SAMPLING MANUAL for FACILITY SURVEYS

For Population, Maternal Health, Child Health and STD Programs in Developing Countries

Anthony G. Turner
Gustavo Angeles
Amy O. Tsui
Marilyn Wilkinson
Robert Magnani

MEASURE *Evaluation* Project

COMPLETE PRELIMINARY VERSION
Version 5
MAY 2000

Table of Contents

Chapter 4 (continued)

Chapter 1
PURPOSES OF FACILITY SURVEYS

This chapter discusses
- # monitoring and evaluation
- # measurement objectives and indicators
- # sampling implications, and
- # data collection implications

Introduction

The purpose of this manual is to present a sampling methodology that can generate estimates of health facilities and their characteristics, and, when desired, tie the characteristics of the sampled facilities to those of the serviced population in a meaningful way. Two sampling designs are proposed and recommended in this manual. The first is for a stand-alone health facility survey (chapter 4), and the second (chapter 5) is for a health facility survey linked to a household survey. The design for the latter requires adopting the same sample areas used to generate household data collected in surveys such as Demographic and Health Surveys or the Reproductive Health Surveys[1]. Both recommended sampling designs provide unbiased estimates of facilities and their characteristics; the linked sampling design, however, provides additional information on the health service environment for resident populations in the household survey sample areas.

The linking of the facility survey to a household offers powerful analytic value for investigating how the presence of health facilities and their services can influence health practices and behaviors of the local populations. However, linking with a household survey sample design does constrain the facility sample plan in several ways. Accordingly, the sample design given in chapter 4 is better in terms of the facility estimates *per se*. It should be applied in countries where the primary objective is only to estimate the distribution of health facilities and their characteristics. If, on the other hand, one of the objective is also to evaluate the impact of large-scale public health interventions on population outcomes, then the second sampling strategy which links the facilities surveyed to a population sampled through a household survey's sample areas, is recommended.

---

[1]See also the DHS manual on sampling - Macro International, Inc. 1996. *Sampling Manual*. DHS-III Basic Documentation No. 6. Calverton, MD.

Analytical perspectives

In considering the sampling needs for facility surveys, it is useful to examine the analytical strategies employed for program evaluation. The effects of interest may be systemic, that is, evaluating and attributing change in health service delivery systems to program improvements, such as for clinical training, commodities logistics management, or infection prevention procedures. Likewise, the effects of interest may be client- or population-based, such as increasing client satisfaction with services or reducing unwanted pregnancies or disease morbidity and mortality. Illustrative hypotheses of systemic outcomes investigated in program evaluation are:

- Health facilities experiencing frequent stockouts have lower client case loads.

- Staff competence influences the facility's ability to provide services at a minimum standard.

- Patients/clients receiving quality services during any consultation are more likely to return for continued care.

Illustrative hypotheses of population-level outcomes that could be investigated are:

- Access to contraceptive services through multiple providers increases the probability of continued use.

- Integrated management of sick children at health facilities reduces infant and child mortality.

- Screening, diagnosis and treatment of STD in prenatal clients can reduce transmission rates.

In terms of sampling, analyses for the first three illustrative hypothesis above require having a probability sample of facilities, their staff and clients, to which relevant survey questions would be administered. Analyses for the last three hypothesis require a sample design that enables information about population behaviors and outcomes to be correlated in some fashion with the information about the provision of health services.

Perhaps the easiest way to envision the units that would need to be considered as part of a sample design is to identify first what population is of interest – health facilities, staff at health facilities, pharmacies or drug retailers, mobile or community-based staff, current or past clients, or populations at risk, whether these are delimited by age, gender or place of residence. This determination helps a sampling statistician develop a protocol that will produce estimates with a minimum of bias and a maximum of precision. The decisions are facilitated in the monitoring and evaluation context where specific indicators of performance are focused. These indicators are generally defined in terms of a characteristic or attribute present in a given population that is measured and monitored for change over time, thereby presenting information from which likely causes can be speculated. An indicator such as the percent of health facilities that have 2 or more staff on site trained in post-abortion care signals the need for a sample of health facilities and information on service-specific training taken from staff records or interviews with staff assigned to the facilities. Decisions about these needs are discussed in more detail in the next section.

Measurement objectives of facility surveys

While the focus of this manual is sampling for facility surveys, the latter cannot be considered apart from the overall survey methodology that is needed to collect data on facility variables and related measures. In turn, the survey methods depend on the measurement objectives, or, in short, what it is we wish to find out. Hence the first step (and requirement) in this or any other survey under-taking is clear and unambiguous specification of those measurement objectives. In statistical parlance they are simply the *estimates* we want. For a facility survey, they may best be seen by considering the substantive items of inquiry in a generic form, as detailed in the subsection immediately following. The survey methods that are necessary to collect the various measures provide a second way of categorizing the *particular* indicators of interest and these are described in a later subsection, entitled, "Indicators and Collection Method."

Basic items of inquiry from the facility survey

For facility surveying of the kind we are treating in this manual, there is a comparatively small number of items of inquiry that differ, fundamentally, from each other in the operational terms that define them for measurement. The short list which follows is thought to be exhaustive.

Note that we distinguish below between a facility, a health care provider, and facility-based staff. The nature of public health programs involves both clinical and nonclinical settings of service delivery. Facilities are self-standing sites of service delivery, such as hospitals, health centers and health posts. Health care providers may be both clinic-based and non-clinic based, such as traditional birth attendants, lay health workers, or retail outlets which dispense medication. Facility-based staff are personnel assigned to facilities who provide on- or off-site/outreach health services and care.

1. Proportion (or percent) of total facilities offering a given service by
    a. Type of facility
        i. public, private
        ii. level (primary, secondary, tertiary)
        iii. service type and range
    b. Compliance with established service protocols or standards
2. Proportion of health care providers offering a given service by

a. Type of providers
    i. mobile, community-based, facility-based
    ii. public, private
b. Service mix
c. Characteristics of providers (location, skills training, gender, etc.)

3. Proportion of facility staff performing a given function or service
    a. By type of staff (physician, nurse, paramedic, etc.)
    b. By characteristics of staff (age, gender, years of service, etc.)
4. Proportion of facility clients receiving a given service, therapy or exposed to an intervention
    a. By type of facility.
    b. By characteristics of clients (age, gender, etc.)

Several observations may be made about these *generic* indicators. First, every specific indicator that has so far been promulgated by USAID and others can be said to fall into one of the generic sets. Lists of indicators are ever-changing, however, and individual countries will no doubt have their own additions or deletions to reflect their health priorities. Second, some of the specific indicators are means or averages (for example, mean distance of clients to nearest facility), but they are, nevertheless, indistinguishable from percentages insofar as the survey and estimation methods are concerned. Third, while there are only a few *primary* types of indicators, it is necessary that those shown as sub-categories be distinguished separately because of their importance for sample design.

Other data - other sources

Evaluation design relies upon a combination of data types from different sources including the facility survey items discussed above. For comprehensive analytical study of relationships between service provision and targeted populations, data from other sources may be needed as well. These may include quantitative data from population surveys, epidemiologic surveillance systems, and community-level information, as well as qualitative data gathered from key informants[3].

While other types of information may be needed in the overall research design for an evaluation of program impact, *they are not all collected in the facility survey*. Population data come from household surveys such as the aforementioned DHS or RHS. Examples of variables or indicators from household surveys are the contraceptive prevalence rate; percent of women ages 20-24 who have had a first birth before the age of 20; the percent of infants under 6 months of age exclusively breastfed; proportion of births spaced 24 months or longer; percent of men or women ages 15-49 who report the use of a condom during the most recent act of sexual intercourse; and the proportion of mothers whose last birth was attended by a trained professional.

---

[3]Qualitative data are not discussed here.

Epidemiologic surveillance systems record the incidence of diseases with major public health import, such as STD, HIV/AIDS, malaria, tuberculosis, polio, or child communicable diseases. Assessing the impact of a public health program focused on one or more of these types of infections can benefit from clinical confirmation and recording of each known infection, as well as its spatial location. These data offer the ability to track outbreaks both as spatial clusters and transmission over time through spatial mapping procedures.

Community-level variables may also be needed to describe the context of the social and economic environment in which individuals reside. These data include information about the community infrastructure and resources, such as schools, roads and transportation and communication systems, agricultural and retail markets, local businesses and industries for labor markets, and water supply and safety. Information is usually gathered on availability of and access to these community resources (measured usually in distance or travel time). It is helpful to assess also the community presence of or its participation in other development programs, such as in education, agriculture, housing, or employment. These may provide competing demands on local public officials' time toward implementing government public health initiatives.[4]

The combination of information of the type described above offers a comprehensive means of evaluating the actual effect of national or large-scale public health program efforts. It does so by isolating, through data collected in a health facility survey and any epidemiological surveillance, the impact of service provision improvements on population-level health outcomes net of effects from community-level improvements.

Indicators and collection method

For purposes of statistical measurement, a typography of the specific indicators from a facility survey, as opposed to the generic sets discussed two earlier subsections, can be generated in relation to the origin of the information source. There are four such sources, including interviews with facility spokespersons (plus reviews of facility inventories), observations by interviewers, interviews with staff, interviews with clients, as follows:

- Service site, or facility, indicators. These are obtained through interviews with spokespersons from the facility and/or through direct examination of stocks and inventories. The indicators are generally of the form, percent of health facilities offering a specific service or having certain necessary items. Representative examples include percent offering family planning methods, percent providing essential obstetric care, percent counseling prenatal clients about STD/HIV risk, percent maintaining various records, and percent with certain equipment or materials.

[4]If community-level data are gathered, their data collection can be easily integrated with a health facility or population survey since such information is likely to relate geographically to boundaries of sample areas used for either type of survey.

- Staff indicators. These are obtained through personal interviews with staff members. The indicators are usually expressed in the form of percent of staff with specific characteristics. Representative examples include percent of staff trained in child care, percent of staff reporting they would not recommend certain [family planning] methods, percent receiving training in maternity-related services in last 3 years, and percent who are able to provide HIV/AIDS counseling, diagnosis and treatment.

- Client indicators. These are obtained through personal interviews with the clients. The indicators are generally of the form, percent of clients receiving a given procedure or intervention. Representative examples include percent of clients satisfied with duration of family planning consultation, percent of caretakers of sick children who were sufficiently satisfied with the care received that they would encourage others to bring in a sick child, and percent who understood the information given about treatment side effects.

- Client-staff interaction indicators. These are obtained through observation by an interviewer who is present during the client's session with the service provider. The indicators are generally of the form, percent of clients receiving a given staff-initiated procedure or intervention during consultation. Representative examples include percent of family planning clients treated respectfully by staff, percent of clients who are asked if they prefer a particular method, percent of those receiving pelvic examinations that have the procedure explained to them, percent of sick children who have their weight checked using a growth chart, and percent of sick children assessed for presence of cough, diarrhea and fever.

Measurement form of the estimates - facilities, staff and clients

The desired estimates from a facility survey are straightforward to define and measure for facilities and staff, but they are somewhat more complex for clients, both conceptually and statistically. As discussed above, most of the facility estimates will be characteristics or attributes, expressed as *percentages* such as percent with certain equipment. However, estimates of *totals* such as total number of facilities offering maternal health services in region A may easily be obtained from a survey designed expressly for facility measurement, and, in most country applications, such totals would likely be tabulated. Similarly, staff estimates of both forms, such as (a) *percentage* of health workers trained in maternity-related care in the last 3 years or (b) *total* number of urban health workers in family planning, may be tabulated from the survey. By the same token, the estimates may be in terms of averages or means, such as the average number of staff with a particular attribute for a specific type of facility or the average number of health facilities per sample area.

The situation is different, however, for client estimates. While the percentage of clients receiving a specific service is wanted, it is understood that the percentage must be expressed in the context of clients visiting facilities. Moreover, in order to measure it statistically, it must be within a specified time frame. It will not be possible for the survey to estimate the *total* number of clients for facilities, nor is the concept itself meaningful when a time dimension is absent. In other words, while an estimate (or the concept) of the number of clients who visit facilities daily (or weekly, monthly, annually) makes sense statistically and conceptually, a timeless estimate of total facility clients is basically meaningless. Thus, to solve the measurement problem, estimates of the denominator of a percentage such as percent of family planning clients receiving a friendly greeting must be made in terms of *client-visits in a specified time period* (see chapters 4 and 7).

