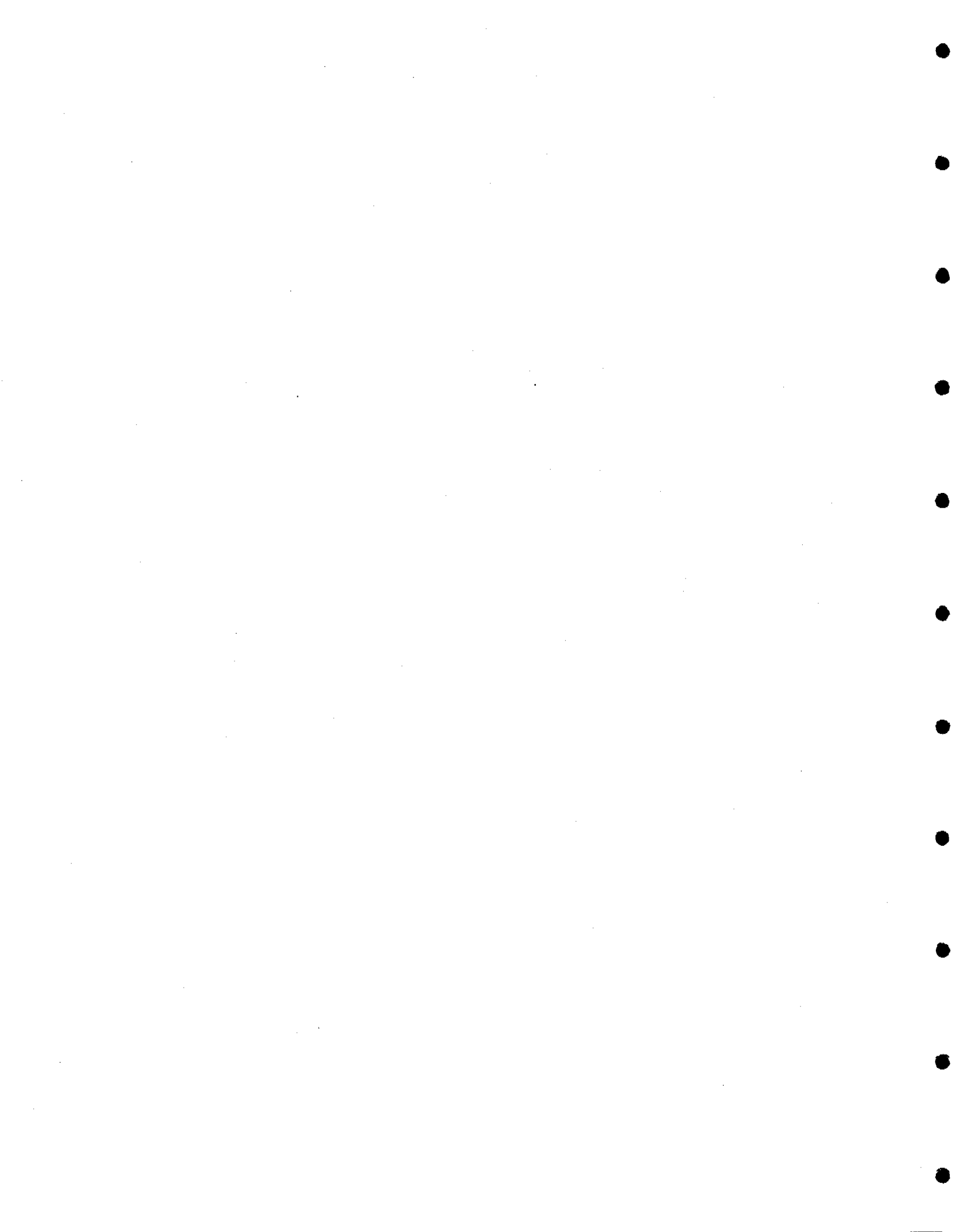
requires more than eight times as much data as is needed for a simple estimate of conditions in a single population. Clearly, information gathered for the latter limited purpose can lead to dangerously misleading conclusions when used in more complex analyses.

Circumstances	Nr. of Observations	
	Each Pop.	Total
One population; no prior standards	100R*	100R
Two populations; no prior standards	200R	400R
One population; prior standards	215R	215R
Two populations; prior standards	430R	860R

*R = "rarity," i.e., the ratio of noncases to cases
OBSERVE: Testing for the significance of differences between populations, given prior standards, requires more than eight times as much data as are needed for a simple estimate of conditions in a single population.

Figure 1

Comparative Sample Sizes



Chapter 4

SELECTING RAPID ASSESSMENT METHODS

Faced with a number of viable assessment methods, how are decision makers to choose among them? How can they avoid the pitfalls of data collection that may burden them with excessive data while yielding too little that is useful? One way is to base selection on the nature of the decisions to be made. In principle this is obvious, but in practice data are often collected without a clear picture of how the information will be used. Careful appraisal of the link between information and decisions will lead to more effective application of assessment methods.

Vital to an informed selection are objective criteria that will allow people to weigh the benefits of each method, freeing them from the need to either take a vast leap of faith or depend upon a familiar, but possibly inappropriate, procedure. As a tool to aid those who must choose from among competing methodologies, this chapter discusses some issues connected with selection.

Informational Issues

As a preliminary step in choosing an assessment methodology, managers and other decision makers must consider the type of information needed, its usefulness in decisionmaking, and the quantity and quality of information needed. It may be useful to ask some of the following questions.

What kind of information is needed and how will it be used?

The type of information to be collected will vary according to its intended use, which determines the *object of measurement*: whether illness or positive health will be the focus of attention, for example. Ideally, an index of health would be more meaningful than a measure of illness, but the dimensions of well-being are poorly defined, and in practice health programs tend to emphasize mortality and morbidity reduction over health promotion.

Will decision makers be better served by quantitative or qualitative data? Surveys recording the magnitude of a condition (duration of a diarrhea episode, perhaps) require smaller sample sizes than those collecting categorical data (presence or absence of diarrhea). Methods that facilitate quantification, therefore, have an advantage. On the other hand, certain important characteristics, notably cultural traits, are impossible to quantify. Methods that succeed in capturing these qualitative features in a meaningful way are therefore to be valued. Moreover, categorical data may more appropriately address certain policy issues: program managers may have less interest in the average duration of diarrhea episodes than in the rate of diarrhea lasting 14 or more days.

It will also need to be determined whether data should be subjective or objective, cross-sectional or longitudinal, and whether they should describe individuals or groups. Methods using groups as the basis for information gathering tend to be more rapid and less costly than those requiring individual contacts, but may produce results that are not truly representative of the populations involved. The applicability of such methods is limited to situations in which group consensus is desired, and then the negative effects of interactions among group members should be minimized.

Who will be the sources of the needed information? For a health survey, investigators will likely choose between consumers or health providers. In general, rapid assessment favors providers, for they already maintain records on persons with problems of interest. Because they may not present a true reflection of the target population, however, consideration must be given to the possible trade-off between speed and convenience on the one hand and validity on the other. For example, provider data on patients may give an unrealistically low estimate of diarrhea rates in the community and may, in addition, give a distorted view of the distribution of the problem by age, sex, and socioeconomic status.

Will action be taken on the basis of overall average conditions or only with respect to specific areas in which conditions are deemed unsatisfactory? In the former circumstance, care must be taken to ensure that subgroups and geographic areas in the target population are similar enough to make aggregation meaningful. Thus, if a hygiene program for a large and stable population of several thousand were under consideration, individual clusters within the total study sample should be as heterogeneous as possible—serving as microcosms of the total population.

Conversely, if decisions will affect individual subgroups, each of them should be as distinct and homogeneous as possible to minimize sampling error and the consequent amount of information needed to make appropriate decisions. Such considerations would be applicable if several diverse squatter settlements within a metropolitan area were being considered for individually tailored hygiene interventions.

Another concern in decisionmaking is whether action is to be taken with respect to a static condition or during the course of monitoring an ongoing process. Assessment methods applicable to the two circumstances differ; in particular, process monitoring relies on frequent collection of small samples to detect trends. Certain techniques borrowed from the field of industrial quality control (such as the control chart) are useful for process monitoring, as they stress the use of information to improve the overall process average over time (Berwick, 1989).

A numerical example shows the merit of this approach. Suppose that performance is normally distributed around an average performance score of 100. A traditional manager might monitor the process to identify the lowest 10 percent of performance and attempt to bring the laggards up to standard. Another manager, however, determining through statistical analysis that this goal would improve average performance by less than 2 percent, may choose a more general effort to improve overall performance by at least 2 percent.

The underlying principle is that process-sampling procedures can be most effective when used to monitor a process that is already in control (Mosteller, 1987) and to orient process management toward general improvement and achievement of greater uniformity.

How will the information be useful for decisionmaking?

Of overriding importance in methodology selection is whether the information it produces will be useful in decisionmaking. Can the method pinpoint the site and the nature of the problem? Can it generate information specific enough to isolate the problem so that appropriate action can be taken?

To be useful in decisionmaking, the information should be applicable at all levels where action is to be taken. Specifically, it should be useful in self-diagnosis, for those in the field may be in the best position to identify problems as they arise and take timely corrective action on their own without the need for outside intervention. Consider the example of a hygiene education program designed to increase hand washing. An assessment procedure that allows the project manager and hygiene educator to determine immediately if the program is "on track" provides much more useful information than a procedure that requires complex analysis, requires outside expertise, and is likely to be delayed in providing the necessary feedback. Thus, simplicity in the interpretation of information is as important as simplicity in data gathering.

How much information is needed and at what level of quality?

The quantity of information to be collected will vary according to the following:

- Whether the decision will be based on an estimate from one group or on a comparison, either with a standard measure, a previous study, or another group studied at the same time.
- The precision necessary. Are errors of omission or those of commission more serious, and what degree of risk is allowable for either error?
- The heterogeneity of the population with respect to the characteristic of interest.
- How much information must be collected from each individual or sampling unit.

The amount of information collected should meet (but not exceed) the minimum necessary to provide sufficient data to the decisionmaking process.

The quality of information relates closely to the quality of decisionmaking, for the correctness of the decision is a function of the reliability and validity (accuracy) of the information that triggers the action. To the extent that data are unreliable and inaccurate, erroneous decisions of two types are possible. *Errors of omission* occur with failure to take needed action, such as when an outbreak of cholera goes unrecognized. *Errors of commission* occur when action is taken unnecessarily. A sample of households, for example, might include by chance a

disproportionate share of children with diarrhea, thereby leading to the misleading conclusion that the incidence of diarrhea in the community at large had increased.

Perfect reliability and validity can be achieved only with complete knowledge of a situation, and this is usually impossibly costly. As a result, errors of omission and commission can seldom be eliminated, but carefully designed procedures for data collection and analysis can limit their risk of occurrence. Errors of omission tend to be more serious than errors of commission, for failure to correct faulty conditions is generally less desirable than action to improve an already satisfactory situation. In certain areas where resources are exceedingly limited, however, the misallocation of resources associated with errors of commission may be the overriding consideration. In practice, therefore, the ramifications of decision errors must be judged on a case-by-case basis.

Methodological Characteristics

Once having satisfied their concerns regarding the nature of the information required, decision makers will need to turn their attention to the methodology itself. Two of its features are particularly important to assess: ease of use and the timeliness of the information the prospective method can produce. The following two sections discuss these elements.

Is the method easy to use (or misuse)?

Methods of potential usefulness in decisionmaking can realize their full potential only if they are readily applicable in practice. Hence, ease of use is another characteristic of interest, relating to the processing, analysis, and interpretation of information, as well as to the initial data gathering.

Unfortunately, simple methods that are easy to use present a trap for the unwary, who may adopt them without understanding their limitations. The ramifications can be serious because the sheer simplicity of the method is likely to impose serious limitations on its use. For this reason, another important characteristic to evaluate is a method's *potential for misuse*.

Will the method yield timely information?

Because the expression "time is money" is more than a slogan, information timeliness represents an extension of the cost-effectiveness factor discussed in the next section. Perhaps good information, if available, could lead to a good decision; if the requisite data are not provided in time for the action to be effective, however, the assessment process is obviously not cost-effective. As a rule of thumb, it is felt that rapid assessment methods should produce useful information for decisionmaking in a matter of weeks, or at most a few months after the need for information has been established.

Cost Considerations

In a world of diminishing resources, few people would select an assessment method without giving some thought to its costs. Decision makers will need to consider not only the total outlay, but also whether the funds are well spent. In other words, how much will it cost to conduct the methodology and will the results justify the price? The final sections of this chapter discuss these two questions.

Is the method cost-effective?

Once the basis for action has been defined, attention shifts to the cost of acquiring the information needed to make the correct decision. Quality often can be purchased for a price; however, the time, effort, and cost of obtaining information should not exceed the benefit derived from the course of action consequently chosen. Inevitably, the manager faces risks of omission and commission and must balance the two. For example, faced with very limited resources, managers may try to detect only the most severe illnesses or the most poorly functioning health centers, even though their failure to detect lesser cases represents errors of omission.

Another question of quality as it relates to cost comes with consideration of whether to consult a few knowledgeable individuals or a larger, more representative sample. The latter is usually recommended in order to minimize bias, although a few key informants can shed considerable light on complex qualitative issues involving group attitudes. Moreover, bias itself is not viewed with uniform displeasure. To illustrate, a survey might be conducted among easily reached families living fairly near a health facility, even though the findings will likely be unrepresentative—the rationale being that if those with ready access to services fail to take advantage of them, there must be a more general problem of usage as well.

The inevitable tradeoff between cost and quality forces a consideration of both factors. Despite some suggestions to the contrary, there is no way in which statistical laws of probability affecting the magnitude of sampling error can be repealed: small samples produce relatively crude estimates of population conditions. Because analysis of significant departures from hypothetical norms requires more information than for estimation purposes without preconceived norms, small samples can be expected to detect only gross departures from the norms. In general, cost savings must be gained by means other than reducing sample sizes (Briscoe, 1988) or, alternatively, rapid assessments based upon limited information must have modest aims in terms of expected precision, risk of error, and complexity of analysis.

Use of more sophisticated sampling procedures can limit acquisition of more data to those occasions when it is really needed, thereby reducing the number of observations on the average. With double sampling, for example, a small initial sample could identify extreme conditions, with the additional cost of a second sample incurred only when necessary (see Chapter 7). Well-defined procedures for normal, reduced, and tightened inspection likewise focus data gathering on the most problematic circumstances (see Chapter 8). Finally, procedures for linking sets of data collected periodically (through control charts, for example)

can depict trends where they exist; where stable conditions prevail, results from the separate samples can be aggregated to produce estimates as good as those available from large samples. Other approaches to sample-size reduction rely upon corresponding reduction in variation among individuals in the population, as discussed in Chapter 3.

What resources are available for the task?

Before selecting a method, it is important to have a clear picture of the available resources. Not only money is significant at this juncture, but also time, personnel, and—sometimes most critically—skills and level of expertise. For example, who will handle design, data collection and supervision, data analysis, interpretation, and presentation? Generally, managers carry out a balancing process that determines the information desired, estimates the resources needed to collect the information, compares this estimate to the available resources, reduces the quantity and/or changes the type of information to be collected, or attempts to increase the resources until those available can accommodate the task.