Geographic levels

A crucial dimension of the measurement objectives for the facility survey concerns the geographic or administrative sub-divisions for which the estimates are desired. As mentioned above, monitoring and/or evaluation plans may be designed to study the provision of health services in a general way for the country or a sub-region, or they may be designed to assess a particular action program. In the case of the latter, the survey would be confined to the geographic areas in which the action program is implemented and any chosen comparison (non-treatment) area. Often this may be a very limited geographic area such as a small set of villages or a single city, in which case the survey estimates would be restricted to the program and non-program areas as a whole, without any attempt to provide separate estimates, say, for each village or parts of the city. When the program and comparison areas are large, however, such as a state or province, project personnel are likely to want estimates disaggregated for important sub-domains, such as districts or communes, for comparison purposes. These so-called *estimation domains* must be clearly specified when the survey is being planned so that the sample design can take account of it.

When the facility survey is general-purpose in nature, that is, not confined to a particular area where an action program is being carried out, it is more likely to be sizable in scope, often covering the entire country. For this case, too, project personnel will undoubtedly require domain estimates (major regions, urban, rural, provinces or states), and the domains must be specified during the survey planning process for proper sampling strategies to be developed.

Estimates of change

It should be observed that evaluation of a program impact often requires estimating change in an individual health behavior (for example, whether and by how much the contraceptive prevalence rate has increased). This rather fundamental but crucial aspect of the evaluation and measurement process is not discussed in this manual because the estimates of change relate to a <u>population</u> variables rather than a program or service level variable measured through the facility survey. On the other hand, change relating to the performance measures of the facilities themselves, such as increase in service quality or utilization, is a necessary dimension of facility surveying in many applications. A separate chapter is devoted to the sampling considerations to estimate change or trends at the service or health facility level (chapter 6).

Implications for sampling

Sampling is used, in various ways, to meet the measurement objectives for a facility survey. Here we will discuss the indicators in terms of their implications for sampling, but *not* the problems, which are taken up in chapter 2.

Referring back to the listed items of inquiry, the first one, "proportion of total facilities offering a given service," requires selecting a valid[5] sample of facilities, or service sites, from the total universe of such facilities. The universe of facilities may be restricted to a sub-region of a country or to a particular action program area (geographically defined), depending upon whether the estimation objective relates to the country as a whole, a sub-region or the specified program area. An initial sample of facilities may necessitate screening to establish which ones offer the services which are of particular interest in this manual, namely, family planning, maternal and child health, or sexual health (STD/HIV/AIDS). This requirement may thus entail two phases of sampling -- the first to find out which facilities provide the targeted services (plus collection of some limited information about the ones that do not), and the second to select a sub-sample of the facilities offering the target services, where the more detailed survey questionnaire would be administered. In addition to the substantive questions which are put to the spokesperson of a given sample facility, questions about number and type of staff, client volume and days and time of operation will have to be included to help the sampler select clients and staff.

The third item, "proportion of facility staff performing a given function or service," requires selecting a valid sample of the health-related personnel working for the facility. For small facilities with only a few staff, all of them may be interviewed, however. Staff in large facilities (those with large numbers of staff) would, most likely, be sampled after first establishing categories, or strata, by type (physicians, nurses, technicians, etc.) and/or size.

The fourth of the basic items of inquiry, "proportion of facility clients receiving a given service, therapy or exposed to an intervention," necessitates selecting a valid sample of the clients of the facility. It is to be noted that the population of clients is not the same as the population living within the facility's catchment area, as it constitutes, by definition, only those persons who visit the facility or otherwise receive its services -- a biased or selected subset. It is perhaps obvious that sample selection of clients presupposes that the facility itself has been selected into the sample, although in theory this need not be so. Cost-saving demands, however, would make it infeasible to select a client (or staff) sample *independently* from the facility

---

[5]When referring to a *valid* sample, we mean one that adheres strictly to the tenets of probability selection methods at every stage of the process (see chapter 3).

sample.[6]

Selection of staff and clients will be as a result of sub-samples from the sampled facilities to which they are associated. Sub-sampling, by definition, means the survey data for staff and clients will require weighting for unbiased estimation (weighting will be necessary to obtain *totals* for facility estimates also). This, as such, does not pose a theoretical problem so long as the sample design is probability-based and the probabilities of selection, and corresponding weights, are properly calculated.

Implications for data collection procedures

Implications for the data collection methodology beyond the sampling methods are discussed in this section. Capturing client-staff interactions and obtaining supplemental data for facility estimation when linked with population survey sample areas are two of the most important procedures.

In terms of field implementation, capturing the requisite information for the clients of facilities in a feasible procedure that is methodologically sound is an issue. Interviewers must visit the facility over a period of hours or days and (a) observe staff-client interactions and (b) conduct an exit interview with clients. Both would be done on a sample basis, except in small facilities that have few clients, where all would be interviewed. From the standpoint of statistical analysis, it is highly desirable (if not mandatory) that the observations and interviews cover the *same* clients. A strategy that departs from this ideal is one for family planning clients where it is thought that the desired information on services is best gathered from observing client-staff interaction of new family planning clients only but interviewing all (new and old) family planning clients upon departure.

Approaches for client-staff observation that have been tried include those, say, with sick children under 5 years old, in which one health worker is followed throughout the day and all his/her patients are observed. This controls for variation in health worker performance across observations but the method does not yield an unbiased sample of clients, which is the desired unit of analysis in this case. As a result, facility assessment of care is biased with this approach. Other methods observe eligible patients or clients on an as-come basis, which is unbiased with regard to client estimates, but is difficult to implement in large facilities where multiple clients show up simultaneously.

For a facility survey that is designed to cover facilities in the same -- or nearby -- sampling areas (primary sampling units, or clusters) of a population survey, construction of the

[6]One exception to this is through population-based surveys which would ask individual respondents if they have visited a health facility for health care in the past month, thereby identifying them as clients.

facility estimates requires that population figures for areas surrounding the sampled facility be obtained. Depending upon the survey design that is used, it may also be necessary to collect supplemental data from the surrounding areas to ascertain the proximity of other facilities. Accordingly, the techniques and procedures take on varying degrees of complexity in field implementation, depending upon the exact nature of the survey/sampling methodology used and these are discussed in chapter 5.

Historical overview of facility surveys

Surveys of facilities offering health and/or family planning services have been undertaken in conjunction with large-scale population surveys on an intermittent basis over the last 30 years or so. This section briefly summarizes the evolution of such surveys.

World Fertility Surveys (WFS)

The proposal for collecting community-level information that could be integrated with individual- and household-level data obtained from conventional household surveys originated in the World Fertility Survey program in a paper by Ronald Freedman in 1974.[7] The rationale for collecting community-level data was to provide a basis for measuring the effects of community-level factors on individual demographic behaviors, effects which individual-level data alone are ill-equipped to measure. Three sets of effects or relationships were of primary interest:[8]

- effects of the provision of family planning and health services on fertility and mortality,
- relationships between economic and social opportunities and fertility, and
- relationships among the three basic demographic variables (fertility, mortality and migration).

This initiative resulted in the development of a schedule of questions about the socioeconomic environment and family planning facilities available to women in the sample points selected. The WFS collected such community data in 17 countries, but concentrated on rural areas (15 of the 17 collected community data in rural areas only).

Demographic and Health Surveys (DHS)

Following the lead of the WFS, the Demographic and Health Survey program has utilized a special questionnaire, the Service Availability Module (SAM), to supplement data gathered in

---

[7]World Fertility Occasional Paper No. 9, 1974.

[8]John Casterline. 1987. "The Collection and Analysis of Community Data." In J. Cleland and C. Scott (eds) *The World Fertility Survey - An Assessment*. London: Oxford University Press.

the individual questionnaire on the availability of family planning and health services. The SAM questionnaire is aimed at collecting information about the facilities available to the population in the sample clusters or segments from which individual women are selected for interview with the standard core questionnaire. Two principal objectives were intended to be served by the collection of service availability data:

- provide a description of the facilities available to women in the country, and
- provide a basis for analyzing the relationship between availability of contraceptive supplies and contraceptive practice.

The first objective would take the form of statistical generalizations such as "68 percent of the women in this country live within 30 minutes of a family planning facility," or, "women typically have access to available family clinics only three days a week," or "the average cost of pills available to women in this country is per cycle," etc. Such descriptive information is of potential value to family planning professionals for program planning purposes. The health services data collected in the SAM questionnaire are subject to similar types of analysis as the family planning information. At the descriptive level, the data permit the estimation of parameters such as the proportion of children who live within 30 minutes of a hospital, clinic or other health service, or who have oral re-hydration services available, etc.

The second objective relates to the more analytical purpose of trying to determine how the availability of supplies relates to the adoption and use of contraception. The concept of availability entails not only the density of contraceptive supplies or the physical proximity of sources of supply to users, but also the components on convenience of access (measured in the SAM questionnaire by the length of time it takes to reach the clinic or the source, and by how often the facility is open), of the variety of methods available, their cost, and medical personnel available. When combined with the individual data on the perceived quality of services and other measures of knowledge of sources and the reputation of different methods, the objective data on availability theoretically should provide a profile of the family planning service environment. With regard to health, the service availability information can be linked with data from the individual questionnaires to address questions such as the relationship between the availability or oral re-hydration services and the women's knowledge of treatment and use of services for children who have had recent episodes of diarrhea. Similar types of analysis can be conducted in connection with the use and availability of prenatal and maternity services.

The decision to include this type of supplementary questionnaire in the DHS was primarily an outgrowth of deliberations about the measurement of the availability of family planning supplies and information. The section of the individual questionnaire on this topic was the subject of much debate in the early days of questionnaire development. Throughout these discussions there was the continuous appreciation of the fact that availability has both subjective and objective dimensions, and that the individual questionnaire was best suited for the collection of data on perceptions of sources and services and the locations where current and former users had obtained contraceptive services.

On the other hand, the feeling was that a mapping of the actual presence of such facilities could best be achieved by a separate data collection procedure that would concentrate on cataloguing the types of services available, their actual proximity to the women in the area, and other characteristics related to transportation time, professional services, methods available, cost, and the days and hours open. Since the DHS is also focused on child health, the inventory of facilities was expanded to include details on the availability and characteristics of hospitals, clinics, health centers, pharmacies and private doctors. In addition to this inventory of family planning and health services, the supplementary community questionnaire also includes items on population site, types of access roads, distance to the nearest city, types of transport available, and the availability of public services such as schools, markets, sewer systems and the like.

The SAM facility selection strategy consisted in selecting the nearest facility of each type if located within 30 kilometers of the center of the DHS cluster. Only one clinic-based health service site of each type was selected in the SAM sample.

As the end of 1999 a total of 45 SAMs has been undertaken in 39 countries.

Situation Analysis

In the early 1990's, a somewhat different type of facility survey, the Situation Analysis, was introduced by the Population Council. The objectives of Situation Analysis studies are as follows:[9]

- to describe the potential of current policies and program standards to promote service delivery of quality services to clients,
- to describe and compare the current readiness of service delivery staff and facilities to provide quality services to clients against the current policies and program standards,
- to describe the actual quality of care received by clients, and
- to evaluate the impact the provision of quality services has on client satisfaction, contraceptive use dynamics, fulfillment of reproductive intentions and, ultimately, on fertility (in expanded research designs, most often using a panel of respondents).

The Situation Analysis approach differed from its predecessor facility surveys in two important respects. One area of departure was in terms of sampling. While both the WFS and DHS chose the facilities about which information was to be gathered on the basis of proximity to sample communities or clusters, facilities are chosen in Situation Analysis studies directly from lists of facilities. This sampling strategy is consistent with the primary measurement object of Situation Analysis - to describe the delivery of family planning and reproductive health services in a given setting. The strategy is, however, less than optimal for the purposes of describing the

---

[9]Robert Miller, Andrew Fisher, Kate Miller et al. 1997. *The Situation Analysis Approach to Assessing Family Planning and Reproductive Health Services*. New York: Population Council, Africa Operations Research and Technical Assistance Project.

services received by the average client or for assessing the impact of the family planning supply environment on service use and contraceptive behavior. It will be noted, however, that at least two attempts have been made to overcome these limitations by linking Situation Analysis data to DHS data.[10]

A second area of departure was in terms of content. In addition to providing information on the physical availability of facilities offering family planning and health services (as was done in both WFS and DHS) and the adequacy of facilities for providing services in terms of infrastructure, trained personnel, and equipment and supplies (as was done in DHS), Situation Analysis studies also included protocols for assessing some of the more qualitative aspects of service delivery. This was accomplished through the observation of transactions between service providers and clients, interviewing clients as they left facilities (that is, exit interviews), and interviewing clients of non-family planning/reproductive health services as to their perceptions and reasons for none-use of family planning/reproductive health services. As a result, Situation Analysis studies produced a more comprehensive picture of service delivery operations at sample facilities than did its predecessor facility surveys.