Chapter 5

CLUSTER SAMPLING

Simple Random Sampling vs. Cluster Sampling

When an assessment is planned, one of the first decisions is how to select a sample of the target group. Several methods are available, but the most straightforward is simple random sampling. This method involves randomly selecting some proportion of all final sampling units (families, households, adults, heads of households, women, as the case may be) for study. However, simple random sampling has clear disadvantages that virtually rule it out as a rapid assessment method. First, analysts must have access to a complete list and detailed map of all final sampling units in the study area. Such lists are rare in developing countries and rarer still in peri-urban areas and, in most cases, would be difficult to compile. Second, in a randomly selected sample, sampling units will likely be found in many communities or jurisdictions, each of which must be contacted and dealt with before the assessment can begin. If the target area is large, this can be time-consuming and wasteful.

As the method yielding the most precise results, simple random sampling can be used as a standard for judging the precision of other sampling methods.

Cluster sampling is a practical alternative to random sampling. In cluster sampling, the sample is selected not from a list of all the final sampling units (*individual respondents*), but from a list of groups (or *clusters*). It is a "hierarchical type of sampling in which the elementary units are often at least two steps removed from the original sampling of clusters" (Lemeshow and Robinson, 1985). Although not generally yielding results as valid as those from simple random sampling, it is a much less expensive and more rapid alternative and is widely used.

Description of the Cluster Sampling Method

In cluster sampling, the study area is first divided geographically into primary sampling units—districts or counties, for example. Certain of these are selected randomly. At the next stage, the units selected are further divided geographically and some of these subunits are chosen randomly. The process is continued, stage by stage, until the lowest subunits—communities, city blocks, or squatter settlements—are identified and individual respondents (*final sampling units*) within each are selected and contacted. In this procedure the time-consuming task of detailed listing and mapping is restricted only to the few subunits, or clusters, that are finally chosen. Mapping these final clusters must be done carefully or the principle of randomization can be compromised and the results distorted accordingly.

If cluster sampling is done properly, each member of the target population, or final sampling unit, has at the outset an equal chance of being selected. This condition will be fulfilled if subunits are chosen with probability proportional to size. Thus, a district with twice as many people as another would be twice as likely to be chosen at the first stage. At the final stage,

each subunit must contribute the same number of final sampling units. If 30 subunits are chosen, 7 final sampling units might be selected from each to produce a total sample of 210.

Representative Uses of Cluster Sampling

Cluster sampling could be used, for example, to assess hand-washing practices in a large peri-urban area comprising many neighborhoods. A predetermined number of these neighborhoods (clusters) would be selected randomly, and only in the neighborhoods selected would it be necessary to make detailed survey arrangements.

Operational and Analytical Difficulties

The operational and analytical difficulties of cluster sampling relate to three factors: the magnitude of differences among subunits (communities or neighborhoods), the pattern of those differences, and procedures for selecting final sampling units (usually persons or households).

Balancing Variation between and within Subunits

Estimation errors are a function of the variation among individual observations, but the effects of variation tend to be averaged out over a series of observations. In cluster sampling, such variation occurs between subunits as well as within them. The larger the number of clusters and the greater the heterogeneity within the clusters, the more the sample variance will tend to approach that of a simple random sample.

However, the main purpose of cluster sampling is to reduce as far as possible the number of subunits that researchers must go into to obtain information, for when communities are entered, community leaders must be informed about the assessment and their collaboration must be sought. For that reason, researchers normally opt for more observations in fewer subunits. Thus in designing a cluster sample, it is necessary to balance desired within-subunit and between-subunit variance.

To illustrate, if a sample size of 100 is decided upon, there are many options as to the distribution of subunits and final sampling units. Deciding upon 20 observations in each of 5 subunits would reduce within-subunit variation, but between-subunit variation would remain large. Specifically, between-subunit variance would be reduced only by a factor of 5. Alternatively, if an individual from each of 100 different groups were chosen, between-subunit variance would be reduced by a factor of 100, and the cluster sample would approximate a simple random sample. While the latter procedure has statistical merit, it could prove to be very costly to enter so many subunits. Thus, the number of observations per cluster represents a compromise between the ideal of one observation per cluster and a logistically simpler, but statistically inefficient large number of observations per cluster.

Design Effect

The ratio of the variance obtained from cluster sampling to the variance that would be derived from simple random sampling for the same population is known as the *design effect*.

For a given level of precision desired, the higher the design effect, the larger the sample must be to compensate. If a design effect were 1, then analysts would know that the cluster sampling methods used were as efficient as simple random sampling; the same size sample would suffice for both.

In practice, the design effect depends on the magnitude of the variance between clusters relative to the variance within clusters. For a given sample size, the more homogeneous the population within the individual clusters, the greater the design effect; the variance between clusters can be reduced by increasing the number of clusters selected. To illustrate with an extreme example, if a survey of access to piped water used cluster sample methods, and if 30 clusters of 7 homes each were selected, the actual sample would really be close to 30 rather than 210 because most of the homes in each cluster would have the same access. Households within a cluster would be relatively homogeneous because piped water supplies are generally a communitywide intervention, and the resulting design effect would be high. Obviously, this high design effect means that the sample size should be increased by increasing the number of clusters selected, rather than increasing the number of homes selected per cluster.

To illustrate further, suppose that analysts are assessing immunization coverage in a certain area. Suppose also that they expect to find that roughly half of the target population is immunized. Using the benchmark formula from Chapter 3 for the size of a random sample (i.e., $n = 100R$ where R is the ratio of noncases to cases), they calculate that $R = 1$ and $n = 100$. If cluster sampling would produce a design effect of 2, the sample size would have to be increased to 200 in order to retain the level of precision available from a simple random sample of 100 children. If it is also determined that 30 subunits in the sample will provide adequate representation of between-subunit variation, the design would call for 7 individuals (final sampling units) from each cluster.

This is essentially the line of reasoning followed by the World Health Organization in establishing the "thirty-clusters-of-seven design" for determining coverage in the Expanded Program of Immunization (EPI). Although the EPI design has been widely followed, its underlying assumption regarding the design effect of 2 should be questioned for each application. *EPI experience suggests that variance may be higher in areas of low coverage, although a design effect of 2 appears reasonable in many other circumstances* (Henderson and Sundaresan, 1982). Analysis of 37 surveys of disease incidence in India revealed a design effect well below 2 for relatively rare maladies in the target areas (tetanus and poliomyelitis), but design effects of 8 or more were found for high-prevalence conditions such as measles and pertussis (Rothenberg et al., 1985).

Heterogeneity of Clusters

Just as the design effect can be decreased by increasing the number of clusters, it can also be decreased by increasing the heterogeneity within each cluster. If, for example, specific ethnic or occupational groups tend to be confined to certain subunits, between-subunit variance will be relatively large. Wherever it is feasible to exercise control over the composition of subunits, each should be made as heterogeneous as possible so that it is reasonably representative of the entire universe of interest (Lemeshow and Stroh, 1988).

In cases where the logical subunits are fixed administrative entities of unknown heterogeneity, preliminary sampling of several clusters could provide an approximation of the design effect sufficient to prescribe the size and composition of the full-blown sample. Reinke (1988, pp. 25-27) suggests one such procedure.

Peri-urban Issues

Peri-urban areas may be characterized by homogeneity within subunits and heterogeneity between subunits. For example, one small neighborhood may be inhabited by migrants from a certain geographical area while a nearby tract is inhabited by those from a different region. In such cases, a large number of small clusters should be surveyed to reduce between-subunit variance. Indeed, each subunit might be made small enough to include all of its residents in the cluster.

Stratification

Where subunits are inherently dissimilar, it may be advisable to combine stratification with cluster sampling. Thus, certain ethnic groups or neighborhoods formed as a result of immigration within the past five years might be identified separately from more stable communities. Cluster sampling techniques could then be applied independently to the two strata or subpopulations.

Alternatively, the total sample size might be reduced by limiting the study to a relatively homogeneous population segment of special interest. Prior studies may have shown, for example, that shanty dwellers or those who share communal sources of water should be especially targeted for hygiene education. Sampling could be limited to these groups, with upper- and middle-class residents excluded.

Bimodal Distributions

Although the magnitude of between-subunit variability can affect the validity of findings, the pattern of variability can make those findings totally meaningless. For example, if one area is free of schistosomiasis while another of similar size exhibits a 60-percent prevalence rate, cluster sampling would produce an estimated rate of approximately 30. Depending on the

sample size and the design of the assessment, the estimate could be made extremely precise, but that would not make it any less misleading. This problem can occur whenever a factor related to the condition of interest is distributed nonrandomly in the population.

The large design effect cited earlier in connection with low immunization coverage rates and high disease prevalence is undoubtedly due in large part to between-subunit variation. Specifically, overall immunization rates can be affected dramatically by a few laggard communities, and a high disease prevalence rate can look like an epidemic. Under these circumstances, it is more important to identify the problem areas than to determine average conditions. For this reason, cluster sampling is unlikely to be informative enough in assessing a cholera epidemic. Other methods, discussed in subsequent chapters, are more appropriate in such situations.

Advantages and Disadvantages/Problems/Warnings

The main advantage of cluster sampling is that it provides a speedy, logistically simple, economical alternative to random sampling. This method makes the potentially difficult task of data gathering feasible and economical in the use of time and resources. As discussed above, however, cluster sampling should not be used, or should at least be adjusted, in cases where there are large differences among subunits or a pattern of variation among subunits that would cause the results to be misleading.

Because the method is intended to develop aggregate estimates, it should not be used to estimate conditions within individual subunits. For example, clusters in a peri-urban area might have widely disparate rates of diarrhea, but these rates should not be compared. Basic cluster sampling procedures are not designed for this purpose because the subunit sample sizes are inadequate. In this situation, sophisticated adaptive cluster sampling procedures (Thompson, 1990) might be employed. Whenever a selected cluster satisfies an exception condition of interest, such as very high incidence of diarrhea, additional clusters are added to the sample from the neighborhood of the first results. The adaptive procedures must be applied with extreme care, of course, because the nonrandom process of sample selection adds complexity to the analysis of findings in order to avoid bias.

Conclusion

Attracted by its simplicity and ease of application, decision makers are using the cluster sampling method in varied settings. Traditional EPI studies are reported from areas as diverse as Burma (Frerichs and Tar, 1988a), Peru (Jaramillo, 1989), and Austria (Weekly Epidemiology Record, 19 June 1987). A related study of immunizable disease has been carried out in Mozambique (Cutts, 1988). A typhoid outbreak in Taiwan has been investigated by means of cluster sample, as has child nutritional status in Burma (Frerichs and Tar, 1988b).

Although these studies have yielded findings of reported value in decisionmaking, there is little evidence of actual validation. One analysis of 446 sample survey estimates of service coverage

concluded that most estimates were probably in error by no more than 10 percent, the level of precision on which the 30×7 procedure is based (Henderson and Sundaresan, 1982). However, this conclusion assumes a normal distribution of between-subunit variation and does not test for bimodal and other irregular patterns mentioned here.

Chapter 6

LOT QUALITY ASSURANCE SAMPLING

Origins in Industry

Lot quality assurance sampling (LQAS) has its origins in industry, where managers needed an inexpensive and rapid method of assuring that a "lot," or batch, of manufactured items met a certain standard or passed a certain test. Developed at Bell Laboratories in the 1920s, lot quality assurance has only recently gained wide recognition in the health field, but it has quickly become popular because it is truly rapid, inexpensive, and easy to carry out (Dodge and Romig, 1959).

In LQAS, the sample can be kept very small because only large departures from a predetermined standard are identified (see Chapter 3). According to this method, if the small sample passes the test, it is assumed that the entire batch (or *lot*) passes. When used in epidemiological surveying, the lot is equivalent to the target group.

Representative Uses of LQAS

Lot quality assurance sampling can only be used in situations in which policymakers can get by with less information. For example, if policymakers wanted to test the success of a latrine promotion project by determining the percentage of people using latrines, this method would be inappropriate. In such a case, the sample would have to be large enough so that policymakers would know, within a reasonable margin of error, what percentage of coverage had been achieved in each health district. But if policymakers wanted to find out which health districts needed help with their latrine promotion programs, the lot quality assurance method could be used. In that case the information required could be reduced to a simple pass/fail test: those districts in which at least 70 percent use latrines pass; the rest fail. The cutoff percentage would be based on policymakers' estimate of what should be expected in the districts given the project activities, budget, and so on.

To use another example, LQAS would be appropriate in ascertaining whether a radio campaign to promote hand washing has reached a majority of the population. If, conversely, a common source of cholera has been identified in an outbreak and policymakers want to make everyone aware of it, the difference between 80- and 70-percent coverage might be important. In that case, an approach more sophisticated than LQAS would be needed.

Lot quality assurance is appropriate when managers or policymakers wish to identify target groups or subgroups with exceptional needs or service units that are performing at substandard levels. This enables them to focus attention where support is especially needed. The method has been adopted by public health decision makers needing information for three interrelated purposes:

- To give rapid feedback on existing conditions
- To identify in those conditions any departures from predetermined goals or standards
- To associate those departures with individual communities or service units

“Significance Testing” Character of LQAS

Cluster sampling is used to make a good estimate about a situation; lot quality assurance is used to test the significance of a departure from a standard.

The sample size required to detect a significant difference between a population and a standard is greater than that needed to estimate an overall condition. Technically, the distinction is between the first level of assessment ($n = 100R$, where R is the ratio between noncases and cases) and the third level ($n = 215R$), explained in Chapter 3. It is the distinction, for example, between simply estimating the percentage of households familiar with ORT ($n = 100R$) and determining whether the percent of households familiar with ORT is significantly less than the target level set by the project as “acceptable,” given the same risk of error ($n = 215R$).

To illustrate the differences in sample sizes required, suppose that 100 observations are needed to estimate within 10 percent the proportion of mothers in a community who are familiar with ORT. Suppose, too, that the population is to be classified as “satisfactory” when at least 75 percent of the households know about ORT, and “unsatisfactory” when fewer than 65 percent are knowledgeable, a difference of only 10 percent. Under these conditions, where there is only a 10 percent difference between “satisfactory” and “unsatisfactory,” a sample of 215, rather than 100, mothers would be required to make such a determination.

However, if the difference between what was considered satisfactory and what was considered unsatisfactory were widened, say from 10 to 35 percent, the sample size needed to reliably detect that large a difference would be much smaller. Lot quality assurance sampling is rapid not because of sampling efficiencies but because of the crudeness of the assessment.

Errors of Commission and Omission

With LQAS, it is very important to decide where to set the cutoff point between satisfactory or unsatisfactory conditions. If unsatisfactory conditions are mistakenly thought to exist in a community and corrective action is taken unnecessarily, an “error of commission” will be committed. If, however, unsatisfactory conditions remain undetected and the decision maker does nothing to correct the problem, an “error of omission” will be committed.

Decision Rules

The decision rule, or the establishment of the cutoff point or pass/fail criteria, can control the likelihood of committing either type of error. However, once the sample size has been set, a decision rule that reduces the risk of one type of error necessarily increases the possibility for the other type to occur. Therefore, the decision maker must determine the relative seriousness of the two errors and balance the risks accordingly.

In this report, errors of omission (failure to take action) are considered more serious than errors of commission (taking unnecessary action) (Lemeshow and Stroh, 1988, p. 29). Therefore, unless otherwise stated, illustrations presented here employ a 5-percent risk of failing to take action and a 10-percent risk of taking unnecessary action. In actual practice, risks must be established to suit existing circumstances, and decision rules and sample sizes modified accordingly. In exceptionally impoverished areas where resources are severely constrained, it may be appropriate to place greater constraints on errors of commission and lower the limit to 5 percent or less.