Other facility surveys in the health sector

A number of facility survey initiatives has been undertaken in the health sector that more or less parallel the Situation Analysis approach in the population/family planning sector. Illustrative of these efforts are surveys undertaken by the University Research Corporation (URC) in connection with Primary Health Care Operations Research and Quality Assurance Projects and the Centers for Disease Control (CDC) in connection with the Integrated Management of Childhood Illnesses (IMCI) project. Like the Situation Analysis, the efforts provide detailed, facility-based information on the quantity and quality/adequacy of services being provided. Due to space limitations, these efforts will not be described here.

Facility surveys under the EVALUATION and the MEASURE Evaluation Projects

The use of facility surveys to monitor trends and improvements in service delivery and to provide a basis for measuring the impact of services on reproductive behaviors has been strongly promoted by the EVALUATION and MEASURE *Evaluation* Projects. The facility survey work undertaken in connection with the EVALUATION Project is notable in two respects. First,

---

[10]Barbara Mensch, Mary Arends-Kuenning, Anrudh Jain and Maria Rosa Garate. 1997. “Avoiding Unwanted Pregnancies in Peru: Does the Quality of Family Planning Services Matter?”: *Studies in Family Planning* 23(1): 21-27; David Hotchkiss, Robert Magnani, Naomi Rutenberg, Luciano Correia, Gwen Morgan and Martha Sutula. 1995. “Access to Family Planning Services, Service Quality, and Contraceptive Use in Northeast Brazil.” Paper presented at the annual meeting of the Population Association of America, San Francisco.

although one of the rationales for collecting service availability data was to permit an assessment of relationships between levels of service availability and demographic and health behaviors and outcomes, such data had been (and continue to be) under-utilized for such purposes. One contribution of the EVALUATION Project was to demonstrate how certain multivariate analytic methods could be applied to DHS/SAM or Situation Analysis data linked with DHS household survey data to produce fairly robust estimates of family planning program impact. Impact assessments were carried out using linked facility and household survey data in Tanzania, India, Brazil, Morocco, Peru and the Philippines under the EVALUATION Project.

A second contribution was the refinement of the methodology for conducting facility surveys. The surveys conducted in Tanzania and India, for example, have advanced sampling methodology on linkage of sample areas for household surveys with those of facility surveys. Repetition of the health facility surveys in Tanzania have further demonstrated the informational and monitoring utility of a facility panel. Monitoring the same health facilities over time[11] has allowed program managers to observe how resources allocated for public health program improvements have, or have not, been implemented at the service delivery level.

Refinements in facility survey methodology have been continued under the MEASURE *Evaluation* Project. Work to date has focussed on two areas. First, the EVALUATION Project survey protocols have been further modified to include observations of service transactions and exit interviews with non-family planning clients and further detail has been added regarding the availability and quality of non-family planning services. Secondly, the sampling approach has been modified to include visits to all facilities located within DHS sample clusters and surrounding ones. The recommended sampling strategy for linked facility/household surveys is described in detail in this document, and a case study of this strategy is provided at the end of this manual. Thirdly, the content of the data collection protocols has been extended to cover more systematically maternal and child health and sexually transmitted infections (STI). The MEASURE *Evaluation* project has also defined a research agenda to investigate selected methodological and substantive issues concerning facility surveys as a method of data collection.

---

[11]Which can be refreshed with a new sample of health facilities at each point of time to preserve the cross-sectional representation feature

Chapter 2
FACILITY SURVEYS - SAMPLING ISSUES

This chapter discusses sampling issues as
- # facility surveys linked or not linked with population data
- # target populations
- # frames
- # sample size
- # stratification, stages of selection

Introduction

In this chapter we discuss basic issues on sampling that need to be considered for the design of facility surveys. We begin by examining the characteristics of health systems and the categories of clinics and health service providers often found in developing countries.

Health systems

In most countries a single agency has major responsibility for the organized provision of health services nationally. In the U.S. it is the Department of Health and Human Services. In other countries it is the Ministry or Department of Health. The allocation of health resources would have limited impact if they were not organized in some way, whether under a government agency, a religious body, an industrial or educational establishment, or in a less organized but directed fashion by powerful market forces, i.e., through the private health care market. An excellent review of national health systems is available in Roemer (1991)[12], wherein he notes that no two ministries of health have identical functions or organizational structures. That said, the central role of government health agencies is to protect the health of the population, a mandate complicated by a changing environment of such factors as health science capabilities, resource levels, disease epidemiology, and internal and external policy and program priorities. Government health ministries often perform key central functions of organizing the training of medical and health personnel; regulation and delivery of health services; procurement and distribution logistics of medical supplies and commodities; health statistics collection, reporting and evaluation; administration, budgeting and financing; and development of health policy and

---

[12] Roemer, Milton L. 1991. National Health Systems of the World. New York: Oxford University Press.

legislation.

Public health programs tend to emphasize the preventive side of health service delivery, focusing on the prevention of communicable disease and environmental hazards and health promotion through nutrition and education. Curative care involves medically specialized and surgical treatment, for which ministries of health are often involved in establishing standards of and regulations for care. Preventive health care often focuses on specific populations, such as infants and small children, women, mothers, or youth. Special programs such as eye or dental care or accidental injury prevention are included as part of preventive health services. The health services of interest to this sampling manual are those for family planning, maternal and child health, and sexually transmitted diseases, most of which are primarily preventive in nature.

Because government health ministries generally play a major role in the protection of population health in less developed countries, the organization and implementation of their reproductive and sexual health programs are important to understand in developing an appropriate sampling methodology. At the same time, it is important for a health facility sampling methodology to address the non-governmental side of health care, which may be provided through non-profit organizations, such as voluntary associations and churches, or for-profit organizations, such as private health maintenance organizations, private physicians, traditional healers, and pharmacies and other retailers. Whether responsive to market or mission forces, health service delivery is often distributed along lines that vary closely with the distribution of population.

Health services are directed towards the people whose health is to be protected; as such they are often organized by levels of primary, secondary and tertiary, where primary embodies most health promotion and disease prevention efforts. WHO has identified a minimum or essential set of primary health services to include:

1. Education concerning health problems and the methods of preventing and controlling them
2. Promotion of food supply and proper nutrition
3. Adequate supply of safe water and basic sanitation
4. Maternal and child health care, including family planning
5. Immunizations against the major infectious diseases
6. Prevention and control of local endemic diseases
7. Appropriate treatment of common diseases and injuries, and
8. Provision of essential drugs.

There is less consensus, formal and informal, on what should be classified as secondary health care. Generally services of the following type are categorized as secondary – specialized ambulatory medical service; common-place hospital care; care by non-medical specialists, and long-term care. Tertiary care is mainly medical and related services requiring highly specialized and skilled personnel and drug or equipment resources. Not surprisingly such services tend to be too expensive to maintain in but a few sites, often located in major urban hospitals. Referral of patients requiring increasingly specialized care from primary to tertiary levels is built into the system of health care delivery. In some places, emergency transport systems have been established to ensure that critically ill patients are moved from primary health care sites to higher level ones as needed, to obtain the appropriate treatment. Secondary and tertiary care facilities are often required for emergency obstetric care where surgical theaters are needed to perform cesarean sections or other assisted procedures of delivery for mothers with prolonged or obstructed labor or other acute complications (e.g., hemorrhage, toxemia, sepsis, or ectopic pregnancy). They are also the site of periodic care for AIDS patients experiencing acute complications or sick children requiring special medical treatment or surgery.

Monitoring the delivery of ongoing or special sexual and reproductive health programs requires that sampling methodologies for health facilities take into account the nature of the health system. As a result of their scope, primary health care tends to be available in service points closest to people's communities, such as community health centers, private physicians' clinics and traditional doctors or healers. Because types 4 and 6 in the above list of primary health services include the reproductive and sexual health services of interest, sampling methods that adequately cover facilities offering primary health services is important. At the same time, the uneven distribution and availability of quality health services in low-income countries draws many rural patients to public and private health facilities in urban centers and large towns. This necessitates including facilities at both the secondary and tertiary levels, to observe improvements in future distribution and utilization of health care.

Categories of clinics and health service providers

The categories of health service providers or facilities for the purposes of this manual are as follows:

- ❑ *health facilities or sites*
  - (Public sector)
    - Hospitals
    - Health centers
    - Dispensaries or health posts
  - (Private sector)
    - Clinics
    - Hospitals or sanatoriums
    - Physicians - consultation offices

- ❑ *Health care providers*[13]
    - Pharmacies
    - Community-based distributors
    - Health workers

[13] In some countries, delivery points may include traditional healers as well.

Staffing for the above is categorized as follows:

- Physicians, doctors
- Nurses
- Auxiliary nurses
- Technicians

Sampling issues

As discussed in chapter 1, facility surveys are conducted within the specific context of evaluation of population and health programs, including both program monitoring and impact assessment. The facility survey is one of several sources of data including program records, service statistics and household surveys that may be utilized to evaluate a family planning health program. In the manual, *Evaluating Family Planning Programs*, it is stated that "the prime objective of facility surveys is to describe the availability, functioning and quality of health and family planning activities." The facility survey thus provides program-based output data used to monitor program performance. The information collected may also be used as one of the data sources to help assess impact. The Manual also points out that program performance variables used to assess impact are the same as the indicators used for program monitoring, and that what differentiates the two are the evaluation design and analytic techniques used.

When considering the sampling implications for the design of a facility survey, it is useful to differentiate, also, among the ways in which the facility indicators are collected and used, as follows:

(a) Linked data
The facility survey data are used in tandem with population data to evaluate both intermediate outcomes such as contraceptive prevalence rate or long-term outcomes such as total fertility rate. Expected changes at the population level may be analyzed in two ways - (a) without attempting to establish whether there is a causal relationship between programs or interventions and an observed change in these outcomes among the study population, or (b) by attempting to establish causality through multi-variate statistical methods.

(b) Linked data *and* linked sample areas
Same as above, except the facility survey data and population data come from the same sample areas (clusters), insofar as possible.

(c) Non-linked survey
The facility survey data are used in isolation to analyze availability, functioning and quality of service provisions - mainly for monitoring purposes. In the analysis the survey results are not linked to population data, for whatever reason, including the fact that relevant population data may not be collected within the same time

frame as the facility survey.

For sample design of a facility survey, the essential differentiation is between the second compared to the other two, or, whether the survey data are linked with population survey sample *areas*, not whether the facility and population *data* are linked in the analysis. If both the facility survey and a relevant population survey are conducted in the same general time frame and if both surveys are based on valid, probability samples, the data can be linked for analysis - irrespective of whether they both select the same sample sites but that will require aggregation of data at some geographical or administrative level. What is important to note here is that this pertains whether the purpose of the analysis is for monitoring or impact assessment.

For program impact assessment the linkage of the facility and population surveys is essential because in that way the facility survey data provide information on the health service supply environment to which the individuals and communities included in the population survey are exposed. The standard method for estimating program impact with non-experimental survey data is multi-level or contextual analysis. The analysis of program impact using these techniques is greatly enhanced using data from linked facility and population survey because it allows to preserve at the individual level the outcome of interest and to consider as explanatory factors individual, household, community and program characteristics relevant to the individuals included in the sample of study. Multi-level analysis can certainly be used with independent non-linked facility and population survey data but to do so it will be necessary to aggregate the individual information at some geographical or administrative level to match the health outcomes of interest to the health service or program information. The unit of analysis will not longer be the individual. This area-level type of analysis is statistically feasible and has been used by program managers and researchers interested in program impact. This procedure, however, imposes several limitations in the analysis. The main limitation of this analysis is the valuable individual-level information that is lost by aggregating data. In fact the number of observations is reduced tremendously by aggregating data. It also begs a theoretical issue that has been called the “ecological fallacy” by sociologist Robert Merton, which argues that causal inference on aggregate units improperly assumes that the dynamics of social behavior can be explained by the actions of these aggregate units.

On the other hand, there are important logistical and cost-saving considerations in having the facility and population surveys conducted together and within the same areas, especially for a program such as DHS. Under that condition the data from the two surveys are “linkable” but so are the sample areas, so that the correlation between service providers and population outcomes may be improved somewhat.

For brevity in discussion we will refer to a facility survey that is designed without regard to the particular sample areas that are used (or will be used) in a related household survey as a *stand-alone* facility survey. “Stand-alone” in this context does not imply that the data cannot be linked with that from other sources for analysis, but rather that the facility survey sample is selected independently of any household survey. For the other type, we use the ponderous term, facility survey linked to population survey sample areas, or shortened simply to *linked survey*.