A decision rule that considers a program satisfactory if it reaches 80 percent of the target population is unlikely to declare a community unsatisfactory if it has actually achieved 70 percent coverage; however, a community with an actual achievement level of 40 percent will almost certainly be considered unsatisfactory.

But what about actual levels of 60 percent and 50 percent, the gray area between the clearly satisfactory and the clearly unsatisfactory? The likelihood that a specified sampling plan will fail to detect different magnitudes of departure from acceptable conditions can be graphed in the form of an operating characteristics curve, as in Figure 2.

Because of the importance of the operating characteristics curve in depicting the risk of error, it should be plotted and examined before endorsing any sampling plan. One can choose a published sampling plan designed to meet certain criteria (Lemeshow and Stroh, 1989, p. 76), or a plan tailored to suit specific present circumstances can be devised. In either case, sampling plans stipulate the number of observations to be made and the maximum number of defectives to be allowed for conditions to be declared satisfactory, for a given risk of error.

Applying Decision Rules and Sampling Plans

The following hypothetical example shows how decision rules are applied and sampling plans decided upon. Suppose that health officials wish to assess mothers' knowledge regarding hand washing. The officials would be satisfied if 80 percent of the women had such knowledge; therefore, they set that level as "satisfactory." Two LQAS plans are considered: one would require 15 observations per community, the other 30. These values blanket the range of sample sizes typically employed in LQAS applications. For either plan, the risk of taking unnecessary action is set at 10 percent or less.

Where the sample size is 15, in order to limit to 10 percent the risk of carrying out unnecessary action, the sampling plan specifies that the limit of defectives be set at 5; that is,

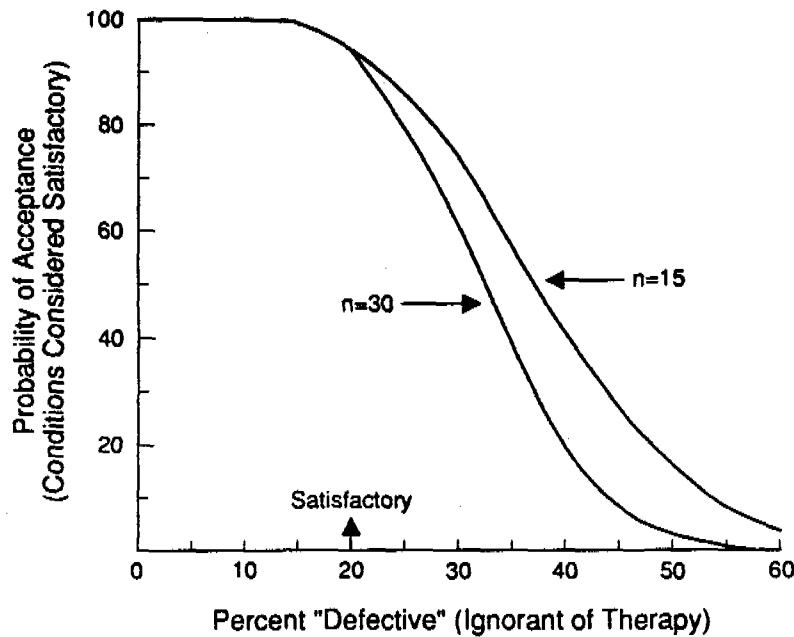


Figure 2

Operating Characteristics Curves

the state of knowledge about hand washing in a community will not be declared unsatisfactory unless at least 6 mothers in the sample are found to be ignorant of the hygienic benefits. Where the sample size is 30, the limit on defectives is set at 9.

The operating characteristics curves for the two sampling plans are plotted in Figure 2. As expected, the plan with the larger sample size is more discriminating, but even then, action is ensured only if more than half the target mothers actually failed to show sufficient knowledge about hand washing.

How successful would either of these sampling plans be in discriminating between truly satisfactory and truly unsatisfactory conditions? The performance of a sampling plan depends on the pattern of community differences encountered. If communities were of two distinct types, those with an 80-percent level of knowledge and those with a 50-percent level, the plan being considered would lead to the correct decision most of the time. What if communities are more varied, however, forming a normal distribution around some average? If the overall level of knowledge in the entire district has reached 60 percent, then several individual communities are presumably at the 50-percent level, a similar proportion have reached 70 percent, and a few are at the extremes. Figures 3 to 5 show how the two sampling plans measure up when the level of community knowledge ranges widely around an average.

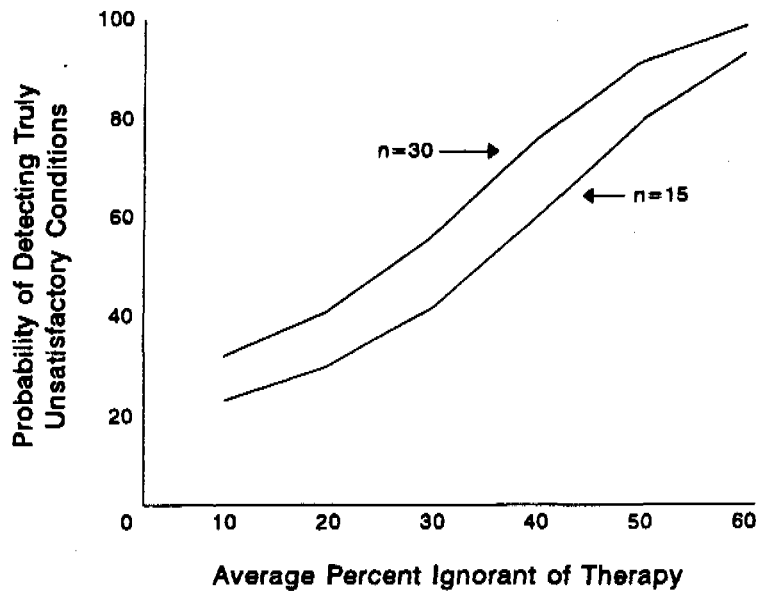


Figure 3

**Ability to Detect Unsatisfactory Conditions
(When Fewer Than 80 Percent Know Therapy)**

Figure 3 shows the proportion of occasions when less than satisfactory conditions would be detected as such. For example, if the overall district average were 70 percent and the true levels for individual communities varied between 50 and 90 percent (level of ignorance of 50 and 10 percent), then about 75 percent of the communities would fail the test, i.e., would not have achieved the desired 80-percent level of knowledge. Yet, as the figure shows, an assessment, even one using the larger sample size of 30, would identify only 43 percent of the communities as problems.

Figure 4 portrays the proportion of all situations—satisfactory as well as unsatisfactory—that would be correctly identified. The U-shaped curve that results is typical. Only when conditions are very satisfactory (when the knowledge level is at least 80 percent in actuality) or very unsatisfactory (when the knowledge level is 50 percent or less) will the district be correctly identified four times out of five. At intermediate (gray area) levels (60 to 70 percent), the likelihood of identifying those districts in which conditions are unsatisfactory (and thus taking some action to improve conditions) declines to about two-thirds.

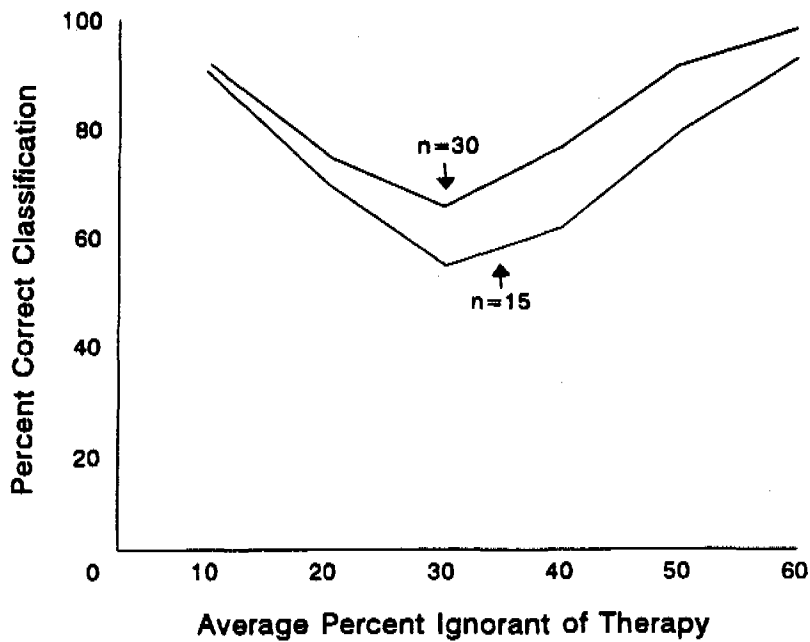


Figure 4

Probability of Correct Classification

Figure 5 shows the relative frequency with which conditions would be declared unsatisfactory on the basis of the assessment in relation to the proportion of communities that are truly below the 80-percent knowledge level. Invariably, both LQAS plans declared a level of unsatisfactory performance that is less than the true level, i.e., the results from both are excessively optimistic. This is true, even though some of the truly satisfactory districts are falsely declared to be unsatisfactory. Both sampling plans, however, show a high (80 percent or more) probability of detecting those districts that are seriously deficient, i.e., those in which 50 percent or more of the mothers are ignorant of hygiene. In fairness to the LQAS method, from a project management perspective it may not be feasible to take action in all districts that are truly unsatisfactory, even if they were known. As a practical matter, it may be that if the extreme exceptions could be isolated with high probability, that might be sufficient to keep the project fully occupied, at least initially. The LQAS method is capable of meeting these modest demands because it can be expected to detect only large departures from acceptable conditions.

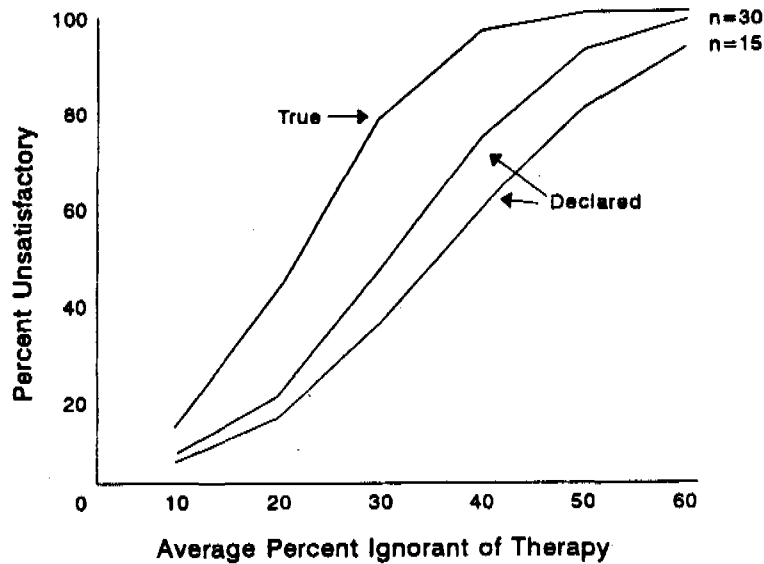


Figure 5

Underestimation of Unsatisfactory Conditions

Advantages

Lot quality assurance sampling is a useful means of detecting gross departures from prevailing standards and is appropriate in cases in which policymakers can live with rough information. It can be used as a valuable screening procedure, permitting detection of the most serious conditions and helping the manager identify sites for priority attention. It is truly a rapid assessment and is inexpensive and easy to administer.

Disadvantages

Lot quality assurance sampling should not be used in the following situations:

- For documenting erratic behavior. Application under conditions of extreme variability can produce misleading results, for the method tends to overestimate the quality of performance that has been achieved. Thus, it might prove useful in

monitoring Vitamin A distribution or immunization coverage but could be unsuitable for assessing distribution of ORS in an epidemic.

- For aggregating data from a small area to obtain an overall population estimate. This is feasible only when conditions are fairly uniform, for then the various samples are merely independent assessments of the same uniform set of circumstances and can be legitimately pooled. Care must be taken, however, to avoid combining heterogeneous data with the expectation of a meaningful aggregate finding (Lemeshow and Robinson, 1985, p. 11).
- For process monitoring. It has been argued that an attractive feature of LQAS is that the small sample sizes make frequent assessment feasible and therefore permit early identification of any "tendency to drift" from an accepted standard (Lemeshow and Stroh, 1988, p. 23). This is true, but other industrial process control methods to be discussed later, notably the use of control charts (see Chapter 11), are likely to be more informative.

Validity of Results Achieved from Small Samples

Several types of LQAS have been used in the health field. To illustrate the diversity, the technique has been used in Peru to identify localities with unacceptably low rates of immunization (Lanata et al., 1990), and in Costa Rica to assess the adequacy and quality of health services provided by primary care units (Smith, 1989, p. 5). In one study, three separate service components were monitored simultaneously: immunization coverage, rate of children under three years of age participating in growth monitoring, and extent of health center outreach into the community (Rosero-Bixby et al., 1988, pp. 14-15).

Interesting as these experiences have been, they have shed little light on the question of the validity of information obtained from the small samples used in LQAS. Findings from another study that was able to evaluate its validity are therefore of special interest (Lemeshow and Stroh, 1988, pp. 21-25). In that study, data on immunization coverage were already available from 294 health posts (HPs), 110 of which had achieved or exceeded the 70-percent target level. Each of the posts was then classified independently on the basis of sample information obtained through application of LQAS procedures. With a sample size of 24, the assessment succeeded in correctly classifying 183 of the 184 HPs that had yet to reach satisfactory levels of performance. Of the 110 posts that had reached such levels, however, only 62 (or 56 percent) were classified correctly through sampling. Thus, 231 HPs out of 294 were subject to management interventions for suspected poor performance, about 1.25 times more than the number truly below the 70-percent target level. One might question the value of an assessment that could lead to such a high level of error of commission.

Chapter 7

DOUBLE SAMPLING

Relationship to Lot Quality Assurance

Double sampling can be seen as a refinement of LQAS. In double sampling, a follow-up assessment is done in those lots or batches in which the results fall in the gray or intermediate area, i.e., where the batch has not clearly passed or failed the test. The follow-up assessment provides more information for policymakers to assess ambiguous circumstances.

Description of the Method

Double sampling schemes specify an initial sample size (n_1), a corresponding value (d_1) for the number of defectives (responses not meeting the decision rule, or criterion), a second sample size (n_2) to be obtained if necessary, and an associated defective value (d_2) that reflects the maximum cumulative number of defectives permitted in both samples, considered together.

The first sample may be thought of as a small (n_1) initial subsample of a larger single sample of size $n_1 + n_2$. Should the number of defectives found in the initial stage indicate clearly whether the entire lot ($n_1 + n_2$) has passed or failed, i.e., should the number of defectives be less than d_1 or greater than d_2 , then it is not necessary to go on to the second sampling stage. However, if the results from the first stage are not clear cut, then sampling continues until the total cumulative number of defectives (from both stages) is greater than d_2 , or until the total number sampled in both stages equals $n_1 + n_2$, whichever comes first. A double sampling scheme, therefore, is a strategy for ensuring that the total number sampled in both stages together will be no more and may be considerably less than that required for a single sample of size $n_1 + n_2$.