Target populations and units of analysis

A facility survey has multiple target populations - facilities, staff, clients, service environments. In addition, an important unit of analysis, is the client-staff interaction, where the target population could be either the staff or the client, depending upon the analytical perspective.

From a sampling standpoint, the prime target population is the facility. The facility must be sampled first. In that respect, important sampling parameters such as sample size and stratification criteria for client or staff samples are determined to a great extent by the facility sample design. It is not possible to have, for example, a scheme for urban-rural stratification for the client sample that is different from the one that will already have been imposed by the facility sample plan.

It is important to define the in-scope staff eligible for inclusion before the survey is fielded. For example, if only health workers that are engaged in the provision of reproductive and health services are in scope for the survey, this should be clearly understood in advance. Sampling of staff would, in general, be recommended only for large facilities. Staff interviews would be carried out on a census basis, that is, 100 percent coverage, in any sampled facility containing only a few staff. Similarly, clients in facilities that serve only a small number would be surveyed on a census basis, but in the context of a particular time frame, such as one full day, through the technique of an exit poll or interview.

As mentioned in chapter 1, weighting of data for clients and staff is a pre-condition for producing unbiased estimates for those populations, because they are sub-sampled. This may be seen by considering the unworkable alternative where *every* staff member and client associated with the sampled facilities would have to be surveyed - that is, to avoid weighting.

Client-staff interaction data should be collected for the client sample as opposed to the staff sample, since unbiased estimates for clients can only be made under that circumstance. Moreover, the client-staff interaction data and the client interviews should come from the *same* sample of clients.

Sampling frame development

Frame development differs for a stand-alone facility survey compared to one linked to a set of population survey sampled areas. For the former, it must rely upon a list of some kind to serve as the sampling frame. A list of facilities by type, together with figures on patient load and staffing, would serve this purpose. A perfect list - an unattainable ideal - must be complete, accurate and current. Frame lists that have been tried in facility surveys have suffered from various flaws. A list of public facilities may be available from authorities such as the Ministry of Health, but these lists are often incomplete or outdated. Lists of private facilities are often nonexistent altogether in many countries.

Size measures for clients and staff, which are needed for the development of efficient sampling plans, are sometimes not available or often of poor quality.

A suitable list of facilities may be constructed, however, but it requires field work in advance of the survey. There are various methods of doing this:

(a) start with an available list and up-date it through a dependent listing operation,
(b) conduct an independent listing operation from scratch,
(c) compile a list (dependently or independently) but restrict it to a sector such as large facilities and use it in combination with a geographic area sample.

For each of these methods, a facet of the task would include obtaining the size measures (client load, staff size). Of the three, (b) may not be feasible in most countries because of the expense of mounting such a task from scratch. For (c) the area sample would be used to cover facilities not on the list (see a complete discussion of area sample usage in chapter 4).

For a facility survey conducted in relation to a population-based, household survey, the requisite facility sampling frame would be developed from the household survey clusters or PSUs (primary sampling units) - a variation of area sampling in (c). It would be done by canvassing the sample clusters to identify in-scope facilities, including those which may be nearby. The technique would assure unbiased coverage of the universe of facilities, but it would likely need supplementation with a list of very large facilities to reduce sampling variance.

Sample size - facilities, staff, clients, client-staff interactions

As with sampling frame development, sample size considerations are different between stand-alone surveys and those linked with population sample areas. With the stand-alone survey it is easier to work out the sample size for facilities in relation to pre-specified precision requirements. In theory, the sample can be made as large as the mathematical formulas dictate, although, in practice, survey budget limits will often take precedence in the actual sample size to be used. By contrast, the number of sample facilities in a survey linked with a set of population sample areas is controlled somewhat by the number of sample PSUs in the population survey.

Sample size is largely determined by whether estimates of level or change are wanted, with the latter requiring a considerably larger sample in general. Examples of estimates of level are the number of facilities in a region or the percentage that provide a given family planning service. A change estimate looks at, for example, differences in facility performance measures at two (or more) points in time. These aspects of sample size are treated separately in two chapters - chapter 4 for level and chapter 6 for change.

Theoretically, the sample size needed to obtain reliable estimates for clients or staff (or client-staff interactions) could be calculated independently of that needed for facilities. In most practical applications, clients and staff would be sampled and interviewed in every facility selected. Realistically, therefore, it is more rational to relate the sample sizes for clients and staff to the facility sample sizes in some ratio.

⇨ *Example*: if the sample size for the facility sample were determined to be 600, the ratio method might lead to sampling and interviewing, on average, *two* staff members per facility, or 1200 total staff, and *four* clients, or 2400. Considerations of budget and survey logistics for client and staff sampling would be the overriding ones, rather than precision requirements. For example, all staff in very small facilities (say, under 4) would be interviewed, while a sub-sample of those in larger facilities would be selected; similarly, for facilities that service a small number of clients versus the larger ones.

When client and staff sample sizes (ratios) are determined in relation to the facility sample size it is irrelevant whether the facility survey is stand-alone or population-linked. Thus, the ratio method would be the recommended approach.

Sample design - rare occurrences

A special issue for sampling and sample size, in particular, is client-related rare occurrences. These would include, for example, new family planning clients, women with problematic pregnancies, patients with STIs and sick children with specific childhood diseases. In many countries, even routine family planning clients are rare. In a general-purpose facility survey, it will not be possible to specify a number of such cases to survey because the client sample will depend upon those who show up at the facility during a specified time period. As a result the number of sample observations is likely to be quite small, and hence the reliability of estimates pertaining to these important sub-groups will be poor.

Study of these rare populations is best reserved for special investigations, rather than trying to cover them adequately in the general-purpose facility sample survey. There may be specific facilities that are designated to see such patients (such as STIs) or specific times of the year when the likelihood of observing certain illnesses is greater. One approach would be to try to ascertain such information in the general-purpose survey and design a special-purpose study for rare occurrences to be done later. Part of the strategy for the special-purpose study would be to have interviewers present in the facility for 5 or more days in order to ensure enough observations per facility.

Sampling efficiencies

Normally, the sampling statistician has a choice of sample design. He/she can introduce various methods to achieve sampling efficiency beyond what would be possible with a simple random sample. Considerations of stratification, use of clusters, optimum allocation, stages of selection, sample size build-up for domain estimates are regularly built into sample plans to improve estimates and/or reduce costs. It is noteworthy that each of these may be considered for a stand-alone facility survey, but limits are imposed on their introduction when the survey sample is defined by the same sample areas (and surrounding areas) used in a household survey.

In a stand-alone facility survey, the use of stratification would be called for to ensure fair representation in the sample for important sub-groups that may differ in significant ways. Size of facilities as well as types - large versus small, public versus private, those offering family planning versus those that do not - are examples of important stratification groupings. The nature and extent of service providers are likely to be differentiated in most countries by facility size and their public-private status. Quality of care may also be affected by these variables. Another important stratification variable is urban-rural, because access to services is likely to be quite different (better) in urban settings where public transportation is more readily available. Very large hospitals are also more likely to be found in urbanized areas. Optimum allocation of the sample to strata can be employed as a design technique.

Related to stratification is the issue of estimation domains - those groups for which separate estimates are wanted in the analysis. Sub-national data may be wanted for various geographic areas or groups of areas, as well as for sub-groups which may have been set up as strata (for example, a stratum of large hospitals, or strata differentiating by type of facility). A sample plan can be designed to accommodate those needs, if is developed independently of the household survey sample areas.

Cluster sampling may also be employed to reduce costs. In the context of a facility survey, this is best applied through the use of an area sample where all facilities within selected areas are inventoried and then surveyed on a census or sample basis. A variation of this approach may be used in either a stand-alone survey or one that is population-linked. The area-based cluster sample is much more difficult to control in terms of the size of sample necessary for sub-groups. For that reason, it would not be recommended (or even feasible) for large facilities; instead, they would be sampled from an existing list or one created specially for the survey.

In general, a one-stage sample of facilities would be the recommended procedure in most applications. In very large countries, however, a first-stage selection of areas might be utilized to reduce the number of locations to be visited, in which case the facilities selected would constitute the second-stage of sampling. For clients and staff, they would, by definition, be selected at a subsequent stage (second, if the facility is the first stage), because they would be sub-sampled from the particular facilities in sample. This of course has implications on survey estimation, as noted previously, since the client and staff samples would necessitate weighting.

Sampling of facilities from a list frame should be done using systematic sampling or systematic in combination with *pps* (probability proportionate to size), although random sampling is a possibility as well. When the list is sorted in geographic sequence, systematic selection will provide both implicit stratification and proportionate allocation, if the sampling rate is constant.

For a population-linked facility survey, considerations of stratification, optimum allocation, domain estimation, stages of selection are all precluded in development of the sample plan. The facilities that are eligible for the sample are determined by the population-based survey clusters or PSUs, and hence by the sample design that gave rise to those PSUs. Whatever strata (geographic or otherwise), allocation schemes, domains or selection stages were developed for the population survey are carried over to the facility survey. It should be indicated that the estimates produced by the linked survey, as recommended in chapter 5, are still unbiased estimates of facility characteristics. They are, however, less efficient that those obtained from an stand-alone survey.

Survey constraints

We conclude this chapter with brief mention that various survey constraints may impose further burdens on optimum sample design for facility surveys. A crucial constraint is that the cost dimension of the facility survey may compromise the sample size or vital developmental activities such as creation of a list frame. It is to be expected that, even when the facility survey is conducted in conjunction with a survey such as DHS, the latter is going to have financial resources larger than the facility survey by, perhaps, orders of magnitude.

When a listing operation is required to obtain the facility frame, it will have to be planned and conducted considerably in advance of the facility survey. Further complications in this connection arise if the survey is to be coordinated with the conduct and timing of the population-based survey.

Interviewer workloads are a crucial survey parameter but difficult to plan accurately in advance of the fielding of the facility survey, especially for client interviews and observations. Certain procedures can be implemented in compilation of facility data - for those facilities which fall into sample - to enable staff and client volumes to be estimated and thus provide a planning tool for estimating interviewer workloads (see chapter 4).

Chapter 3
GENERAL PRINCIPLES AND CONSIDERATIONS

This chapter discusses
- # probability methods
- # national coverage
- # simplicity of design
- # self-weighting designs
- # techniques to avoid

Introduction

In this brief chapter we discuss the basic parameters of good sampling practice that ought to be used in any facility survey used for monitoring and/or evaluation purposes, and whether linked with population survey sample sites or not. Consequences of not using accepted methods are indicated. Recommendations, as such, are not made in this chapter. Instead, many of the features discussed are presented as requirements for facility survey sampling.

Probability, or formal, methods

Sampling is used in lieu of gathering information about the entire population of facilities to economize resources. Instead, one collects the facility information from a part of the facility population. The population is referred to as the *universe* and the part surveyed as the *sample*.

Scientific sampling requires that the sample be selected in such a way that valid inferences about the universe can be made from the sample estimates. There are myriad ways of doing this, and the sampling statistician has many choices in the particular sample design chosen. For the design to be a probability one, however, it must satisfy the mathematical theory that every member of the universe of facilities has a known, calculable and non-zero chance of being selected into the sample.

When the sample design meets the probability criterion, sampling error can be evaluated. *Sampling error* is the error that occurs because the survey estimate (for whatever variable) comes from a sample instead of the whole universe. It can be estimated directly from the sample data after the survey. In this way, confidence intervals around the survey estimates can be constructed, to enable the analyst to evaluate the range of sampling error associated with them.

Formal sampling methods require the use of probability techniques. They are characterized, not only by the possibility to estimate sampling error, but also by the following features:[14]

(1) Clearly defined selection procedures
(2) Use of lists (including lists of areas) as sampling frames
(3) Applicability of sampling theory.

REQUIREMENT 3.1

⇨ Use probability sampling at every stage of selection.

This manual presumes that formal, probability sampling methodology is to be used to select the facility sample and at every stage of selection. A number of different sampling techniques are employed, including systematic sampling, list sampling, area sampling, stratified sampling, cluster sampling, sampling with probability proportionate to size and multi-stage sampling. See the glossary for definitions of these terms.

National and sub-national coverage

The facility surveys that are discussed in this manual are intended, mainly, to provide estimates of facilities and their characteristics (a) at the national level and (b) for important sub-national areas such as regions, large provinces, urban-rural. The techniques described for sample design - chapters 4 through 6 - may also be applied to a sub-region of a country (rather than the nation as a whole), especially when action or intervention programs are confined to the sub-region and the government wishes to monitor or evaluate it. For either purpose - national estimates or program monitoring of a particular area - *coverage* must be as complete as possible.