Representative Uses

Suppose policymakers wished to determine whether household soap use in an area departs significantly from the target of 80-percent utilization, and a single sample of about 50 would be required to detect a significant departure from the 80-percent target. A double sampling scheme with the same objective would specify: $n_1 = 35$, $d_1 = 10$, $n_2 = 70$, and $d_2 = 26$. If an initial sample of 35 households reveals that 10 or fewer are without soap, conditions can be declared to be satisfactory and sampling in that area will be discontinued. Similarly, if more than 26 of the households have no soap, sampling will be discontinued because standards have clearly not been met. In all other situations, i.e., when the sample of 35 reveals that 11 to 26 households have no soap, a second sample of 70 will be selected and acceptance will depend on finding a cumulative total of 26 (d_2) or fewer of the 105 households (35 from the

first round and 70 from the second) without soap. By increasing the sample size in a second round, policymakers are able to improve the validity of their results from the first round.

Analysis of the scheme described above reveals that a sample of 35 will nearly always be adequate if more than three-fourths or fewer than one-fourth of the households use soap. If most communities are expected to fit this description, the double sampling scheme would provide low-cost assurance of detecting the occasional exception to this pattern. Otherwise the uniform selection of a single sample of 50 individuals would produce equally satisfactory results.

Another example, drawn from Lemeshow and Stroh (1988), illustrates the advantages of double sampling. In this example, the EPI manager wishes to know how many children ages 12 to 23 months have received all of their immunizations in each of the health post areas. He thinks that the coverage level nationwide is about 60 percent, but reports from the 294 HP areas range from 20 to 100 percent. The manager suspects that the reports may not be completely accurate. To check the situation, a survey will be conducted. (The health centers have an average population of 2,500, with approximately 88 children in the target age group.)

After determining that a sample of 13,818 would be needed for a survey if the "conventional" stratified random sampling method were used, the manager chooses to use a double sampling method (see Lemeshow and Stroh, 1988, page 21, for the details of this sample size calculation). He decides that any HP with coverage below 70 percent will be considered a poor performer and will be identified for increased supervision. This particular scheme uses $n_1 = 10$, $d_1 = 0$, $n_2 = 14$, and $d_2 = 3$. In each HP area, 10 children will be surveyed. If all are found to have completed all of their immunizations, the HP will be declared "acceptable." If more than 3 defectives (children who have not completed all of their immunizations) are found, the HP will be declared "unacceptable." In either of those circumstances no further sampling would be required.

However, if 1 to 3 defectives are found, a second sample of 14 additional children will be surveyed. In that second round, as soon as 4 defectives are found (including those from the first round), the survey is stopped and the HP is declared unacceptable. If the entire second sample is surveyed and no more than 3 defectives are found (again, including those from the first round), the HP area is declared acceptable.

By using this double sampling method, the sample size can be kept much smaller than in stratified random sampling. Although the results are not as revealing, they may be just as useful in certain circumstances: "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 quality assurance sampling (QAS) scheme may be an alternative approach worthy of consideration" (Lemeshow and Stroh, 1988, p. 25).

Advantages

The chief advantage of double sampling is that data-gathering efforts are automatically expended where they are needed most. In other words, it is a way of obtaining more precise information than can be obtained with LQAS, without increasing the sample size across the board.

Disadvantages

A double or sequential sampling scheme is administratively more complex than a single sample plan. Also, the resulting differences in subunit sample size make comparative analysis among subunits more difficult.

Sequential Sampling

Considering the desirable features of double sampling, it would seem that triple sampling would be even better. Such plans are in use. Also, some assessments use sequential sampling, which provides for continuous assessment of the need for additional information before taking action (Reinke, 1988, pp. 709). The decision boundaries are two upward-sloping parallel lines plotted on a chart in which the horizontal axis represents the number of observations made and the vertical axis the number of defectives found (see Figure 6).

Sample results are plotted in sequence on the chart on which the boundaries are plotted. Each observation moves one unit along the horizontal axis. An observation that yields an undesirable result causes a one-unit shift up the vertical axis. As long as the combination of samples selected and defectives found remains within the parallel bands, sampling is continued. Acceptance occurs when the boundary to the right is passed, i.e., when few defectives have been found in a sufficiently large sample. The converse, a rash of defectives encountered in a small sample, moves the units above the upper band into reject territory.

Values for the intercepts and slope of the lines forming the decision band are determined by the defined levels of acceptable and unacceptable performance, along with the associated risks of error of commission/omission specified to be tolerable. Relatively simple formulas are used to determine the precise positions of the decision lines. Their application reveals that large differences between what is termed satisfactory and unsatisfactory serve to move the intercepts closer together, thereby narrowing the band within which sampling is continued. High standards of acceptability reduce the slope, so that a rash of defectives will quickly force a movement above the upper limit into the reject area. These effects occur automatically, of course, with calculation of the boundary equations.

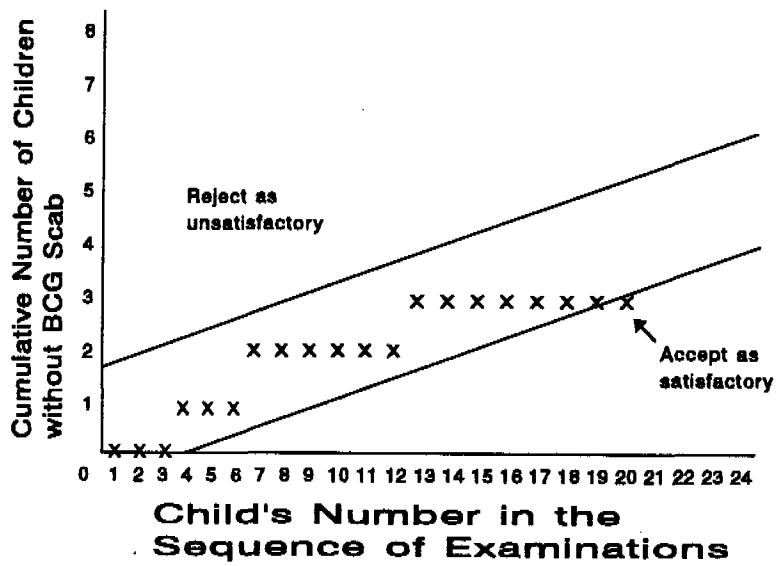


Figure 6
 Sequential Sampling for BCG Immunization

Chapter 8

REDUCED AND TIGHTENED INSPECTION

Reduced and Tightened Inspection Compared with Double Sampling

Reduced and tightened inspection, like double sampling, is a method designed to keep the sample size as small as possible without compromising the validity of the results. But, while double sampling focuses on individual units that are resurveyed if they fail to meet certain standards, this method either tightens or loosens the level of scrutiny while the assessment is taking place.

Results from the surveys of individual subunits are carefully monitored. If they fall outside certain prescribed limits, the sample size and number of defectives allowed are adjusted in a dynamic process based on the results that are being received. Like lot quality assurance sampling and double sampling, the amount of information obtained in reduced and tightened inspection is small because the sample size is small: we can ascertain only if a criterion is or is not being met.

Reduced and tightened inspection is suitable for assessments in which many units are being surveyed, and is used to assess a current situation or to monitor a situation over time.

Description of the Method

Reduced and tightened inspection is based on a specified acceptable quality level (AQL), with a limited risk of errors of commission when quality at such levels is maintained. To show how this method works, we will assume that policymakers wish to discover whether 85 percent of the households in a given area have access to a latrine, perhaps to judge the success of a program. The area will be surveyed subunit by subunit.

The sampling plan, based on a 10-percent chance of error of commission and a 5-percent error of omission, calls for a sample size within each subunit of 75 households. For conditions to be deemed satisfactory, at least 56 must have access to a latrine. Put another way, only 19 defectives are allowed if a subunit is to be deemed satisfactory.

As the survey moves from subunit to subunit, a running tally is kept of the cumulative situation observed. If any subunit records fewer than 56 households with latrine access, or if the average level observed drops below 81 percent, trouble is suspected and tightened inspection is instituted. For example, if the first three subunits showed 82-, 86-, and 72-percent coverage, giving an overall average of 80 percent, tightened inspection procedures would be initiated.

With tightened inspection, the sample size continues at 75 but now at least 62 of the sample households must have latrine access for conditions to be deemed satisfactory. In other words, the number of defectives allowed is reduced to 13. If fewer than 62 households in any subunit have latrine access, the survey is stopped. Policymakers can assume they have results showing

that the area does not meet the 85-percent decision rule. (There are other possibilities, however. For example, it is possible that the sampling plan is defective or that the area is not homogeneous enough to use reduced and tightened inspection.)

If none of the subunits falls below 62 acceptances, tightened inspection continues until the overall results reach the AQL (85 percent). Then, tightened inspection is discontinued and the normal level of inspection resumed (i.e., at least 56 households out of 75 must have latrine access).

On the other hand, if the overall average coverage rate goes as high as 89 percent, reduced inspection is permitted. In reduced inspection, the sample size is reduced to 15 (one-fifth of the original) and only 7 or more households must have latrine access. Reduced sampling is continued as long as each successive sample meets the required number of acceptances and the overall average exceeds the 85-percent standard of acceptability. Otherwise a reversion to normal inspection is indicated.

To summarize: If the results are falling between 81 and 85 percent, normal inspection is carried out; if they fall below 81 percent, tightened inspection is instituted or the study is discontinued; if they fall at or above 89 percent, reduced inspection is carried out. Once reduced inspection is instituted, it continues as long as the running total is 85 percent or greater. Once tightened inspection is instituted, it continues until the running total reaches 85 percent or the survey is stopped.

The specific percentages and numbers of defectives permitted will differ, of course, according to the standards of quality that are operative, but the procedures to be followed are unvarying. The reduced and tightened inspection method takes advantage of past experience, either with a given survey that is progressing or with a survey used, for example, on a yearly basis. For that reason, this method can reduce the amount of new information needed for decisionmaking, if the experience continues to be relevant.

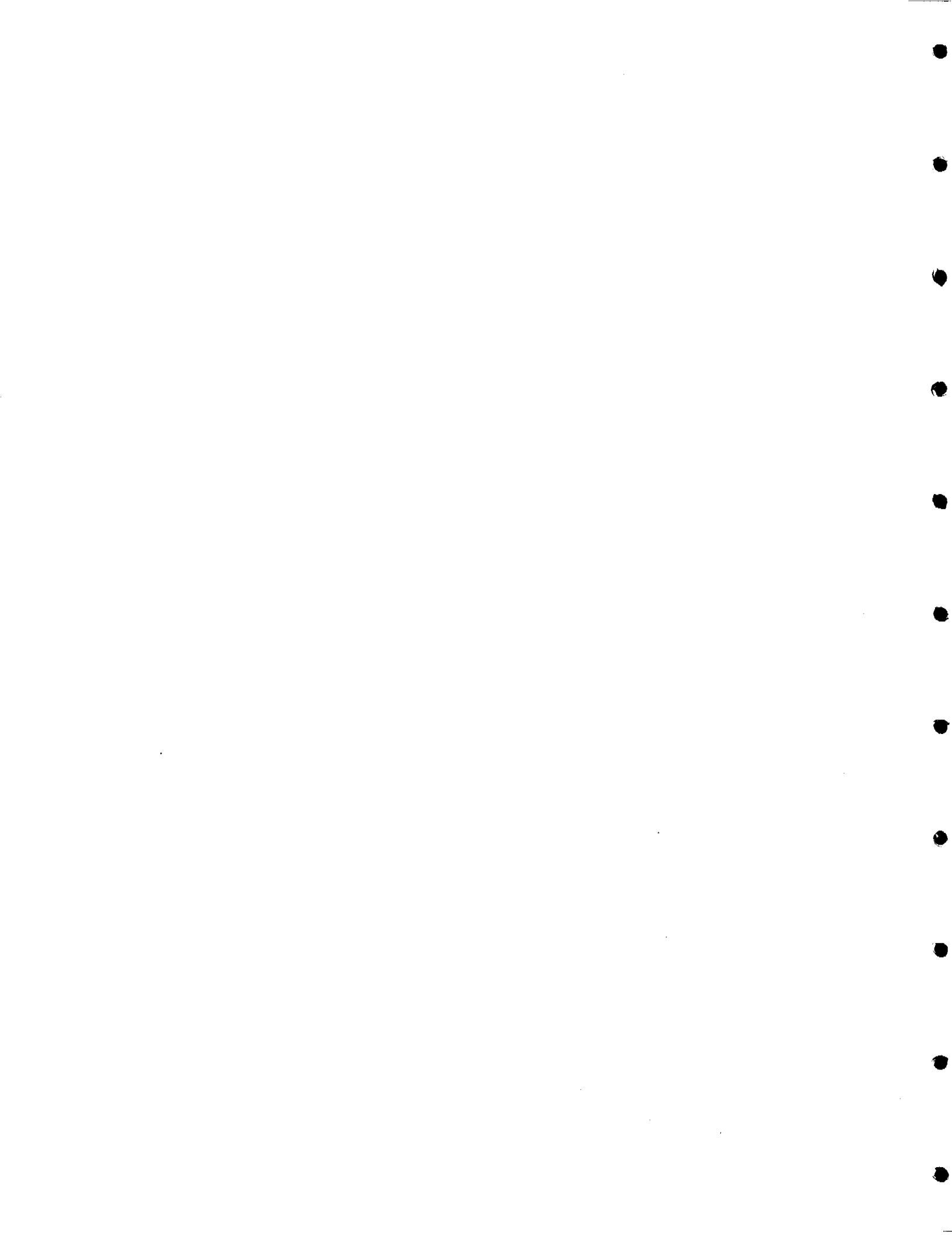
Representative Use of Reduced and Tightened Inspection

An illustrative application of reduced and tightened inspection is a national trachoma survey carried out by means of a community-by-community sweep through the country. Policymakers might want to find out, for example, if the current level of hygiene education activities has kept trachoma incidence within certain bounds. In areas of the country where little or no disease has been found, the sampling level is reduced until an area of higher disease prevalence is encountered. Then, the level of scrutiny is increased according to well-defined decision rules.

Evaluation of the Method

Reduced and tightened inspection is useful only under reasonably stable conditions, i.e., those in which performance is generally satisfactory. In such situations, reduced and tightened inspection provides a rapid, economical method that does not sacrifice the ability to detect exceptional circumstances in which standards are not being met. When conditions are erratic and essentially unpredictable, however, the advantages of these procedures do not outweigh the additional administrative burdens they entail.

The method's principal advantage is increased efficiency in data gathering through an objective appraisal of the need for information. The principal drawback is that the prescribed adjustment in the sample size is valid only when recent past experience is relevant for current decisionmaking.



Chapter 9

EPIDEMIOLOGICAL SURVEILLANCE

Description of the Method

Epidemiological surveillance, a system of collecting health information on a certain population to monitor conditions, permits policymakers to discern trends and identify problems as they arise. This kind of surveillance is defined as the continued watchfulness over the distribution and trends of incidence [of disease] through the systematic collection, consolidation, and evaluation of morbidity and mortality reports and other data (Smith, 1989, p. s5).

Epidemiological surveillance can also collect data on precursors of disease as well as on service activities carried out to alleviate health problems.

Because the goal of epidemiological surveillance is to enable managers to respond quickly to emerging problems and trends, this method typically focuses narrowly on a specific condition and/or population and seeks to provide managers with a thorough understanding of causal factors and other associated circumstances. For example, attention may be directed to a few reportable communicable diseases and may include detailed investigation of case contacts.

Representative Uses of the Method

Epidemiological surveillance might be used, for example, to discern seasonal patterns of diarrheal disease throughout a community during outbreaks and trends in ORS use. To obtain the information, women of child-bearing age in certain communities would be contacted weekly and questioned about episodes of diarrhea and actions taken.

Surveillance Sites

In epidemiological surveillance, the information sought regards populations rather than problems. Hence, data are normally obtained from designated sentinel sites and nonrandom methods of data collection are employed (Smith, 1989, p. s7). For example, certain clinics may be identified as sentinels because they are known to have competent staff and good record-keeping systems. Such clinics must be otherwise typical, however, to ensure the validity of representative data. If the high quality of information obtainable from a clinic is associated with a high quality of care, which in turn might affect patterns of service utilization, the site is probably not typical and the data will not be representative.

Broadly speaking, the matter of generalizability relates to the nature and number of sites selected. The sites might be client oriented, including the population of an entire community or district, or they might focus on certain types of providers, such as hospitals, health centers,

or individual practitioners or on other institutions, such as markets or schools (Richards, 1989, p. 936).

Multiple sites are sometimes selected to reflect dissimilar circumstances, while in other cases they are chosen to replicate seemingly similar conditions. To illustrate, Lobet et al. (1987) employed sophisticated methods of statistical analysis to distinguish districts according to population density, unemployment rate, availability of health resources, and so on, and then chose sentinel districts purposefully to reflect separate combinations of circumstances. In contrast, others have selected a number of apparently representative sites and examined subsequent findings statistically to detect differences that could not be anticipated in advance (Andersson et al., 1989).

How Epidemiological Surveillance Can Be Made Rapid

Since policymakers and managers will often wish to monitor multiple sites intensively on a continuing basis, epidemiological surveillance can easily become anything but rapid and inexpensive. Moreover, the goal of decentralized monitoring and evaluation (Andersson et al., 1989, p. 203) leads to the expansion of overall information gathering. Nevertheless, there are ways to make this method rapid and inexpensive.

First, keep the database simple so that the requisite information can be compiled routinely by workers already employed for other purposes (Richards, 1989, p. 940). Second, link data collected at different times to minimize the amount of information collected at any one time. For example, if each individual or family has a unique identification number, basic information on demographic and socioeconomic status and on sanitation practices and other environmental conditions need not be repeated. Such a system has been employed successfully in the elaborate diarrheal research effort carried out in Matlab, Bangladesh, during the past 30 years.

Active and Passive Systems

Methods of surveillance are diverse, but the most basic distinction is between active and passive systems. Active systems employ health workers to collect information; passive systems rely on the individuals involved to report events as they occur.

As might be expected, passive systems are usually characterized by substantial underreporting. In India, for example, fewer than 550 cases of guinea worm disease were passively reported in 1981, whereas an active search the following year uncovered just under 43,000 victims (Richards, 1989, p. 936). On the other hand, the smallpox eradication effort showed that passive reporting can sometimes provide useful order-of-magnitude estimates of disease levels and trends, even if it does not provide completely accurate indicators of incidence (Richards, 1989, p. 936). Passive systems can be reliable for gathering information about cholera or other diseases for which treatment is likely to be sought.

Variations of active and passive surveillance have been devised to meet the specific needs of the monitoring effort. The attempt to control guinea worm disease in Africa serves as a useful case in point. (Guinea worm surveillance is especially useful for the water and sanitation sector. Since guinea worm is transmitted solely by contaminated drinking water, a decline in the number of cases in a defined area can serve as a reliable indicator of the effectiveness of the area's water program [Richards, 1989, p. 935].)

Guinea worm surveillance in Africa is carried out in four phases. In phase 1, those communities in which guinea worm is being transmitted are identified through an active process in which all communities are surveyed. In phase 2, the effectiveness of control activities is documented. Data collection is concentrated in communities found to be affected, and baseline case counts are obtained so that progress in reducing incidence can be charted. In this phase, reliable indicators of change are more important than a precise absolute measure of incidence. This contrasts with the emphasis during phase 3, after control measures have been instituted and the rarity of the disease demands an accurate count of new cases occurring annually. When the country is declared free of guinea worm, phase 4 (passive surveillance) is introduced to detect promptly any reemergence of the disease.

Similar phased systems have been used in sentinel sites in connection with epidemics of diseases that are likely to be distributed unevenly, such as cholera. Surveillance activities then serve to target resources to particular affected areas.

Range of Applications

Epidemiological surveillance has been used in the United States in a highly specific sentinel cities program to identify impending influenza epidemics, and in Thailand to monitor refugee health status (Smith, 1989, p. 56). Extensive experience with sentinel site surveillance in Mexico and Central America has led to the creation of a one-page data collection instrument from which indicators of coverage, impact, and cost can be calculated (Andersson et al., 1989, p. 203).

Epidemiological surveillance is used mainly to monitor communicable diseases but has been readily adapted to suit broader management concerns. Surveillance techniques developed in other settings, especially industry, likewise have a significant role to play that has not yet been realized (see Chapter 11).

Peri-urban Applications

Epidemiological surveillance is particularly useful in peri-urban settings where communicable diseases are a significant danger and rapid changes in incidence and distribution must be detected promptly. For example, Islam (1991) and Mosley (1979) have observed that in cities endemic for cholera, a single case of cholera may cause an explosive outbreak of the disease. Failure to detect minor fluctuations in background rates may result in significant—and potentially preventable—epidemics.

Likewise in Pakistan, Jalil (1989) observed marked (nearly two-fold) seasonal fluctuations in diarrheal rates and rates of acute respiratory infection. Again, attention to variation in these rates could have substantial implications for intervention. Finally, minor geographic differences may result in wide fluctuations in disease rates even in the same city. Thus, in a study of 51 slum communities in Dhaka, Bangladesh, over a six-month period, Stanton and Clemens (1987) noted differences in the incidence density ratio of diarrhea disease in excess of 25-fold (0.07-1.82).

Chapter 10

DEMOGRAPHIC SURVEILLANCE

Difficulties of Demographic Surveillance in Developing Countries

The monitoring of births and deaths over time represents another form of surveillance with a long history. Indeed, the experience gained by demographers is so rich that it has been embodied in separate literature.

Demographic surveillance is fraught with difficulties, particularly in developing countries. Trends in birth and death rates, while among the most important indicators to be monitored, are often underreported. This problem is especially severe in the case of early infant deaths, in which both the birth and the death are likely to be overlooked.

When responsibility for the registration of vital events is vested in the health system, records are usually incomplete because many births and deaths are not attended by health personnel. When families are asked to report events to another government agency, registration is unlikely to be complete unless some benefit outweighs the inconvenience involved. Thus, attempts to provide complete enumeration of vital events have been notoriously unsuccessful.

Sample registration systems have sometimes been effective. A computerized system developed in Bangladesh, for example, combines field-level data-entry procedures with a user-friendly software analytical package that regularly updates demographic and service indicators (Phillips et al., 1989). Despite encouraging experiences such as this, the reality remains that sample registration systems are hampered because the infrequency of births and deaths requires large numbers of respondents to be surveyed for adequate numbers of births and deaths to be recorded within a reasonable period of time.

Techniques for Rapid Demographic Assessment

Demographers have devised a number of ingenious techniques for rapid assessment. Three are described here.

Pregnancy Histories

The most common of the three techniques involves obtaining pregnancy histories from a representative sample of women of child-bearing age. Estimates of age-specific fertility rates as well as child mortality rates can be obtained from such histories. The validity of the results depends on uniformly complete recall of all types of events occurring at all times in the past. Because it is unrealistic to assume that people can remember everything, a limited recall period is used, and the interviewer probes extensively for events of interest. This reduces the amount of information obtained and increases the cost. Typically, the probing includes reference to local historical events in order to stimulate recall of vital events and to establish their timing.

Reference to natural disasters has the added advantage of stimulating recall of a period when death and disease were especially likely.

Routine Encounters

In the second approach, routine encounters with mothers—for example, at the time of delivery of a baby or vaccination of a child—may be used to ascertain the outcome of a previous birth. Some inexpensive investigations of this type (Bicego et al., 1989) have produced satisfactory approximations of the probability of death before the age of two, but they require contact with a representative subgroup of the target populations. For this reason the approach would not work if those seeking immunization services differed in some important respect from nonusers of the services.

Survival of Close Relatives

A third approach is to collect data in censuses and surveys about the survival of spouses or close relatives of respondents and to calculate a conventional measure of survivorship from these (Timaeus, 1986). For example, in two surveys of the same population conducted five years apart, the proportion of first-survey respondents in each age group with living mothers can be compared with the equivalent proportion for second-survey respondents who are five years older.

Verbal Autopsies

Simple methods devised to verify the occurrence of a death are likely to leave the cause undetermined. Somewhat more sophisticated, yet still feasible, methods have been developed for soliciting from caretakers of the deceased a description of the circumstances of death. These verbal autopsy methods have been shown to yield results of acceptable accuracy, especially in connection with child deaths due to well-defined causes such as pneumonia (Kalter et al., 1990).

Limitations of Rapid Demographic Surveillance

Any methods that rely upon recall of past events must be sensitive to problems of underreporting. More complete information is likely to be obtained when events are reported as they occur. One solution is to employ village informants to collect information and to pay them according to the number of verified events reported.

Chapter 11

INDUSTRIAL PROCESS CONTROL METHODS

Process Control Methods—Not Yet Used in the Health Field

Among the many techniques for monitoring industrial processes, control charts have had the most extensive and varied use. They are obviously applicable to health surveillance but, surprisingly, have not yet been so used. Therefore, the following discussion must necessarily draw upon hypothetical examples.

Process control procedures have been devised to monitor performance involving both *attributes* (e.g., ORS packets in stock or not) and *continuous variables* (e.g., number of packets distributed last month). Applications involving the latter, because they call for very small samples, are especially simple and rapid, as well as informative.

Description of the Method

In industrial process control, performance is monitored by means of a chart or graph on which the control limits are shown. Whenever the results fall outside the limits, a problem is indicated, i.e., the process is out of control. For example, to monitor the incidence of chronic diarrhea in a certain subdistrict, a chart showing the number of cases could be plotted with easily calculated upper and lower control limits. New cases would be recorded on the chart every week. If the number of cases went above the upper control limit on the chart, an epidemic would be immediately apparent. Trends would also be clearly shown on such a chart.

Types of Process Control Methods

Several types of process control methods exist. All feature frequent selection of very small samples (5 to 20 observations each) to ensure prompt signaling of trouble, calculation of simple indicators from the findings, and visual display of the results. They are discussed one by one below.

Ranges and Averages

In industrial process control, ranges, rather than more complicated standard deviations, are generally used as measures of variation. It can be shown mathematically that the difference between the minimum and maximum values in a random sample of five observations from a normal distribution will average 2.3 standard deviations. Therefore, one may easily calculate control limits to serve as decision points from the average range. These can be interpreted just like more sophisticated statistical measures.

Taking a concrete example, suppose a system of home visiting is carried out by community health workers supervised from health posts, each of which looks after five workers. These workers average 28 visits per week, and the average range between the best and poorest weekly performance within a health post is 7 visits. The district health officer would like to monitor the system in several ways. In particular, he wishes to do the following:

- Identify promptly any post that exhibits exceptional performance levels (good or bad) during any week.
- Note any post in which variability among workers is extreme.
- Become aware of important performance trends at any post.
- Compare average levels among posts over time.

These concerns can be monitored by means of control charts for averages and ranges constructed along the lines of Figure 7. Published sampling plans for use in industrial process control indicate that, for samples of five, individual ranges should be no more than 2.11 times the average range, and separate averages of five from a stable process should almost never deviate from the overall process average by more than 0.58 times the average range. Using the present example, then—

$$\text{Upper limit on ranges} = (2.11)(7) = 14.8$$

$$\text{Upper limit on averages} = 28 + (0.58)(7) = 32.1$$

$$\text{Lower limit on averages} = 28 - (0.58)(7) = 23.9$$

Any weekly result outside these limits indicates that the subject health post is out of control, i.e., its pattern of performance differs significantly from usual practice.

For example, an observed *range* of 15 or more between the best and worst results would indicate the presence of an exceptional (either superior or incompetent) worker at the subject post. An “out-of-control” *average* means that the post as a whole is exceptional. The health post depicted in Figure 7 is well within the upper control limit for ranges.

The visual display of plotted results makes it possible to discern trends and compare performance among posts without resorting to sophisticated statistical analysis. Useful rules of thumb have also been devised to be used with the charts. For example, since the probability is less than 1 in 100 that seven consecutive results will fall by chance on the same side of (above or below) an average, such a result is treated as out of control, even if no single observation is beyond the upper control limit. On these grounds, the health post depicted in the lower chart of Figure 7 gives evidence of inferior performance from the third week on.

Obviously, this type of industrial process control is applicable to a wide range of water and sanitation activities from, for example, the maintenance of pumps by environmental engineers to the number of latrines inspected or the number of hygiene education activities carried out

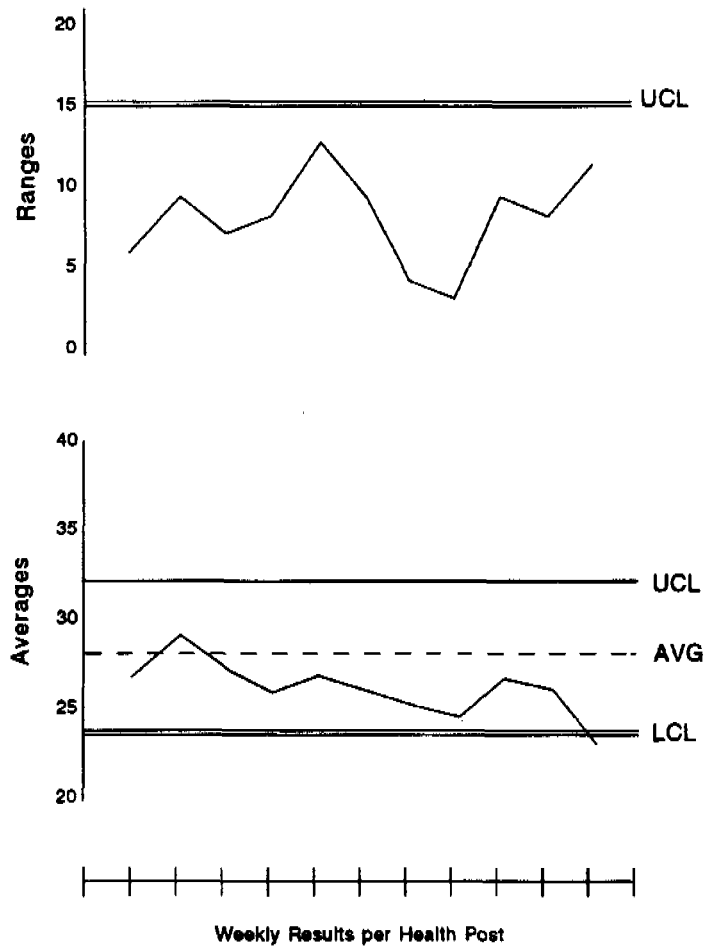


Figure 7
Control Charts for Home Visiting

by community health workers. The use of control charts at sentinel sites or elsewhere in the monitoring of communicable diseases would permit prompt detection of epidemics in a way that is easily comprehended by minimally trained personnel, yet is objective enough to form a natural basis for resource allocation.