Complete coverage requires ensuring that all geographic areas of the study universe are given a chance of inclusion in the survey. The situation is often encountered, after a sample has been selected, that certain sampled areas are deemed inaccessible, out-of-scope or unsuitable for the survey for one reason or another. The survey sponsors may wish to exclude areas that constitute a security risk to the survey team, areas too difficult to reach because of the terrain, areas that are inhabited by refugees or ethnic groups not considered part of the target population, and so forth. For proper sampling it is important to identify any such "out-of-scope" areas before the sample is selected and then select the sample from the redefined universe, rather than eliminate out-of-scope areas after sampling or substitute other areas for them. The former will yield an unbiased estimate for the smaller universe, but the latter will yield a biased estimate for the full universe.

---

[14]Scott, C. "Sampling for Monitoring and Evaluation," A Technical Supplement to *Monitoring and Evaluation of Agriculture and Rural Development Projects,* Washington: The World Bank, 1985, p.5.

> **REQUIREMENT 3.2**
>
> ⇨ Design a survey for national coverage.

In sampling from a list frame of facilities, it is also useful to identify, in advance of sample selection, facilities which may be out of scope for the survey, though the coverage issue is less of a problem. If, say, acupuncture facilities or faith healing centers are not to be included in the facility survey it is better to pare them from the list before selection, so that the sample size of in-scope facilities can be controlled precisely. On the other hand, there is no coverage bias if they remain on the list and are screened out, when selected, at the interview stage. Sample size can be controlled and hence reliability, though not precisely, by estimating the proportion of out-of-scope facilities on the list and over-sampling accordingly, the expectation being that the same proportion would be encountered in the field during the screening interview.

Simplicity of design

The main audience for this manual is practitioners in developing countries who are planning facility surveys, either stand-alone or in conjunction with a household survey sample. In many instances the practitioners may not be sampling statisticians, and they may not be able to access the services of sampling experts to assist in the design.

The manual is not intended to show, definitively, what the sample design should be for a particular country application, since each country will have its own requirements, conditions and resources. It is also not intended to teach someone how to become a sample designer for facility surveys. It is intended, however, to provide useful guidelines that may be used by the survey team to plan a sampling strategy. The various tasks required, as seen more completely in the remainder of this manual, are nevertheless very demanding with respect to the sampling skills needed. For that reason it is strongly advised that sampling expertise be sought to assist in the sample planning and implementation if such expertise is not present among the project team.

> **REQUIREMENT 3.3**
>
> ⇨ Use simple, as opposed to complex, sample methods.

The methodology of sampling for a stand-alone survey is fairly straightforward on a theoretical level, although there are significant issues regarding frame development and other implementation problems. Complications arise when the facility survey is conducted in the sample PSUs, plus adjacent ones, of a household survey. Nevertheless, an overriding consideration in the development of the manual was to select methods and procedures which are do-able and cost-efficient. Simplicity of design is thus a guiding principle not only for this manual but for any facility survey sample plan that a country might develop on its own. More elegant or complicated designs increase, substantially, the likelihood that field implementation will be error-prone, thus increasing *non-sampling* error.

Use of self-weighting designs

Generally, it is recommended that samples be designed to permit survey estimates that can be made without the necessity for weighting. This would allow percentage distributions, for example, to be made directly from the sample counts. Weighting of the data is not needed when all the ultimate sample units are chosen with the same probability of selection, except in cases where differential adjustments such as may be necessary for different non-response categories may be indicated.

Unfortunately, a feasible facility survey cannot likely be designed on a self-weighting basis - either as a stand-alone survey or population-linked. The stand-alone survey will necessitate using a combination of facility list(s) and area canvassing in such a way that the probabilities of selection between the two are apt to be quite different. In a population-linked survey, a self-weighting design is precluded by definition since the probabilities of selection are determined by the sampling clusters selected in the population survey, which are virtually always selected in proportion to their sizes.

Weighting itself is not problematic either statistically or practically, however, given the capability of today's computers. It is only necessary that the weights be calculated properly, for insertion into the data files to tabulate the survey estimates. Some available software requires that "standardized" weights be used, whereby the survey weights are re-scaled so that their sum equals the original sample size[15].

Informal methods - techniques to avoid

Informal sampling methods are the complement to the formal, probability methods discussed above. They go by various names including purposive samples, convenience samples, judgment samples, quota samples, case studies or expert samples.

REQUIREMENT 3.4

⇨ Avoid purposive, convenience, judgmental or quota samples.

⇨ *Examples:* Such a sample in the context of facility surveying would be one in which facilities are chosen on the basis of their convenience for interviewing, as opposed to the use of a random selection technique. Another example would be a selection of, say, the first 5 clients that show up at a facility - a quota methodology.

---

[15]A procedure for calculating standardized weights is illustrated in *Sampling Manual*, DHS-III Basic Documentation 6, Macro International Inc., Calverton, Maryland 1996 and *FANta Sampling Guide* , Food and Nutrition Technical Assistance, Washington, 1999.

These informal methods suffer from many flaws that preclude valid inferences about the whole population of facilities - or their clients and staff - to be drawn from the “sample” results. The flaws include the fact that they do not adhere to probability selection criteria; and, as such, selection probabilities cannot be ascertained. Sampling frames are often not used in a systematic way; frame exclusions for various cost or administrative reasons will likely lead to a difference between the target and the survey populations[16]. Sampling errors cannot be estimated from the sample data themselves. Finally, valid confidence intervals showing the margin of error around the survey estimates cannot be constructed.

For these reasons, informal sampling methods are not recommended for the kinds of facility surveys covered in this manual.

[16]Even when a frame is inaccurate or incomplete, probability methods ought to be used on the available frame and users of the resulting data informed about known frame exclusions.

Chapter 4
SAMPLING FOR STAND-ALONE FACILITY SURVEYS

This chapter discusses
- # preferred dual-frame design
- # list and area frame sampling
- # sample size, stratification and selection methods
- # staff and client sampling
- # alternative approaches

Introduction

In this chapter suggested guidelines are given in detail on sampling approaches for facility surveys or assessments when the design is independent of any population survey to which its data may (or may not) be linked at the analysis stage. That is, the facility survey sample design is not determined by the sample PSUs or clusters of the relevant population survey. Sampling for facility surveys that are confined to the population survey sample areas, and surrounding areas, is taken up in chapter 5.

A primary measurement goal of the sample design for the so-called stand-alone facility survey is to be able to make an unbiased estimate of the *number* of facilities, as well as their characteristics and services provided. Two of the most important subsidiary goals include making estimates of the various types of facilities and providing the frame for sampling the staff and/or clients of the selected facilities. Sample implementation, however, may be fraught with practical problems, and these are discussed, along with suggested solutions, throughout this chapter in connection with the recommended sampling design approaches.

Preferred sample design - essential features

In this section we will recommend the essential features of a sample design for what would likely be the usual situation in most of the countries that may conduct a stand-alone facility survey for monitoring and/or evaluation - either (1) national in scale or (2) program monitoring or evaluation of a study universe that encompasses a substantial part of the country (for example, a major region, one or more provinces or districts). A subsequent section will offer sampling approaches for the fairly specialized situation when monitoring or evaluation is for a target area of an action program or project that is confined to a comparatively small area of a country such as a compact set of villages or a single city.

For the "usual situation" case, the essential feature of the sample design to meet the goals mentioned in the first paragraph of this section is the use of a *dual-frame sampling methodology*. A list frame should be used for the selection of large facilities, and an area frame should be used for all others.

Design details are discussed in subsequent sections of this chapter, but the preferred sample plan can be summarized in the following steps:

RECOMMENDATION 4.1

⇨ Use dual frame sample for stand-alone survey.

List frame sample

(1) Compile a list of the large (or otherwise significant) facilities for the study universe.
(2) Stratify the list in an appropriate way.
(3) Select a sample, systematically (or random), from the list and conduct the survey interview in the selected facilities.
(4) In the selected facilities, further select a sample of clients and a sample of staff, systematically or at random, for interview.

Area frame sample

(5) Compile a list of geographically-defined areas that cover the entire study universe (example, census enumeration areas).
(6) Stratify the list of areas in an appropriate way.
(7) Select a systematic sample of these areas and canvass them to identify facilities and service delivery points - SDPs, but excluding any which appear on the list frame.
(8) Conduct the survey interview in all facilities identified except those which appear on the list frame.
(9) In the sample facilities, further select a sample of clients and a sample of staff.

It is important to recognize that the design is made up of two samples - one from a list frame of large or other important facilities and one from an area frame for the balance. In terms of the sequence of operations, it is not necessary to select (or interview) the list frame sample prior to the area sample, but it is necessary to create the list frame first, that is, before the area frame sample is implemented. This is because any facility which appears on the list frame is not to be included in the area frame interviews - step (h). Note that this procedure applies not merely to *sampled* facilities on the list frame but rather the entire list, since any facility on the list frame will have had its proper chance of inclusion in the sample. Operationally, it is necessary therefore to supply the interviewing teams working on the area frame with the names, addresses and locations of the list frame facilities in their areas, so that these may be excluded during the area frame canvass - step (g). Un-duplication of the two frames is further discussed in a subsequent subsection of this chapter.

It is useful to note also that the field task of canvassing to locate facilities applies only to the area frame. The large facilities that make up the list frame would be identified in other ways exclusive of field canvassing. See further subsections in this chapter that provide more detailed procedures on the development of the list frame and canvassing for the area frame.

Advantages and disadvantages

As mentioned, implementing the steps above can pose a number of practical problems to be discussed subsequently, but the conceptual design is straightforward and has much to recommend it. The principal advantage is the capability of the design to provide an unbiased estimate of total facilities, and by type, without having to start with a complete and accurate facility list at the outset. (The list of *areas* for the area frame must, however, be complete.) Even when the list frame is faulty, the area sample will theoretically cover the entire universe anyway, which is the reason that un-duplication is necessary - step (h). In other words, while the list frame should be as complete and accurate as practical in order to assure that survey variances are acceptably low, even when it is incomplete or contains inaccurate information the area frame will compensate by providing unbiased coverage of the universe.

A disadvantage of using a list frame is that the facilities selected from it will be scattered, geographically, throughout the study universe. This may require travel of consequential distance to each location containing a list sample unit. For example, if 60 hospitals are chosen, it might require significant travel expense to visit each of them, since they are likely to be widely dispersed. By contrast, the facilities selected from the area sample frame will be "clustered," so that the travel dimension will be between sample areas rather than sample facilities.

The remainder of this chapter discusses sample size and details of the nine-step preferred sample plan, along with recommended guidelines on stratification schemes and implementation methods. It is necessary to begin with the issue of sample size, since the other sampling issues depend so much on it.

Sample size for facilities

Determining sample size is a complex subject for any survey. For a facility survey there is added complexity because the survey has three target populations - facilities, staff and clients. Each of these requires its own sample and, thus, its own sample size considerations. The complexity of sample size determination is due to the often elusive issue of *precision* requirement, but survey budget and constraints, plus the likelihood of having to produce reliable estimates for domains compound the issue.

When the precision, or *margin of error*, needed for the estimates can be stated in advance it is fairly easy to determine the necessary sample size, using well-known mathematical formulas, assuming that some reasonable assumptions about the unknown parameters can be made. There are always scores if not hundreds of estimates, however, that will be produced by the survey, and so it is necessary to specify the most important ones in order to calculate the sample size. Even so, each of the main estimates will, no doubt, require a *different* sample size. A procedure that has been used successfully is to calculate the sample size requirements for all of the important estimates and choose the largest.[17]

Determining the sample size strictly in terms of precision requirements (that is, without regard to available budget and other resources) should be straightforward for estimating facilities. This is so because the facility estimates are all of the form, "percent with attribute X." It is sometimes difficult to get the sponsors of a survey to focus their attention on the precision requirements, but, suppose as an example, the required precision is specified at plus or minus 5 percentage points and, further, it is anticipated that *each* of the key estimates of the survey will be in the range of about 40 to 60 percent. In that case, the largest sample size that would be needed is when the percent with the given attribute is exactly 50, and that is the sample size that should be used. Then, the survey result, based on the sample size so calculated, might be, for example, 55 percent of facilities reporting the presence of appropriate examination areas for clients. In this example, the survey estimate, taking account of its margin of error (*standard*, or sampling error), would be 55 percent plus or minus 5 percent - that is, a confidence interval of 50-60 percent (at the 95 percent level of confidence).

However, one or more of the key estimates of interest may constitute a small percentage of the total, in which case a fixed percentage as large as 5 points for the margin of error to be tolerated would not make sense. For example, an important survey estimate could be percent of facilities reporting client record card systems. If that percentage were anticipated to be around 5, allowing a margin of error of 5 percentage points would result in a survey estimate with a confidence interval of 0+ to 10 percent, a result not likely to be very informative to users.