P-Charts for Proportions

In addition to control charts for averages and ranges, p-charts for proportions are used. Procedures for handling proportions are much like those employed in analyses described earlier involving discrete variables, but p-charts retain the control chart advantages of simplicity, visual clarity, and ability to identify trends through frequent comparative sampling.

For example, control charts for chronologically tracking the proportion of diarrhea cases treated properly with ORT could be used, not only to display apparent trends but also to determine whether they are statistically significant.

C-Charts for Case Counts

Events that are relatively uncommon but have a constant likelihood of occurrence tend to form a so-called Poisson distribution pattern, which is based on the fact that departures over time from the average number of occurrences are determined solely by the average itself. Conditions will seldom differ from the average by more than two times the square root of the average. Thus, a hospital that admits 25 patients per week on the average will seldom experience a week in which more than 35 patients are admitted. Should such an exception occur, it would be promptly detected and its cause (e.g., an outbreak of shigella or the closure of a nearby hospital) readily determined.

Information regarding the average count of events is all that is needed to provide a complete description of the full range of conditions to be encountered under normal circumstances. Observations outside this range are indicative of an out-of-control situation. Thus, basing a c-chart on the Poisson distribution does not require knowledge of the population size or the probability of occurrence of the subject event in that population. To derive the above understanding about hospital patient admissions, it was not necessary to know the precise catchment area of the hospital or the probability that an individual member of the population would be admitted during a particular week.

Clearly, Poisson distributions lend themselves to simple and rapid methods of assessment. The c-chart applied to Poisson distributions represents one such method that captures control chart advantages as well. Consider, for example, the surveillance of new cases of chronic diarrhea in a certain subdistrict where the average number reported per month has stabilized at around 36. A c-chart plot of monthly figures would place control limits at 24 and 48 cases, 12 units (2×6) on either side of the average. Thus, a monthly report of more than 48 new cases would suggest a significant increase in incidence, whereas fewer than 24 cases would provide evidence of genuine improvement or lax reporting.

Moving Averages

Although control charts for averages can be very useful in depicting trends, moving averages are sometimes employed to give an even clearer picture. A moving average is an average computed of a specified number of observations, say the most recent 10 observations. Because a moving average tends to form a smoother line than the connection of individual points, trends appear more striking in plots of moving averages. Care must be taken in interpretation, however, for the various points on the line are not independent of one another; therefore, an apparent trend may simply portray the effect of one maverick observation persisting in succeeding averages. Thus, for example, the use of moving averages would be illogical in the tracking of epidemic diseases, such as cholera, which exhibit a substantial, but time-limited, upsurge in cases.

Exponential Smoothing

Exponential smoothing is a more recent innovation that has gained considerable attention, especially for forecasting. The thinking behind this concept is that the dynamics of a particular situation may be such that recent experience may be more relevant than earlier observations in assessing the current situation; yet a single recent observation is likely to be an unreliable indicator of present underlying conditions. Therefore, exponential smoothing uses past data but reduces exponentially the weight given to the data as one goes back in time. Specifically, with a weighting factor of 0.9, weights are applied as follows:

Latest observation: $1.0/10 = 0.100$

Second-last observation: $.9/10 = 0.090$

Third-last observation: $.9^2/10 = 0.081$

(etc.)

An attractive feature of exponential smoothing is that it provides a simple way to update averages. Again using a weighting factor of 0.9 for illustration, the previously calculated exponential average gets that weight in determining the new average, while the latest single result gets a weight of 0.1. The resulting average is therefore somewhat like a 10-point moving average, whereas a weighting factor of 0.95 would produce results akin to a 20-point moving average. The weight selected thus depends on how rapidly circumstances are thought to be changing.

Exponential smoothing could be especially useful in monitoring an endemic disease (say chronic diarrhea) in which a control measure introduced for another purpose (improved sanitation, perhaps) is expected to have some impact on prevalence of the disease in question.

Average Outgoing Quality Limit

Sometimes process monitoring is coupled with rectification of conditions found to be unsatisfactory. Suppose, for example, that mothers attending a clinic for children under five are to be instructed in hand washing and given a bar of soap. Subsequent home visits then ascertain whether mothers have actually received the soap; if not, they are given a bar. The proportion of mothers to be checked depends on evidence of the quality of the clinic activity; if many mothers are found to be without soap, more will be investigated.

The final proportion of "defective" mothers, i.e., those ultimately without soap, will be low if the clinic does a good job in distributing the soap in the first place. Final quality will likewise be high on the average if clinic performance is exceedingly poor, for then sampling is likely to be intensified, more "defective" mothers will be found, and the faulty condition will be rectified. Intermediate levels of clinic performance will produce the poorest final outcome on the average, for moderately poor performance will often not be detected and rectified.

The intermediate level of quality that results in the most unsatisfactory final outcome is known in industry as the Average Outgoing Quality Limit (AOQL). Sampling plans have been designed to reduce this limit to minimally satisfactory levels. They apply the following general decision rules, specific values of n and f being determined by the AOQL selected:

- Begin with 100-percent inspection until n units in succession are found without defect.
- After this happens, inspect only every f th unit until a defect is found.
- Then revert to 100-percent inspection until n consecutive units are again found without defect.

If it has been decided, for example, that at least 98 percent of mothers should have the bar of soap, this translates into an AOQL of 2 percent. The plan devised for such circumstances calls for $n = 76$ and $f = 20$ (i.e., 5 percent of cases). This means that home visits will be made to each of the first 76 mothers who visit the clinic. If all can produce a bar of soap, a 5-percent sample of mothers will be visited thereafter. Whenever a sample mother is found to be "defective," the practice of 100-percent home visiting is resumed.

Assessment of the Method

Industrial process control methods hold a lot of promise for epidemiology; they are truly rapid and easy to administer. Because the data is collected frequently and the sets of data from each inspection are linked, the amount of data needed in each set is minimal. This permits quick and easy calculation of indicators, prompt signaling of trouble, and clear discernment of trends from visual displays of results. The displays also facilitate comparison of communities or service units.

Control charts are intended to be used after process stability has been achieved. Thus, they are not advised when conditions are erratic (for example, in the early stages of a program, when numerous "bugs" are still being worked out). They can be helpful, however, in detecting an occasional outbreak of disease or in identifying a few health workers experiencing unusual difficulties in meeting performance standards.



Chapter 12

CASE-CONTROL ANALYSIS

Locating Cases of Rare Conditions

In sample survey methods, the sampling unit is not necessarily the same as the unit of analysis: that is, samples may be selected from a listing of households in an area, but the ultimate focus of interest may be children with diarrhea. When the condition of concern is uncommon, as in the case of rare diseases, it may be necessary to contact a large sample of households before enough cases are identified to permit meaningful analysis. Contacting such a large sample is anything but rapid and inexpensive. Under these circumstances more efficient methods of data collection are sought. The case-control approach is one widely used possibility.

Description of the Method

Application of the case-control method begins with the selection of readily available *cases* and *controls*. For example, members of the population who are on health center rolls for a specified reason could be the cases. Note that the "cases" need not suffer from a disease condition; they may be users of a certain type of service. Controls are individuals who are in many respects similar (perhaps the same age and sex) to the cases but who differ in one important regard: they do not exhibit the condition that distinguishes a case.

The purpose of the study or investigation is usually to test a hypothesis about factors associated with the condition. Again, the factor need not be a toxic substance thought to cause disease; it might be socioeconomic status, conditions of crowding, or the like. Ideally, cases and controls are similar in most respects but differ in their exposure to the factor that is hypothesized to be associated with the condition being analyzed.

To illustrate, a peri-urban study in the Philippines compared diarrhea cases seen at certain clinics with controls seen at the same clinics for respiratory problems. The two groups were compared as to their source of drinking water (Baltazar et al., 1988; Briscoe et al., 1988).

Representative Use of the Method

Because of its severity, chronic diarrhea is of considerable interest, even though it is a relatively rare condition. Its rarity makes it difficult and costly to identify enough cases in the general population to permit meaningful analysis. A community survey of the problem would require too large a sample and too much time to be feasible. Using the case-control approach, cases could be identified from among hospital admissions. If the target group were children under five years of age, for example, a control group of children of the same age would be selected from among those admitted to the hospital for other reasons. The two groups would

then be compared with respect to their usual source of drinking water to determine whether this environmental factor was associated with increased risk of contracting diarrhea.

Finding Representative Cases and Controls

Often it is difficult to find cases and controls that are representative of the target group (Buck et al., 1988, p. 537). Readily available cases may not be representative; for example, diarrhea cases seen at the health center may differ in some important way from those not receiving such attention. Similarly, the controls might not be representative of the population at large, since they, too, are often selected from a convenient source.

Avoiding Distortion

Case-control studies can also be affected by more subtle forms of distortion, three of which are noteworthy.

Effects of Confounding

First, effects of confounding could occur if an important factor is more often present in one study group than in the other (Briscoe et al., 1988, p. 446). It is possible, for example, that diarrhea cases seen at the health center tend to be malnourished toddlers of low socioeconomic status, whereas patients with respiratory illness are drawn from a broader spectrum of children.

Interaction Effects

Interaction effects can introduce a second type of difficulty into the analysis. Referring to the example above, even if a similar proportion of the children in both groups is malnourished, it is likely that the combination of malnutrition and poor sanitation exerts an especially potent effect on the incidence of diarrhea. The combined effect may be hard to discern apart from the influence of the two factors separately.

Separating Cause and Effect

Observed differences between cases and controls also might be due to the course of the illness (or other condition), rather than the pattern of exposure prior to the onset of illness (Buck et al., 1988, pp. 535-36). The danger is especially great for chronic diseases. For example, individuals blinded by trachoma might be seen to be following a much different life-style at the time of identification as "cases" than would have been discerned at an earlier time. Such a difference might be identified as a cause of their condition when it is actually an effect.

The association among bottle feeding, diarrhea, and malnutrition is another that has confounded retrospective analysis. Although the association is well established, the sequence of cause and effect is by no means clear.

Avoiding Difficulties

To avoid these difficulties, one might follow a representative sample of the target population over time and identify conditions of interest as they occur. While analytically sound, this approach could become very costly and time-consuming, especially if the conditions are relatively rare and develop slowly.

Problems of confounding and interaction could be brought under control if a so-called factorial design were employed in sample selection. This would further complicate the data collection and analysis, of course.

Odds Ratios

Case-control studies rely on techniques of analysis that are less sophisticated but by no means simple (Rodrigues and Kirkwood, 1990). For example, errors of estimation are likely if one wishes to determine the incidence of disease in a population exposed to a certain health hazard compared with the incidence in another group not exposed to the risk. This is so because disease incidence cannot be determined precisely in a case-control study. Therefore, other measures, notably the odds ratio, are employed. Because of the "proxy" nature of these measures, their interpretation requires some knowledge of epidemiology (Briscoe et al., 1988, p. 444).

Here is how the odds ratio is used. In the example given in the table below, 40 percent of the members of a community of 500 healthy individuals are exposed to a certain toxin in drinking water (e.g., naturally occurring arsenic). Within one year, 30 of the exposed persons develop a certain liver condition, and by the end of the second year the number has tripled to 90. Among those not exposed, the condition is less common and develops more slowly: 15 cases are observed after one year, and 30 are identified by the end of the second year.

ILLUSTRATIVE CONDITIONS						
	Total	Cases @ Time -			Ctrl @ Time -	
	Pop	0	1	2	1	2
Exposed	200	0	30	90	17	35
Unexposed	300	0	15	30	28	85

Because the situation is dynamic, the values obtained for any measures of interest will depend on the time at which measurements are taken. After one year, the illness rates in the exposed and unexposed groups will be 15 and 5 percent, respectively. The relative risk associated with exposure would therefore be judged to be 15:5, or 3.0. After two years, however, illness rates of 45 and 10 percent would produce a relative risk of 4.5.

Realistically, of course, the population denominator would usually be unavailable to calculate illness rates. The most that could be expected would be identification of all 45 cases after one year or 120 cases after two years. Equal numbers of controls might then be obtained. If the mix of exposed and unexposed individuals in the control group matched that in the healthy population at large, there would be 17 exposed and 28 unexposed controls selected in year 1; in year 2, the corresponding numbers would be 35 and 85.

Case-control analyses typically determine odds ratios from numbers such as these. Using year 2 data for illustration, the odds for contracting disease among the exposed were found to be 90:35, or 2.57. Similarly, the odds among the unexposed were 30:85, or 0.35. The ratio of the two odds, $2.57/0.35$, is 7.3. Thus, it would be estimated that those who are drinking water with naturally high levels of arsenic are more than seven times as likely to contract the liver disorder than those who are not exposed. In the illustrative case, the error in the estimate may be determined because the true relative risk is known to be 4.5. The discrepancy is large because the example dealt with a rather common condition in a small population; in any case, however, the odds ratio tends to overestimate relative risk.

Additional errors could be introduced, of course, if not all of the cases were identified, if the controls were not strictly representative of all healthy individuals, or if the time when the cases were selected is significant. The latter point can be especially troublesome when cases and controls move repeatedly in and out of the "diseased" state, as happens, for example, with diarrhea.

Experience with the Case-Control Approach

Because of the method's attractive features, a considerable body of experience with case-control applications has accumulated. Reference has already been made to the Philippine study of the association between diarrhea and water use (Baltazar et al., 1988; Baltazar and Solon, 1989), and a similar assessment in Malawi could be cited (Briscoe et al., 1988). Two different case-control studies in India have indicated that repeated incidents of severe diarrheal disease and/or heatstroke can contribute substantially to the risk of blinding cataract (Minassian et al., 1989). A similar case-control approach was employed in Colombia and Haiti to answer much different questions regarding the use and nonuse of health services (Smith, 1989, p. s10). Here the issue was the more appropriate targeting of scarce resources in order to achieve desired service outcomes. Analyses of the vaccination status of individuals suffering from immunizable diseases have relied upon the case-control approach, noting that the odds ratios derived from the data provide approximate measures of vaccine efficacy.

Validation of the Findings of Case-Control Studies

Whereas numerous case-control applications have been recorded, reported validation of their findings is much less common. One vaccine efficacy study and the Philippine investigation of diarrhea are notable exceptions. Results in the former case were found to be similar to those obtained from a classical cohort study (Smith, 1989, p. s11), while in the Philippines and Malawi, long-term, more expensive prospective studies are under way for comparative purposes (Briscoe et al., 1988, p. 444).

Two case-control studies carried out in Bangladesh likewise produced evidence of validity. In one study, substantial reduction in mortality among children receiving the measles vaccine (Clemens et al., 1988) was later confirmed by a cohort analysis of the larger data set from which the cases and controls had been selected (Koenig et al., 1990). Another case-control study in urban Bangladesh found the incidence of diarrhea among children to be associated with three water and sanitation practices: lack of hand washing before food preparation, open defecation by children in the family compound, and inattention to proper disposal of garbage and feces (Clemens and Stanton, 1987). Subsequent educational intervention directed at these three factors produced an improvement in hand washing practices and a reduction in the incidence of diarrhea, thereby tending to validate the case-control findings (Stanton and Clemens, 1987).

Assessment of the Case-Control Approach

Despite the limitations of the case-control approach, it is frequently used as a quick and convenient way of assessing uncommon events that are already known to have occurred instead of having to wait for them to happen. Moreover, it may be the only feasible approach on ethical grounds, as when a prospective study would require that a presumably efficacious intervention would have to be withheld from a control group (Rodrigues and Kirkwood, 1990, p. 205).