To avoid the above problem, it is useful when calculating the sample size to specify the precision requirements in terms of the width of the confidence interval around the estimate of concern. This further implies looking at *relative* errors instead of standard errors. For example, a country may want to estimate each of its important items at the 95-percent level of confidence with, say, a relative error of 10 percent; for a 20-percent item this would translate into a standard error of 2 percentage points, while for a 40-percent item it would be 4 percentage points, and so forth. A relative error, also known as *coefficient of variation* (*cv*) of 10 to 20 percent is, in fact, commonly specified as the precision needed for the key estimates of a survey, no matter what their magnitude. Statistically, the coefficient of variation is equal to the standard error of the survey estimate divided by the estimate. A related measure, the relative variance, or rel-variance, is equal to the variance of an estimate divided by the square of the estimate.

---

[17]See, for example, *Multiple Indicator Cluster Surveys Handbook*, UNICEF, New York, 1995, and revised version 2000; and *FANta Sampling Guide*, previously footnoted.

We use the measure of rel-variance to determine sample size as follows:

RECOMMENDATION 4.2

⇨ Set width of confidence interval for key estimates for facilities at p ±.1p, p ±.15p, or p ±.2p.

$$n = \frac{3.84 f}{V^2} \frac{q}{p}, \text{ where}$$

n is the sample size we wish to calculate,
p is the anticipated proportion of facilities with the attribute of interest,
q is equal to $1 - p$,
f is the so-called design effect (shortened from *deff*),
V^2 is the relative variance, (square of the relative error), and
3.84 is the square of the normal deviate (1.96) needed to provide an estimate at the 95 percent level of confidence.

Note that formula used here is one of several known formulas that produce the same result. The design effect, f, is a value that reflects the ratio of sampling variances, where the numerator is the variance of the sample design being used for the particular facility survey in question, and the denominator is the variance that would result if a simple random sample of facilities with the identical sample size had been used. The design effect reflects the effects of stratification, stages of selection and degree of clustering used in the facility survey. Generally, the clustering component, which is a measure of the degree to which two facilities in the same cluster have the same characteristic compared to two selected at random in the population of facilities, contributes the biggest effect. The interpretation of the design effect is that it shows how much more unreliable the sample is compared to a simple random sample of the same size. If the design effect were 1.2, for example, the facility sample would have sampling variance 20 percent greater than an alternative design using simple random sampling.

Unfortunately, unless information is available from a prior survey, design effects are not known until after the survey has been conducted and the sampling errors have been calculated. Moreover, each estimate in a survey has a different design effect. The sampler is thus faced with a dilemma for the calculation of sample size. It is thought, however, the design effects for most of the facility survey estimates of interest will be very low because (1) the list sample will not be clustered at all and (2) both the cluster sizes (that is, number of sample facilities) and the intra-cluster correlations in the area sample will be small (see more about the list and area samples below). We will assume, therefore, for purposes of calculating sample sizes that the value of f will be about 1.2 at the maximum.

RECOMMENDATION 4.3

⇨ In calculating n for facilities assume sample design effect is low, i.e., 1.2.

It should be noted that when there are

comparatively few facilities in the universe, so that the sample size becomes a significant proportion (say, 5 percent or more) of that total, the calculated sample size should be reduced by the factor, 1-n/N, where N is the number of facilities in the universe. This is an important variation of the formula given above for small study universes. The formula, as given, assumes n is very small relative to N, and the correction factor has therefore been omitted.

⇨ *Examples*: Suppose we set the width of the confidence interval at p ± .15p at the 95-percent level of confidence. Then relative error or coefficient of variation is 15 percent, or 0.15. The relative variance is the square of the relative error. So, V^2 is $(.15)^2$ or .0225. Thus if the estimate, p, that we want to obtain is thought to be 50 percent, or .5, the sample size would be calculated as

n = [3.84 (1.2) (.5)] / (.0225) (.5), or 205.

Similarly if p is thought to be 15 percent, or .15, the sample size would be

n = [3.84 (1.2) (.85)] / (.0225) (.15), or 1161.

To avoid using the formula, sample sizes necessary for various values of p are given in Table 4.1. The values in the table give results for confidence intervals of p ± .1p, p ± .15p and p ± .2p, but assumes the design effect is fixed at 1.2.

Table 4.1. Sample size, n, for varying estimated percentages, p, (facility estimates) - with confidence interval p ± .1p, p ± .15p, and p ± .2p

Value of item, p	confidence interval	confidence interval	confidence interval
	p ± .1p	**p ± .15p**	**p ± .2p**
0.8	115	51	29
0.7	197	88	49
0.6	307	137	77
0.5	461	205	115
0.4	691	307	173
0.3	1075	478	269
0.25	1382	614	346
0.2	1843	819	461
0.15	2611	1161	653

A country that wishes to use a confidence interval other than the ones shown in the table would have to calculate sample sizes using the formula. If a country has information from a previous survey that suggests the value of the design effect is different from 1.2, that should also be used to calculate the sample sizes, rather than relying on the table values. The default value, 1.2, for the design effect used in the table may be too liberal for some types of designs; for example, it should be approximately 1.0 for stratified samples drawn completely from list frames without clustering. Also, when the number of facilities in the whole population of facilities is relatively small, the additional factor, 1-n/N, must be inserted into the formula to obtain the correct sample size.

The table reveals quite clearly how sensitive the sample size is to the confidence interval, as well as the proportion parameter, p. It would be expected that among the key estimates for most countries some of them would be in the lower ranges for p - 10 to 25 percent. Budgetary considerations would also have to play a role in deciding upon the precision of the results and the associated confidence interval to specify.

Adjusting sample size for non-response

It should be noted that the sample sizes calculated are estimated and based on assumptions which may not hold up strictly when the survey is conducted (the design effect, for example). For that reason, the values shown should be considered as orders of magnitude.

⇨ *Example*: The tabled value of 614 for a 25 percent characteristic with confidence interval 25 percent ± 3.75 percent should not be taken as a literal target sample size. It could be rounded up to 615, 625 or down to 600 without noticeable change in precision.

Moreover, an important assumption for the tabled values is that response will be 100 percent. Since complete response is rarely attainable in the field, the calculated sample size should be increased by a factor to reflect the anticipated non-response rate.

⇨ *Example*: if non-response is expected to be about 10 percent, the sample size would be increased by 1.1. In the preceding example, the calculated figure of 614 would thus be increased to 675.

Adjusting sample size for domains

The survey design will most likely require that the estimates be disaggregated for important estimation domains - urban-rural, major regions, facility types. If there is

Box 4.1 Sample Size Calculations - Summary

- ✓ Ascertain main estimates of interest
- ✓ Identify those with small value of p
- ✓ Look up n in Table 4.1, or use formula
- ✓ Choose largest n; round it
- ✓ Adjust n upward for non-response
- ✓ Evaluate n in relation to budget, field constraints; revise if necessary

particular interest in obtaining very reliable data for a given domain, it may be necessary to increase the sample size in that domain.

⇨ *Example*: If *equally* reliable data were wanted for urban and rural areas separately it will be necessary to sample the two areas disproportionately to assure the *same* sample size in each (except in the unlikely case where they each contain a 50 percent share of total facilities). By way of illustration, use of a proportionate sampling scheme when the urban-rural distribution is 65 and 35 percent respectively would give a sample size for the urban part that is about twice as big as the rural part, in which case the reliability[18] of the urban sample would be twice as good. The desire for equal reliability in this case would demand that the rural sample size be increased commensurately.

In general, the sample size for domains when equal reliability is wanted for each necessitates multiplying the calculated sample size (that assumes one domain) by the number of domains. Thus, if equally reliable estimates were wanted for, say, 5 regions, the sample size would be about 5 times the values shown in Table 4.1. The survey budget would likely preclude such a large sample, so certain compromises would have to be made. One such compromise is to relax the confidence interval criterion for the domain estimates. For example, if the national level confidence interval is set at $p \pm .15p$, then the confidence interval for domains might be set at $p \pm .2p$. Another compromise is to select the most important domains for the stricter reliability and allow the others to be measured with whatever reliability a proportionately allocated sample would yield.

An alternative approach for determining domain and overall sample sizes is to carry out the calculations from the formula above separately for each domain of interest. The total sample size would then be the sum of the domain samples.

List frame

Development

It is first necessary to ascertain the study universe. If the survey is intended to focus on evaluation or program monitoring in a specific area that has been targeted for an action or intervention program, the facility list must be compiled for the program area so defined. If it is a global or general-purpose monitoring survey covering an entire country or sub-region, the facility list must pertain to that designated area.

---

[18]It is the *sampling variance* of the rural sample that would be about twice as big as the urban variance. The sampling error - the square root of the variance - would be about 40 percent larger.

For the study universe, however defined, ease of facility list compilation will vary by country. In all countries it is likely that public facilities will be easier to pin down than private ones. A key point to remember about the list frame is that it is the large ones which are of main concern; for that reason, it may be a fairly simple task to find appropriate authorities who can provide the needed list. Small facilities or health service providers that are missing from the list(s) compiled will not pose a problem for the survey statistically, because they will be covered by the area frame.

Box 4.2 List Frame Development

Contact

- ✓ Ministry of Health
- ✓ World Health Organization
- ✓ Community leaders

List

- ✓ Public hospitals
- ✓ Other large facilities

Record

- ✓ Name
- ✓ Location
- ✓ Type
- ✓ Size information

Large facilities will also be covered, in theory, by the area frame, but it is better to give them a very high probability of selection, which can only be done if they are on the list frame. The sampling variances - in particular for the staff sample and client sample (discussed in subsequent subsections) - would be unacceptably high if large facilities were given a low probability of selection through the area frame.

⇨ *Example*: If hospital A has client volume of 100 patients per day, the survey estimates for the client sample will be substantially improved by giving this hospital a probability of being selected that is equal or close to certainty (probability 1.0), which cannot be attained through the area sample.

It is recommended that the Ministry of Health (MOH) be the initial contact point for locating a list of large or otherwise significant facilities: hospitals, hospital centers, medical colleges. All public hospitals should be included by definition, irrespective of size, since the MOH would likely have a list of the public ones. For other facilities, either public or private, only the large facilities should be included. For this purpose “large” would vary according to the country’s own definition and requirements, but a suitable rule-of-thumb for inclusion would be private hospitals with 10 or more beds and other facilities with health-related staff size of 10 or more.

In addition to the MOH, international organizations such as the World Health Organization may be able to supplement the list, especially of the private hospitals and facilities. It would be useful, also, in conjunction with the area sample units, to inquire from appropriate community spokespersons the name and location of the nearest hospital. Those mentioned that are not on the list frame would be added to the latter, and sampled accordingly (or covered on a census basis).

Information necessary for sampling purposes - for the staff and client sampling operations

- that ought to be obtained for each facility on the list, at the time it is being constructed, includes

(a) its name and location
(b) type of facility
(c) (if available) estimated daily client volume
(d) (if available) staff size (health professionals and technicians), and
(e) (for hospitals) number of beds
(f) times of operation (days of week open to clients and hours of operation).

When information on client volume and/or staff size is not available at the time of frame construction, it will be collected when the facility sample is interviewed. Daily client volume, item c, may not be available but most facilities would have information on annual client volume from which the daily volume can be calculated. For some services of interest, such as family planning clinics, that are only available on certain days it would be necessary to find out the days of operation - item f.

Sample size and stratification

The overall sample size for a facility survey will vary from country to country, depending upon conditions, precision requirements and need for domain estimates. Still, the sample size from the list and area frames combined is likely to be several hundred facilities in most applications. It is also quite likely that the number of facilities on the list frame may be comparatively small, even if the study universe is an entire country. This is because the list frame includes only public hospitals and other large facilities. In many applications of the facility survey, therefore, the list frame may comprise 100 or fewer facilities.

It is recommended that all facilities on the list frame be included in the sample if the frame contains 100 or fewer entries. In other words, each one would be selected with 100 percent probability of inclusion, constituting a census in effect. Alternatively, rather than setting a somewhat arbitrary cut-off of 100 facilities, it might be advisable to calculate the sampling rate. If the latter is larger than one-half, then take all facilities. The alternative strategy may have some advantages when deciding sample size for domains.

RECOMMENDATION 4.4

- ⇨ If list frame contains 100 or fewer facilities, select all, or
- ⇨ Take all if sampling rate is greater than one-half.

It should be noted that optimum allocation procedures may be used to determine the appropriate distribution of sample facilities between the list and area frames. However, it is anticipated that adequate cost figures, needed for optimum allocation, for sampling and interviewing between the two frames is not likely to be available in most countries. It is nevertheless clear that travel costs, per sample unit, for the list sample will be much higher than the area sample; and, for that reason, it is felt the sample size for the list sample should be limited.