Case-control studies will continue to be attractive because they are feasible and sometimes the only practical option. Their inherent limitations regarding the correctness and completeness of findings should be recognized and minimized.

With regard to correctness, consideration must be given to the possibility of bias in the selection of both cases and controls, and the likelihood of errors (overestimation) in estimates of relative risk or other indices must be acknowledged. Ingenious methods may have to be devised to protect the integrity of the approach. For example, a case-control analysis imbedded in a representative population sample survey may minimize selection bias while retaining the advantages of speed and convenience (Smith, 1989, p. s11).

With regard to the issue of completeness, the small sample sizes usually associated with case-control studies make it exceedingly difficult to isolate the effects of confounding or interaction (Briscoe et al., 1988, p. 446). In this respect, however, case-control studies are no different from other methods of rapid assessment. As previously noted, the precision with which

combinations of effects can be estimated depends on the size of the sample, not the total size of the subgroup of interest; therefore, a reasonable amount of information must be available from a specified set of circumstances if the composite influence of those circumstances is to be assessed with any reliability.

Chapter 13

SELECTING THE MOST SUITABLE METHOD

This report has reviewed nine techniques for rapid assessment of health conditions in peri-urban areas, eight of them in some detail. The review has focused on examining *when-to* rather than *how-to* use each of these methods. The objective has been to provide guidance to planners and project managers who must collect and analyze the information needed to allocate resources and manage health programs in developing countries. Descriptions of the assessment methods are relatively nontechnical and are intended to help managers select the most efficient method for a particular information need. Additional illustrations and guidance on implementation can be found in the articles cited in Chapters 5 through 12 and in the Bibliography.

As has been stressed throughout this review, the key to cost-effective information gathering is first, careful specification of both the decision to be made and the minimum amount of information required for that decision, then second, selection of the data collection method best suited to providing that information in a valid, reliable, timely and efficient manner. To be cost-effective, different types of decisions call for different data collection strategies. The following scenario illustrates how a district manager might make use of a rapid assessment strategy to design, monitor, identify and resolve problems, and evaluate a proposed project or set of activities. In this illustration, the manager employs a series of different methods for data collection and analysis, each appropriate for the particular stage of the project and the specific decisions to be made. Together the set of methods constitute an efficient strategy for obtaining the information needed for management decisionmaking.

Suppose a new low-cost loan program for latrine construction is being implemented, along with improved hygiene education at neighborhood health clinics; neighborhoods in which fewer than 60 percent of the households have access to an appropriate latrine are eligible for phase 1 of the loan program. An LQAS survey is used to identify the set of eligible neighborhoods. For subsequent before-and-after evaluation of the program, a cluster sample survey of households in the eligible communities establishes baseline estimates of existing household water supply and sanitation facilities, recent diarrhea morbidity, use of ORT, and hygiene knowledge and practice. (Note that preliminary screening by the relatively small LQAS survey means that the baseline cluster survey can be limited to the eligible neighborhoods.)

The implementation process itself is monitored using process control charts showing trends in loan approvals, construction starts and completions, loan repayments, and so forth. Emerging problems (say, loan defaults) are spotted, and particular neighborhoods are identified where performance is significantly below the normal range of variation. These are targeted for special problem-solving attention. At the same time, data from the existing epidemiological surveillance system are monitored for trends and seasonal variation in the number of diarrhea

cases seen in all health facilities and the number of severely dehydrated diarrhea cases admitted to the hospital serving the area, as well as in the number of diarrhea-related deaths identified by community nurses in verbal autopsy reports. A sudden rise in any of these indicators triggers increased priority for instruction in ORT and hygiene education both at the neighborhood clinics and by outreach workers.

At the end of the first year, another brief LQAS survey tells the managers whether or not a projected intermediate coverage target of at least 50 percent has been met, and points to the neighborhoods where greater efforts are needed during the second year. At the end of the second year, a follow-up cluster survey permits a before-and-after assessment of change in the number of households with latrines and in hygiene knowledge and practice, and so forth.

Many different illustrative scenarios could be spun out, but there are certain regularities in the types of questions asked by managers, and therefore in the most efficient methods to use to obtain the information needed to answer those questions. To aid planners and managers in selecting the most suitable, the assessment methods described in this document have been organized according to the type of questions each is best suited to help answer. These methods are presented in the following table.

Table 1

Selecting Rapid Assessment Methodologies

When you need to	Consider using ...
<p>Estimate population parameters (means, proportions, variances)</p> <ul style="list-style-type: none"> ■ 1 population, 1 time period <p>Ex. % of households with piped water in the house</p>	<ul style="list-style-type: none"> ■ Cluster sample survey
<p>Test significance of differences between populations:</p> <ul style="list-style-type: none"> ■ 1 population, >1 time period ■ >1 population, 1 time period ■ >1 population, >1 time period <p>Ex. % of households with soap, before and after a health education campaign</p>	<ul style="list-style-type: none"> ■ Cluster sample surveys
<p>Test significance of departure from a predetermined standard</p> <p>Ex. Has the program reached its target of at least 60% of mothers who report having used ORT during the most recent diarrhea episode?</p>	<ul style="list-style-type: none"> ■ LQAS survey ■ LQAS survey with double sampling of ambiguous subunits ■ Sequential sampling ■ Reduced and tightened inspection
<p>Identify specific subunits (villages, health facilities, etc.) that depart significantly from a predetermined standard</p> <p>Ex. List of specific peri-urban neighborhoods in which fewer than 80% of households have access to a functioning toilet within the building</p>	<ul style="list-style-type: none"> ■ LQAS ■ LQAS survey with double sampling of ambiguous subunits ■ Process control charts (ranges)

When you need to	Consider using ...
<p>Identify trends (numbers, rates, means, proportions, variability):</p> <ul style="list-style-type: none"> ■ in morbidity ■ in vital events ■ in service delivery performance ■ in service outcomes ■ in client KAP ■ etc. <p>Ex. Is there a seasonal trend in household expenditures for water?</p> <p>Ex. Has there been a reduction in the % of hospital admissions for severe dehydration over the past three years?</p>	<ul style="list-style-type: none"> ■ Epidemiological and demographic surveillance systems: <ul style="list-style-type: none"> □ Active: <p>repeated or continuous surveys (convenience, purposive, cluster, etc., sample)</p> □ Passive: <p>repeated or continuous monitoring of events at fixed sites</p> ■ Process control charts (means, proportions)
<p>Identify emerging problems:</p> <ul style="list-style-type: none"> ■ in health conditions ■ in service delivery processes <p>Ex. Verbal autopsies this month indicate three deaths probably due to measles, complicated by diarrhea and dehydration.</p> <p>Ex. Clinic attendance records show a sudden increase in the number of measles cases this week.</p>	<ul style="list-style-type: none"> ■ Surveillance systems (health, demographic) ■ Process control charts (ranges, number of cases)
<p>Compare subunits with respect to variability in performance</p> <p>Ex. Pumping station number two was out of service on nine days last month; that is more than three times the average number of days lost each month by the system as a whole.</p>	<ul style="list-style-type: none"> ■ Process control charts (means, proportions, ranges)

When you need to	Consider using ...
<p>Monitor service delivery outcomes and correct deficiencies concurrently</p> <p>Ex. The target is 98% of mothers who know how to mix ORS. Mothers' knowledge will be monitored as they leave the clinic to ensure that they have been properly instructed in home care for diarrhea. The first 76 mothers know how to prepare ORS; thereafter, every 5th mother is interviewed. Should any subsequent mother be unable to give the proper instructions for mixing ORS, health workers are asked to take corrective action. Monitoring 100% of mothers is begun again and the cycle continues.</p>	<ul style="list-style-type: none"> ■ AOQL (Average Out-going Quality Limit), i.e., reduced and tightened inspection of outcomes, with corrective action
<p>Determine whether some factor is related to a specific health problem</p> <p>Ex. With widespread use of ORT and other fluids in the management of diarrhea in the home, cases coming to the clinic with severe dehydration are becoming relatively rare. Are children with severe dehydration more likely to have been treated with antidiarrheals before coming to the clinic than those with only moderate or no dehydration?</p>	<ul style="list-style-type: none"> ■ Case-control study



BIBLIOGRAPHY

- Agbodaze, D., C.A. Abrahams, and S. Arai. 1988. "Enteropathogenic and Enterotoxigenic Escherichia Coli as Aetiological Factors of Infantile Diarrhoea in Rural and Urban Ghana." *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 489-91.
- Andersson, N., E. Martinez, F. Cerrato, et al. 1989. "The Use of Community-based Data in Health Planning in Mexico and Central America." *Health Policy and Planning*, 4(3): 197-206.
- Bailey, R., C. Osmond, D. Mabey, et al. 1989. "Analysis of the Household Distribution of Trachoma in a Gambian Village Using a Monte Carlo Simulation Procedure." *International Journal of Epidemiology*, 18(4): 944-51.
- Baltazar, J., J. Briscoe, V. Mesola, et al. 1988. "Can the Case-control Method Be Used to Assess the Impact of Water Supply and Sanitation on Diarrhea? A Study in the Philippines." *Bulletin of the World Health Organization*, 66(5): 627-35.
- Baltazar, J., and F. Solon. 1989. "Disposal of Faeces of Children under Two Years Old and Diarrhea Incidence: A Case-Control Study." *International Journal of Epidemiology*, 18(4-Suppl. 2): s16-s19.
- Bentley, M., G. Pelto, W. Straus, et al. 1988. "Rapid Ethnographic Assessment: Applications in a Diarrhea Management Program." *Social Science and Medicine*, 27(1): 107-16.
- Benyoussef, A., J.L. Cutler, R. Baylet, et al. 1973. "Migration et Urbanisation Une Etude Collective au Senegal." *Bulletin of the World Health Organization*, 49: 517-37.
- Bertrand, W. 1985. "Microcomputer Applications in Health Population Surveys: Experience and Potential in Developing Countries." *World Health Statistics Quarterly*, 38: 1.
- Berwick, D.M. 1989. "Continuous Improvement As an Ideal in Health Care." *The New England Journal of Medicine*, 320(1): 53-6.
- Bicego, G., A. Augustin, S. Musgrave, et al. 1989. "Evaluation of a Simplified Method for Estimation of Early Childhood Mortality in Small Populations." *International Journal of Epidemiology*, 18(4-Suppl.2): s20-s32.
- Black, R.E., K.H. Brown, S. Becker, and M. Yunus. 1982. Original Contributions. "Longitudinal Studies of Infectious Diseases and Physical Growth of Children in Rural Bangladesh. I. Patterns of Morbidity." *American Journal of Epidemiology*, 115(3): 305-13.
- Bledsoe, C., D. Ewbank, and U. Isiugo-Abanihe. 1988. "The Effect of Child Fostering on Feeding Practices and Access to Health Services in Rural Sierra Leone." *Social Science and Medicine*, 27(6): 627-36.

- Briscoe, J., J. Baltazar, and B. Young. 1988. "Case-control Studies of the Effect of Environmental Sanitation on Diarrhea Morbidity: Methodological Implications of Field Studies in Africa and Asia." *International Journal of Epidemiology*, 17(2): 441-47.
- Buck, C., A. Llopis, E. Najera, et al. 1988. "The Challenge of Epidemiology: Issues and Selected Readings." Pan American Health Organization, Scientific Publication No. 505.
- Carlson, B. 1985. "The Potential of National Household Survey Programs for Monitoring and Evaluating Primary Health Care in Developing Countries." *World Health Statistics Quarterly*, 38: 1.
- Carstairs, V., and M. Lowe. 1986. "Small Area Analysis: Creating an Area Base for Environmental Monitoring and Epidemiological Analysis." *Community Medicine*, 8(1): 15-28.
- Clemens, J.D., and B.F. Stanton. 1987. "An Educational Intervention for Altering Water-Sanitation Behaviors to Reduce Childhood Diarrhea in Urban Bangladesh. I. Application of the Case-Control Method for Development of an Intervention." *American Journal of Epidemiology*, 125(2): 284-91.
- Clemens, J.D., B.F. Stanton, J. Chakraborty, et al. 1988. "Measles Vaccination and Childhood Mortality in Rural Bangladesh." *American Journal of Epidemiology*, 128(6): 1,330-39.
- Cohen, A. 1989. "Urban Unfinished Business." *Journal of Public Health Policy*, 214-21.
- Coreil, J., A. Augustin, E. Holt, et al. 1989. "Use of Ethnographic Research for Instrument Development in a Case-control Study of Immunization in Haiti." *International Journal of Epidemiology*, 18(4): s33-s37.
- Crockett, S., K. Heller, J. Merkel, et al. 1990. "Assessing Beliefs of Older Rural Americans about Nutrition Education: Use of the Focus Group Approach." *Journal of the American Dietary Association*, 90: 563-67.
- CUS (Centre for Urban Studies). 1989. "The Urban Poor in Bangladesh: Phase 1: Sector: Comprehensive Summary Report." Prepared for the GOB/UNICEF/World Bank.
- Cutting, W.A.M., and K. Gopa. 1988. "Children in Third World Slums." *British Medical Journal*, 296: 1,683-84.
- Cutts, F. 1988. "The Use of the WHO Cluster Survey Method for Evaluating the Impact of the Expanded Programme on Immunization on Target Disease Incidence." *Journal of Tropical Medicine and Hygiene*, 91: 231-39.
- Dalkey, N.C. 1969. *The Delphi Method: An Experimental Study of Group Opinion*. Memorandum RM-5888-PR. Santa Monica, Ca.: Rand Corporation.

- Delbecq, A.L., and A.H. Van de Ven. 1971. "A Group Process Model for Problem Identification and Program Planning." *The Journal of Applied Behavioral Science*, 7(4): 466-92.
- Dodge, H.F., and H.G. Romig. 1959. *Sampling Inspection Tables (2nd ed.)*. New York: John Wiley and Sons.
- Engle, P. 1989. "Women, Work, and Child Welfare in the Third World." AAAS Selected Symposium 110. *American Association for the Advancement of Science*, 179-200.
- Fassin, D., and E. Jeannee. 1989. "Immunization Coverage and Social Differentiation in Urban Senegal." *American Journal of Public Health*, 79(4): 509-11.
- Fosu, G.B. 1989. "Access to Health Care in Urban Areas of Developing Societies." *Journal of Health and Social Behavior*, 30: 398-411.
- Frerichs, R. 1988. "Rapid Microcomputer Surveys." *Journal of Tropical Pediatrics*, 34: 147.
- Frerichs, R. 1989. "Simple Analytic Procedures for Rapid Microcomputer-Assisted Cluster Surveys in Developing Countries." *Public Health Reports*, 104(1): 24-35.
- Frerichs, R., and R. Miller. 1985. "Introduction of a Microcomputer for Health Research in a Developing Country—the Bangladesh Experience." *Public Health Reports*, 100(6): 638-47.
- Frerichs, R., and K. Tar. 1988a. "Use of Rapid Survey Methodologies to Determine Immunization Coverage in Rural Burma." *Journal of Tropical Pediatrics*, 34: 125-30.
- _____. 1988b. "Breast Feeding, Dietary Intake and Weight-for-Age of Children in Rural Burma." *Asia-Pacific Journal of Public Health*, 2: 16-21.
- _____. 1989. "Computer Assisted Rapid Surveys in Developing Countries." *Public Health Reports*, 104(1): 14-23.
- Gould, J., and R. Frerichs. 1986. "Training Faculty in Bangladesh to Use a Microcomputer for Public Health: Follow-up Report." *Public Health Reports*, 101(6): 616-23.
- Grant, J., M. Kinney, and C. Guzzetta. 1990. "A Methodology for Validating Nursing Diagnoses." *Advances in Nursing Science*, 12(3): 65-74.
- Grosse, R.N. 1980. "Interrelation between Health and Population: Observations Derived from Field Experiences." *Social Science and Medicine*, 14C(2): 99-120.
- Hansluwka, H. 1985. "Measuring the Health of Populations, Indicators and Interpretations." *Social Science and Medicine*, 20(12): 1,207-24.
- Henderson, R., and T. Sundaresan. 1982. "Cluster Sampling to Assess Immunization Coverage: A Review of Experience with a Simplified Sampling Method." *Bulletin of the World Health Organization*, 60(2): 253-60.