Selecting all the facilities on the list frame for the survey offers certain operational advantages in sample design and implementation. First, it precludes the need to stratify the list frame by size or type of facility, and, consequently, the need to determine optimum sample sizes by stratum. Operationally, this is a decided advantage if information on size and type of facility is not readily available when the frame is constructed or compiled. Also, urban-rural stratification is irrelevant when all the facilities are sampled with certainty. Third, problems of determining the sample sizes needed for domain estimates (and a suitable allocation scheme) are precluded when the facility sample is a census.

If the list frame is sizeable, however, that is, containing several hundred facilities, a sample should be selected after appropriate stratification. Stratification criteria that should be considered are geography, facility type and size. The size criterion may be any one of the following, whatever is available from the compiled list: number of staff, client volume (expressed as a daily, weekly, monthly or annual average) or number of beds (if a hospital).

RECOMMENDATION 4.5

⇨ For a large list frame, use stratification variables in combination, together with systematic sampling, for implicit stratification.

The stratification variables may be used in combination and, that, in fact is recommended, so that systematic sampling may be used in selecting the sample in order to achieve so-called *implicit* stratification. Either of two prototypical stratification schemes might be constructed, emphasizing again that it would be used only when there are large numbers of facilities on the list frame. The first, scheme A, is as follows:

Stratification scheme A (when measure of size is not available)

- Urban
  - Administrative area A
    - Public hospitals
      - Facilities listed geographically in serpentine[19] order
    - Other facilities
      - Facilities listed geographically in serpentine order
  - Administrative area B
    - Repeated as above
  - Administrative area C
    - Repeated as above
  - Etc.
- Rural
  - Above categories repeated.

[19]Serpentine ordering entails arranging the geographic units in a stringed, snake-like sequence, commencing at any corner (example, northeast).

Scheme B is identical except for the additional entry of the size measure, as follows:

<u>Stratification scheme B (when measure of size is available)</u>

	Measure of size
Urban	
Administrative area A	
Public hospitals	
Facilities listed in serpentine order	xx
Other facilities	
Facilities listed in serpentine order	xx
Administrative area B	
Repeated as above	
Administrative area C	
Repeated as above	
Etc.	
Rural	
Above categories repeated.	

When the facilities are stratified in schemes such as the ones above, and, in fact, recorded on a sample selection sheet in the sequence shown, then a systematic sample may be applied to select the facility sample. That would be the recommended approach if the sampling methodology called for proportionate allocation of the sample by stratum. Another technique would have to be used to select a disproportionate sample by stratum (for example, applying different sampling rates for the urban and rural parts - see further discussion below).

With scheme A, a systematic, equal probability sample, *epsem*, would be used, that is, each facility would be given the same chance of selection without regard to its size, since information for the latter is not available anyway. The sample would be selected by (1) calculating the interval, I = N/n, where N is the number on the list and n is the sample size, (2) choosing a random start between 1 and I and (3) applying the RS and I, successively, to the facilities on the list.

⇨ *Example*: Suppose the list contained 1240 facilities (N) and the sample size is 100 (n). The sampling interval, I, to be applied would be 12.4 (1240 ÷ 100), carried to one decimal place. A random starting point between 0.1 and 12.4 would be selected from a random number table.

Under scheme B - where the measures of size are available - either *epsem* or *pps* sampling (probability proportionate to size) may be used. *Epsem* is recommended whenever the facilities on the list are approximately equal in their size measures. The *pps* method is the appropriate method if the sizes of the facilities on the list are widely variable, but the effect of a *pps* sample is to over-sample the largest facilities, and is therefore biased unless compensated at a later stage of selection. The facility sample is a one-stage selection procedure, however, so the compensation would have to be handled through appropriate weighting.

For the reasons stated earlier with respect to choosing all list-frame facilities when there are 100 or less, the suggested sample size to use for large list frames is 100. Again, it is noted that the number of sample areas in the area frame, together with the list sample facilities, constitute the number of separate *locations*, as opposed to facilities, that must be visited in the survey. That number is expected to be a few hundred, at most - in the range of 300-500 - and not more than 100 of them should be selected from the list frame, because of the much higher travel costs, per facility, associated with the latter.

RECOMMENDATION 4.6

⇨ If list frame is large, select 100 facilities, using

⇨ *epsem* when measures of size are absent or roughly equal, or

⇨ *pps* when known measures of size are variable.

Sample size, allocation options

Other options exist for allocating the sample size for the large list frame. One such option is to sample all the facilities of a certain type with certainty, or probability 1.0. For example, all metropolitan hospitals and/or others that exceed a specific cut-off size would be selected. Then, subtract the number of such facilities from 100, and select the balance in accordance with stratification scheme A or B above.

Another option that may likely be quite applicable in many countries is to allocate the sample equally between urban and rural strata. In that case, differential sampling rates between the two strata would be required.

⇨ *Example*: As an illustration, if the overall sample size from the list frame is 100, 50 facilities would be allocated to each of the urban and rural strata. Suppose the number of urban facilities is 212 and the number of rural facilities is 75. The sampling interval, I_u, for the urban stratum would be equal to 212/50, or 4.24, and its random start would be between 0.01 and 4.24; for the rural stratum, I_r, would be 75/50, or 1.5, with a random start between 0.1 and 1.5.

Area frame

Defining areas

The areas making up the frame must possess three characteristics without exception. First, they must comprise the entire study universe without gaps. Second, they must be geographically defined with identifiable external boundaries. Third, they must be big enough, geographically, so that the sample size, in terms of the number of facilities contained within their boundaries, is sufficient to meet the precision requirements of the survey. There are other characteristics the frame areas should have in order to aid the sampling process, but these are less essential if not met exactly. One of these is that the areas should have already been mapped.

Another is that a relevant measure of size should be available for each area. Each of these five features is discussed below.

Box 4.3 Area Frame Features

- ✓ Complete coverage of study universe
- ✓ Geographically defined areas
- ✓ Areas big enough to meet sample size requirements
- ✓ Areas already mapped; or to be sketched
- ✓ Size measures to be determined

The geographic areas that make up the area frame must encompass the entire study universe, so that survey *coverage* is complete. If certain areas are omitted for whatever reason - lack of internal security, inaccessibility, etc., - the sample will be biased. It will not represent the facilities that are located in the excluded areas, and to the extent that these have different characteristics from the included areas, survey bias, usually of unknown magnitude, will result. The survey design team may decide there are certain areas which have to be excluded because of various difficulties in conducting the survey. In that case, they must be excluded *before* sampling and the study universe must be re-defined. It is especially important when presenting the results to state in the report exactly which areas are included in the study universe and which are not. If possible, the report should also provide a crude estimate of the number of facilities that are located in the excluded area(s).

The areas must be geographically defined with clear, unambiguous external boundaries. This is necessary so the interviewers can locate the sample areas and canvass them properly. The boundaries should be identifiable features such as streets, roads, lanes, rivers, streams, railroad tracks, important landmarks. It is also useful, if maps are available, for internal boundary markings to be delineated, as this helps the interviewer to follow a prescribed path of travel when canvassing. Census enumeration areas are often useful geographic areas that meet these criteria, but they may not be big enough (see next paragraph).

The requirement that the areas be big enough for a sufficient sample of facilities may be difficult to meet in practice. Paradoxically, it must also be small enough that it can be canvassed in a reasonable amount of time. The issue of the geographic size of the sampling areas and sample size, in terms of facilities, is discussed further below in the subsection on this topic.

Another useful feature to look for in deciding upon the definition of sample areas is whether maps are available. The maps, like good boundary demarcations, are valuable in helping interviewers locate the area and canvass it. In many countries, however, suitable maps do not exist or are seriously obsolete. Sketch mapping must be undertaken in that case, in order to provide a mechanism for quality control of the canvassing operation. The sketch map would entail the interviewer drawing a sketch of the sample area, including internal boundary markings, either prior to (preferred) or during the canvass operation, and indicating on the sketch the location of facilities.

RECOMMENDATION 4.7

⇨ Prepare sketch map of sample area when no suitable map is already available.

The final useful feature mentioned in the first paragraph of this section is that a relevant measure of size should be available for each area in the frame. The measure of size may be needed to establish the selection procedure (when *pps* sampling is used with systematic selection), but *its main use is to establish an estimate of the average size* (number of facilities) for area sampling units. Yet, a suitable measure of size for the frame units may not be available in many practical situations. First, the concept of relevant measure of size is troublesome. Ideally, number of facilities (even a crude approximation) would be the measure of size of choice, but that may rarely be available. By contrast, population size, which is often readily available for areas such as villages or census enumeration areas, may be inversely related to the number of facilities present. For example, facilities, especially in urban settings, may often be located in commercial areas where the average population size is small. Unless a variable is available that could be used as a proxy for number of facilities - for example, number of doctors or number of pharmacies - then the sample of areas would have to be selected with equal probability (which is not necessarily a bad solution - see more about this in the section below on sample size in relation to area sampling).

Stratification (urban, rural)

The nature of health delivery systems is likely to be quite different in rural communities as compared to urban ones. The number and distribution of facilities plus such important subjects as the quality of care given may be dramatically different. For this reason, it is recommended that any sample of facilities consist of urban and rural strata. The sample design including sample size considerations and sample selection procedures are likely to be different between these two strata.

RECOMMENDATION 4.8

⇨ Stratify the area sample by urban, rural.

Sample size - number of areas, number of facilities

The recommended sampling procedure in the area sample is as follows:

(1) Stratify areas in geographic order by urban-rural at the first level and administrative regions at the second.
(2) Select areas in a one-stage operation - primary sampling units, or PSUs.
(3) Canvass sample PSUs
(4) Locate and record all facilities or service providers within sample PSUs and conduct interviews in each of them.

The sample size that is needed in the area frame is equal to the total sample size less those on the list frame sample. It is quite likely that the average number of facilities to be found in an area on the frame will not be known in advance. Moreover, the average is likely to differ for urban and rural areas. Both are issues which have an important bearing on sample design. The average, if known, dictates the number of areas to be sampled.

⇨ *Examples*: If it were determined that the area frame portion of the sample in the urban stratum should contain 500 facilities (after subtracting, say, 60 from the list frame), and the average number of facilities per urban area is 4, then about 125 areas would need to be selected. If the average is only 2, then 250 areas would be needed.

The *number* of areas thus has to be evaluated because of the travel costs involved. It can be seen, therefore, that the average number of facilities per area is a function of its geographic definition and size.

⇨ *Example*: An area defined as comprising 4 city blocks would thus be expected to contain 4 times as many facilities as one defined as a single block.

Box 4.4 Definition of Sample Areas

✓ Big enough to meet sample size needed, but

✓ Small enough to be conveniently canvassed

An optimum methodology must therefore be developed to balance the geographic size of the area (that is, its definition or limits) to be canvassed against the number to be visited. The optimum would no doubt be different for the urban and rural strata. The issue is further exacerbated when no information exists on the average number of facilities per area, however the latter may be defined.

⇨ *Example*: In the Uttar Pradesh application[20] where sample areas were defined as villages or urban blocks, the average number of facilities of the type, "fixed service delivery point," per area was only 1.1, whereas the average of "individual service agents" was about 10. While the number of fixed points may be a reasonable average to assume for villages or blocks in other countries, the Indian health delivery system may be unusual with respect to definition and number of individual service agents and may not be a good model for other countries.

It was mentioned above that census enumeration areas often meet certain criteria that make them suitable candidates as areas for the area frame. Typically, the boundaries have been delineated and a recent map exists. The EA may also have a known measure of size in terms of counts of households and/or persons, though, as mentioned above, this may not be the

---

[20]See Singh, K. *etal*, "Estimating the Population and Characteristics of Health Facilities and Client Populations Using a Linked Multi-Stage Sample Survey Design," *Survey Methodology*, 23, 2, pp. 137-146, Statistics Canada, December 1997.

appropriate measure of size for facility sampling. The geographic size of an EA is usually conveniently constructed so that an enumerator or a team of enumerators can cover it in a day. In many countries, the census EA is defined as a village or an urban block, which further enhances its candidacy as an area frame unit. However, the EA may not contain enough facilities, on average, to justify its use. Preliminary work undertaken in Tanzania in preparation for its 1999 Facility Survey showed there is only one facility per 3-4 EAs.

RECOMMENDATION 4.9

⇨ In area frame, conduct small pretest to determine average number of facilities per urban area and per rural area.