- HKI (Helen Keller International Institute of Public Health Nutrition). 1985. "Bangladesh Nutritional Blindness Study 1982-1983." Dhaka, 2-11.
- Islam, A.B.M.Q., A.K. Siddique, K.A.H.M. Akram. 1991. "A Hospital Outbreak of Cholera in Dhaka, Bangladesh." *Tropical and Geographic Medicine*, 43(2):117.
- Jalil, F., J. Karlberg, L.A. Hanson, and B.S. Lindblad. 1989. "Growth Disturbance in an Urban Area of Lahore, Pakistan Related to Feeding Patterns, Infections and Age, Sex, Socio-Economic Factors and Seasons." *Acta Paediatr Suppl*, 350: 44-54.
- Jaramillo, A., R. Dios, and A. de Quilcate. 1989. "Respuesta a la Tuberculina en Escolares de Pacasmayo, Peru." *Bol of Sanit Panam*, 107(5): 388-95.
- Kalter, H., R. Gray, R. Black, et al. 1990. "Validation of Postmortem Interviews to Ascertain Selected Causes of Death in Children." *International Journal of Epidemiology*, 19(2): 380-86.
- Khanjanasthiti, P., and J.D. Wray. 1974. "Early Protein-Calorie Malnutrition in Slum Areas of Bangkok Municipality, 1970-1971." *Journal of the Medical Association of Thailand*, 57(7): 366-75.
- King, C., C. Chen, S. You, et al. 1989. "Community-wide Epidemiological Investigation of a Typhoid Outbreak in a Rural Township in Taiwan, Republic of China." *International Journal of Epidemiology*, 18(1): 254-60.
- Koenig, M., Kahn M.A., Wojtyniak B., et al. 1990. "Impact of Measles Vaccination on Childhood Mortality in Rural Bangladesh." *Bulletin of the World Health Organization*, 68(4): 441-47.
- Kok, P. 1986. "Cluster Sampling for Immunization Coverage." *Social Science and Medicine*, 22(7): 781-83.
- Kroeger, A. 1983. "Health Interview Surveys in Developing Countries: A Review of Methods and Results." *International Journal of Epidemiology*, 12(4): 465-81.
- Lanata, C.F., G. Stroh, R.E. Black, and H. Gonzales. 1990. "An Evaluation of Lot Quality Assurance Sampling to Monitor and Improve Immunization Coverage." *International Journal of Epidemiology*, 19(4): 1086-90.
- Lapau, B. 1987. *Epidemiology of Treatment Seeking Behavior in Lilirilau Subdistrict Soppeng Regency South Sulawesi*. Unpublished doctoral thesis.
- Lemeshow, S., and D. Robinson. 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): 65-75.
- Lemeshow, S., and G. Stroh. 1988. *Sampling Techniques for Evaluating Health Parameters in Developing Countries*. Board on Science and Technology for International Development, Working Paper, Washington, D.C.: National Academy Press.

- _____. 1989. "Quality Assurance Sampling for Evaluating Health Parameters in Developing Countries." *Survey Methodology*, 15(1): 71-81.
- Lemeshow, S., A.G. Tserkovnyi, J.L. Tulloch, et al. 1985. "A Computer Simulation of the EPI Survey Strategy." *International Journal of Epidemiology*, 14(3): 473-81.
- Lobet, M., A. Stroobant, R. Mertens, et al. 1987. "Tool for Validation of the Network of Sentinel General Practitioners in the Belgian Health Care System." *International Journal of Epidemiology*, 16(4): 612-18.
- Lopez de Romana, G., K.H. Brown, R.E. Black, and H.C. Kanashiro. 1989. "Longitudinal Studies of Infectious Diseases and Physical Growth of Infants in Huascar, an Underprivileged Peri-Urban Community in Lima, Peru." *Journal of Epidemiology*, 129(4): 769-84.
- Lwanga, S., and N. Abiprojo. 1987. "Immunization Coverage Surveys: Methodological Studies in Indonesia." *Bulletin of the World Health Organization*, 65(6): 847-53.
- Minassian, D.C., V. Mehra, and J.D. Verrey. 1989. "Dehydrational Crises: A Major Risk Factor in Blinding Cataract." *British Journal of Ophthalmology*, 73: 100-05.
- Moskowitz, H., and H. Tsai. 1988. "A One-sided Double Screening Procedure Using Individual Unit Misclassification Error." *Management Science*, 34(9): 1,139.
- Mosley, W.H., and M. Khan. 1979. "Cholera Epidemiology—Some Environmental Aspects." *Prog Water Tech*, 11(1/2):309-316.
- Mosteller, F. 1987. "Assessing Quality of Institutional Care." *American Journal of Public Health*, 77(9): 1,155.
- Okun, D.A. 1988. "The Value of Water Supply and Sanitation in Development: An Assessment." *American Journal of Public Health*. 78: 1,463-67.
- Patel, R.B. 1989. "Care of Low Birth Weight Babies in Slums." *Indian Journal of Pediatrics*, 56: 231-37.
- Pearson, R. 1989. "Rapid Assessment Procedures Are Changing the Way UNICEF Evaluates Its Projects." *Hygie*, 8(4): 23-25.
- Phillips, J.F., A.B.M.K. Mozumder, D. Leon, and M. Koenig. October 1989. "The Application of Micro-Computer Technology to Longitudinal Studies of Survival: Lessons from a Field Study in Bangladesh." Presented at Community-Based Health Research in Mexico: A Proposal for the Initial Stage. NIPH, Mexico.
- PRICOR Report. June 1989. "The Evolution of PRICOR's Operations Research Approach." Bethesda, Md.
- Reinke, W.A. March 1988. *Industrial Sampling Plans: Prospects for Public Health Applications*. Occasional Paper No. 2. Baltimore: The Johns Hopkins University, Institute for International Programs.

- Richards, F., and D. Hopkins. 1989. "Surveillance: The Foundation for Control and Elimination of Dracunculiasis in Africa." *International Journal of Epidemiology*, 18(4): 934-43.
- Rietveld, C. November 1988. "Water Supply and Sanitation in Fast Growing Cities." Paper presented in the inaugural session of the Collaborative Council, The Hague, Netherlands.
- Rodrigues, L., and B. Kirkwood. 1990. "Case-control Designs in the Study of Common Diseases: Updates on the Demise of the Rare Disease Assumption and Choice of Sampling Scheme for Controls." *International Journal of Epidemiology*, 19(1): 205-13.
- Rohde, J.E. 1983. "Why the Other Half Dies: The Science of Politics of Child Mortality in the Third World." *Assignment Children*, 61/62: 35-7.
- Rosero-Bixby, L., G. Stroh, C. Grimaldo, et al. July 5, 1988. "Lot Quality Acceptance Sampling for Monitoring Primary Health Care Programs." Unpublished document.
- Ross, D., and J. Vaughan. 1986. "Health Interview Surveys in Developing Countries: A Methodological Review." *Studies in Family Planning*, 17(2): 78-94.
- Rothenberg, R., A. Lobanov, K. Singk, et al. 1985. "Observations on the Application of EPI Cluster Survey Methods for Estimating Disease Incidence." *Bulletin of the World Health Organization*, 63(1): 93-9.
- Sapire, S. 1989. "Rapid Evaluation Methodology (REM) Maternal/Child Health and Family Planning." Unpublished paper.
- Shuval, H., R. Tilden, B. Perry, et al. 1981. "Effect of Investments in Water Supply and Sanitation on Health Status: A Threshold Saturation Theory." *Bulletin of the World Health Organization*, 59(2): 243-48.
- Singh, M., and V.K. Paul. 1988. "Strategies to Reduce Perinatal and Neonatal Mortality." *Indian Pediatrics*, 25: 499-509.
- Smith, G. 1989. "Development of Rapid Epidemiologic Assessment Methods to Evaluate Health Status and Delivery of Health Services." *International Journal of Epidemiology*, 18(4): s2-s15.
- Stanton, B.F., and J.D. Clemens. 1987. "An Educational Intervention for Altering Water-Sanitation Behaviors to Reduce Childhood Diarrhea in Urban Bangladesh. II. A Randomized Trial to Assess the Impact of the Intervention on Hygienic Behaviors and Rates of Diarrhea." *American Journal of Epidemiology*, 125(2): 292-301.
- _____. 1988. "The Influence of Gender on Determinants of Urban Childhood Mortality in Bangladesh." *International Journal of Epidemiology*, 129-35.
- Stanton, B.F., J.D. Clemens, and S. Ahmed. 1990. "Methodological Considerations in Defining Chronic Diarrhoea Using a Distributional Approach." *International Journal of Epidemiology*, 19(2): 439-43.

- Stanton, B.F., J.D. Clemens, and T. Khair. 1988. "Educational Intervention for Altering Water-Sanitation Behavior to Reduce Childhood Diarrhea in Urban Bangladesh: Impact on Nutrition Status." *American Journal of Clinical Nutrition*, 48: 1,166-72.
- Stanton, B., D.R. Silimperi, K. Khatun, et al. 1989. "Parasitic, Bacterial and Viral Pathogens Isolated from Diarrhoeal and Routine Stool Specimens of Urban Bangladeshi Children." *Journal of Tropical Medicine and Hygiene*, 92: 46-55.
- Sudre, P., J. Breman, and J. Koplan. March 24, 1990. "Delphi Survey of Malaria Mortality and Drug Resistance in Africa." *Lancet*, 722.
- Tekoe, B. 1990. "Households, Resources, and Child Health in a Self-Help Settlement in Cairo, Egypt." *Social Science and Medicine*, 30(8): 929-40.
- Thompson, S.K. 1990. "Adaptive Cluster Sampling." *Journal of the American Statistical Association*, 85(412): 1,050-59.
- Timaeus, I. 1986. "An Assessment of Methods for Estimating Adult Mortality from Two Sets of Data on Maternal Orphanhood." *Demography*, 23(3): 435.
- Timaeus, I., T. Harpham, M. Price, et al. 1988. "Health Surveys in Developing Countries: The Objective and Design of an International Programme." *Social Science and Medicine*, 27(4): 359-68.
- Trachtenberg, A., A. Washington, and S. Halldorson. 1988. "A Cost-based Decision Analysis for Chlamydia Screening in California Family Planning Clinics." *Obstetrics and Gynecology*, 71(1): 101.
- Tuomilehto, J., and P. Puska. 1987. "The Changing Role and Legitimate Boundaries of Epidemiology: Community Based Prevention Programs." *Social Science and Medicine*, 25(6): 589-98.
- Valadez, J. March 1988. "A Test for Lot Quality Acceptance Sampling for Monitoring Primary Health Care Coverage and Costs." Preproposal submitted to the Thrasher Research Fund.
- Walton, M. 1986. *The Deming Management Method*. New York: Perigee Books.
- Weekly Epidemiology Record. 19 June 1987. "Expanded Programme on Immunization (EPI), Immunization Coverage Survey." 25: 183-85.
- Wheeler, D.J., and D.S. Chambers. 1986. *Understanding Statistical Process Control*. Knoxville, Tn.: SPC, Inc.
- Wiquhart, G. "Lot Quality Assurance Sampling (LQAS) for Monitoring Health Care Practices." Unpublished paper.
- World Health Organization. 1981. *Development of Indicators for Monitoring Progress towards Health for All by the Year 2000*. Health for All Series No. 4, Geneva: WHO.

World Health Organization. 1985. *Primary Health Care in Urban Areas: Reaching the Urban Poor in Developing Countries.* A State-of-the-Art Report by UNICEF and WHO. UNICEF.

Zaman, S.S., N.Z. Khan, S. Islam, et al. 1990. "Validity of the 'Ten Questions' for Screening Serious Childhood Disability: Results from Urban Bangladesh." *International Journal of Epidemiology*, 19(3): 613-20.

Mathematical Appendix

SAMPLE SIZE DETERMINATION

The amount of information needed to draw meaningful conclusions about the presence of a particular condition of interest is dependent upon three factors as follows.

- R - is the *rarity* of the event of interest, i.e., the ratio of noncases to cases. For example, if six "well" children are found for every "case" of diarrhea, then R is 6.
- T - is the maximum *tolerable* error as a proportion of the true prevalence rate. For example, if a 50-percent prevalence rate is to be estimated with an error of not more than 5 percent, T is 0.1.
- D - is a statistically derived factor based upon how *definitive* the conclusions are to be. For example, if there is to be 95 percent assurance (confidence) that the tolerable error is not exceeded, statistical tables specify that D is 4.

The sample size (n) required to accommodate specified values of these factors is given as:

$$n = DR/T^2.$$

Decision makers typically consider that 95-percent confidence is adequate in making statements about the precision of a sample estimate. Furthermore, as a rule of thumb herein, we require that the risk of errors of commission not exceed 10 percent and that the risk of errors of omission not exceed 5 percent. These decision criteria translate into D constants as indicated below.

D	Decision Situation
4.0	To estimate the prevalence of the condition of interest in one population
8.0	To estimate the difference between two populations in the prevalence of the defined condition
8.6	To assess the significance of an observed departure from a specified standard in a population
17.2	To assess the significance of an observed difference between two populations in the prevalence of a defined condition

A benchmark situation is usefully defined as that in which a true prevalence of 50 percent is yestimated with maximum error of 10 percent, for then $R=1$, $T=0.2$, and

$$\begin{aligned}
 n &= (4.0)R/(0.4) \\
 &= 100R \\
 &= 100.
 \end{aligned}$$

Using the benchmark situation as the starting point, the necessary sample size can be modified according to the rarity of the event of interest. To illustrate, if every fourth observation is expected on the average to produce a "case" (i.e., three noncases to one case), then $R=3$ and

$$\begin{aligned}
 n &= 100R \\
 &= 300.
 \end{aligned}$$

In contrast, a condition found in only one percent of the population would call for a sample of 9,900 (i.e. $100 \times (99/1)$). Thus we see the difficulty and high cost of assessing the incidence of rare conditions with any precision.

The illustration applies to the first decision situation cited in the table above. Other, more complex decision situations relate to the benchmark case as indicated in the following table.

Selecting Sample Sizes

Decision	Situation	Nr. of Observations	
		Each Pop.	Total
Estimation	One population; no prior standards	100R*	100R
Estimation: difference between two populations	Two populations; no prior standards	200R	400R
Significance testing	One population; prior standards	215R	215R
Significance testing: difference between two populations	Two populations; prior standards	430R	860R
Significance testing: difference between two populations in departure from standard	Two populations; prior standards	860R	1720R

*R = "rarity," i.e., the ratio of noncases to cases
OBSERVE: Testing for the significance of differences between populations, given prior standards, requires more than eight times as much data as are needed for a simple estimate of conditions in a single population.