The likelihood that a relevant measure of size does not exist *a priori* for the sample areas, however defined, suggests two strategies which should be followed, in finalizing the sample plan in a given country or setting:

(1) A small pretest should be carried out in the areas that are being proposed as PSUs, in terms of the *definition* of the area, for example, census EAs. The pretest would ascertain the number of in-scope facilities present in each EA chosen for the pretest. Its purpose would be to get a rough idea of the average number per sample area, so that the *number* of areas to include in the main sample could be calculated. The pretest could be confined to 20-30 areas, purposively chosen, but representing, in equal number, both rural and urban communities. It would probably be useful to gather counts on fixed SDPs and individual service agents, separately, in the pretest. The average number of facilities would then be calculated, separately for urban and rural areas, and the results would be used either (a) to fix the number of sample areas to select in the main sample or (b) to change the definition of the sample area to a larger or smaller unit, geographically. For example, if the computed average dictated that 800 areas would have to be sampled, it would be prudent and more cost effective to widen the definition from, say, an EA to the next highest administratively defined unit such as census crew leader district (usually a combination of 3 or 4 census EAs, communes, etc.). It is recommended that the starting point, however, for the pretest be the census EA.

(2) For the main sample, the sample areas (EAs or re-defined areas after the pretest) should be selected with *epsem.* It is expected that the number of facilities per sample area is not likely to vary very much, except perhaps by rural and urban, but the latter will be treated as separate strata anyway. Moreover, the number is expected to be fairly small (1 to 4 perhaps), in which case use of a *pps* sample would not be worthwhile.

RECOMMENDATION 4.10

⇨ Select area sample PSUs with *epsem*.

In the unlikely event that a country considering a facility survey will have relevant measures of size available for the proposed sample areas (EAs, villages, communes, etc.), the pretest (point 1 above) would be unnecessary of course. Further, if those measures of size suggest considerable variation within the two strata (but not between), then a *pps* sample might be considered over *epsem* (point 2).

Canvassing

The area sample will entail, as mentioned previously, canvassing the entire geographic area of each of the area sample units. Canvassing requires locating, hopefully with the aid of a map, the sample area, following a systematic path of travel within the sample area and identifying all facilities or service providers that are present. It will also require asking questions of likely respondents to ascertain the presence of a facility. While this probably would not necessitate knocking on the doors of structures that are obviously intended to be residential only, it may be necessary to inquire at every non-residential building. In villages or other rural EAs it is also useful to conduct an interview with the village chieftain or other knowledgeable elder about the location of various facilities before proceeding. As mentioned, it is not likely there would be more than a few - 1-4 - facilities in each sample area, so the canvassing must be as thorough as possible to ensure complete coverage.

Box 4.5 Canvass Procedures

- ✓ With map, locate sample PSU
- ✓ Follow defined path of travel to locate facilities
- ✓ Inquire from community leaders
- ✓ Conduct substantive interview in every facility found
- ✓ Collect information needed to sample staff and clients

It is to be expected that the number of facilities found in each sample area will be small enough in number so that canvassing and interviewing for facility information can take place in a single visit. In some unusual cases, however, the number of facilities may be too large to cover. A second visit may be necessary.

In each facility identified in the canvassing operation, survey instruments will be administered (a) to elicit substantive data about the facility and the nature of services provided and (b) to obtain a list of staff, by type, and an estimate of client volume to be used for sampling purposes (see subsections below on staff and client sampling). It is recommended that the data under (b) be collected on (or transcribed onto) a facility listing sheet, apart from the substantive questionnaire. Note that it is important to obtain a list of the staff, rather than merely a count, since a sub-sample of staff will be selected in large facilities (more than, say, 3 staff members).

Regarding sample size, it may appear at first glance that the canvassing procedures do not permit adequate control of the sample size. Recall, however, that in the preceding subsection a procedure was given to determine the average number of facilities per area, as part of the overall sample plan. The *total* sample size is therefore appropriately specified and hence controlled fairly precisely, though there is likely to be quite variable sample sizes from one area to another.

Two-phase sampling

In some countries, the survey operation may involve two-phase sampling of facilities, depending upon the measurement objectives. This would occur when a country is interested in obtaining some basic, limited information about all facilities during the canvass, but only those which are determined to be within the scope of the survey will be interviewed or sampled for administration of the main survey questionnaire. An example might be an initial canvass to identify facilities and find out which ones provide family planning or maternal health services. The initial interview may involve only a few questions, while those facilities determined to be in-scope would be given the full interview. If the full interview is administered at the time of the initial visit to all in-scope facilities, the sample would be one-phase. However, if the in-scope facilities are sub-sampled (preferably back in the central office by sampling staff), it would be two-phase. In this case, two visits would be necessary to obtain the facility data. It should be noted that a two-phase sampling operation such as this (with sub-sampling from the initial list of facilities identified in the area frame), would likely be necessary only when large numbers of certain types of facilities may be found (for example, individual service agents). *Epsem* sampling would be recommended when sub-sampling is needed in a two-phase operation.

The question of whether two-phase sampling may be applicable for screening pertains to the list sample as well as the area sample.

It should be noted, also, that for field planning, more than one visit to the facility will most likely be necessary irrespective of whether the facility sample is one-phase or two-phase. This is because the facility will serve as the focal point for the staff and client samples, both of which will most likely be selected at a later date, thus necessitating a subsequent visit.

Un-duplication with list frame

Because the area sample covers the universe in its entirety, the facilities which are on the list sample frame have duplicate chances of selection, *whether actually selected or not*. During the canvass operation, the interviewer should be provided with a list of the facilities that are on the list frame for his/her sample area. Any such facility encountered during the area canvass should be skipped (not interviewed), since it will have already have had its chance of inclusion in the list sample. Whenever an area sample facility is excluded for this reason, appropriate notation of this fact must be made by the interviewer in order for supervision and quality control to be effective.

RECOMMENDATION 4.11

⇨ Compare each area sample facility with list frame and do not interview if on list.

Staff sampling

The canvass operation, described above, will provide a list of staff for each facility identified and selected from the area frame sample. In addition, the same kind of information will be collected for the facilities selected from the list frame sample. These data will constitute a sub-frame to be utilized in selecting the sample of facility staff.

Sample size

The sample size to be used for interviewing facility staff must be at least as big as the facility sample but probably larger. Thus, if 600 facilities are selected, at least 600 staff interviews must be obtained as well, since we will want to have staff data associated with each of the facility sample units.

The maximum sample size for staff, on the other hand, would occur under a sampling strategy where all relevant staff (health-related occupations) would be interviewed in each sampled facility. This in fact is the recommended strategy in small facilities, those containing 4 or fewer staff. For large facilities, sub-sampling of staff should be used, so that an average of about 3 staff members per facility is selected and interviewed. The sample sizes and selection rates to be used in large facilities are shown in Table 4.2.

Table 4.2 Sample selection rates for staff sampling within facilities

Number of staff members	Sample selection rate	Number of sample cases
4 or less	1 in 1 (select all)	1 to 4
36652	1 in 2	3 or 4
36747	1 in 3	3 or 4
36842	1 in 4	3 or 4
14-16	1 in 5	3 or 4
17-19	1 in 6	3 or 4
20-22	1 in 7	3 or 4
23-25	1 in 8	3 or 4
26-28	1 in 9	3 or 4
29-31	1 in 10	3 or 4
etc.	*****	3 or 4

The last column of Table 4.2 shows that the number of sample staff members in each facility will be either 3 or 4. The exact value will be depend on the random starting number.

The choice of a fixed number of staff per facility has two advantages. Overall sample size is controlled exactly to agree with the desired sample size. Interviewer workloads are also equalized per facility. The determination of the number of staff to select into the sample should be based on budgetary considerations, in terms of the number that can be conveniently interviewed. This, in turn, is based on the number of facilities in sample. Thus, if 500 facilities are selected, based on considerations of precision requirements discussed earlier, and it is determined that 2000 staff interviews can be accommodated, the sample size per facility would be computed as 2000/500, or 4 - again with facilities containing 4 or fewer staff being designated as "take-all" cases.

> **RECOMMENDATION 4.12**
>
> ⇨ Select all staff members in facilities with 4 or fewer, and
>
> ⇨ Select 3, on average, in larger facilities.

An alternative strategy would be to select staff at a fixed rate in every facility. An advantage would be that different rates of selection would not have to be figured for each facility and thus some errors in field implementation might be avoided. Disadvantages are that neither the overall sample size nor interviewer workloads would be controlled.

Stratification and sample selection

Stratification for staff sampling will be necessary for large facilities only, since all staff members will be chosen in small ones. Most of the large facilities are expected to be those that are selected from the list frame, although a few may be sampled from the area frame as well.

For sample selection, which is to be done at the level of each selected (large) facility, it is recommended that the list of staff members compiled in the facility interview be stratified by arranging them on (another) list in a particular sequence, such as:

Staff member	List (by type)
1	physician A
2	physician B
3	physician C
4	physician D
etc.	
7	nurse A
8	nurse B
9	nurse C
10	nurse D
etc.	
15	technician A
16	technician B
etc.	

> **RECOMMENDATION 4.13**
>
> ⇨ Stratify staff by type in large facilities.

Then, from the ordered list, a systematic sample should be chosen in accordance with the values in Table 4.2.

➩ *Example*: Suppose the ordered list contains 18 staff members. Table 4.2 shows the sample rate should be 1 in 6. Choose a random starting point between 1 and 6 from a table of random numbers; suppose it is 2. Select staff members number 2, 8 (that is, 2+6) and 14 (8+6) for the staff sample.

Ordering the list in a prescribed way, such as the one above, and then using systematic sampling for the selection method achieves implicit stratification, as mentioned previously in the discussion about selecting facilities from the list frame.

In some applications of the procedures, instances may arise where the sample of staff is selected but not interviewed until after sampling and interviewing for clients and client-interactions have been completed. This may present situations in which the selected staff members will no longer be available for interview that day for one reason or another. In such cases the temptation to substitute should be avoided if at all possible, as the procedure is biased. Instead, at the time of selection arrangements should be made with the selected staff member to schedule an appointment for his/her interview, either later that day or the next.

Client sampling

Sample selection of the client sample associated with the sampled facilities poses data collection issues that are more difficult to resolve than the staff sample. Organizing the field work to collect data for the staff sample is comparatively simple, since the staff members selected can be expected to be present at the facility during normal working hours and can thus be contacted and interviewed then. By contrast, clients of the facility come and go at will. An interviewer wishing to interview clients at the facility is constrained by the number and type of clients that happen to show up when the interviewer is present. In a given day, for example, it may not be uncommon in a small facility for no one to show up for consultation.

As discussed in chapter 1, the estimates wanted for the client sample are percentages receiving certain services or with certain attributes. The denominator of the percentage must be estimated in terms of client-visits in a specified time frame, a necessary requirement in order to properly calculate the sample weights for the client sample. This issue and those mentioned in the above paragraph have implications on sample size, selection methods and interview methodology, discussed in the subsections below.

Sample size - precision issue

The sample size for the client sample is difficult to determine in advance of the facility sample operation. Information such as client volume that is needed to determine client sample size is not likely to be available for the area frame until after the interview for the facility sample has been conducted. Some information about client volume may be available for the list frame units, though even that is likely to be incomplete or inaccurate. As a result, it is likely that the sample size for the client sample will have to be determined after initial results are available from the facility sample data collection.

In the facility sample interview, therefore, information about client volume must be collected, as mentioned previously. This will then be used by the sampler and survey team to decide on the size of the client sample and the best method for selecting the sample and conducting the interviews. Client volume information that should be entered on the facility survey listing sheet includes average number of clients (for example, total, maternal health, children). The averages may be expressed, depending upon the information available at the facility, as daily, weekly, monthly or annual. The days of the week that the facility is open for patients and its hours of operation should also be collected in the facility interview (see appendix 2 for illustration of client data sheet).

Although it might be expected that survey sponsors would want the same general order of magnitude in the reliability of the client estimates as the facility estimates, it is actually more sensible to determine a feasible number of clients to sample per facility, rather than base the overall sample size on a predetermined precision requirement. These considerations are similar to those discussed above in choosing the sample size for the staff sample. A sample in the range of 3-6 clients per facility is a reasonable target to consider. The reliability of the client estimates would be expected to be somewhat better than facility estimates, with a sample in that range, even though the design effect for the client sample may be higher than the facility sample. The larger sample would more than compensate, however, for the increased design effect. For example, with an expected sample size of 4 clients per facility, the sample size for the client sample would be, obviously, 4 times the facility sample size.

Sample size per facility

As recommended above, the sample size per facility should be in the range of 3-6 clients, on average. Depending upon objectives in a particular country, the number of clients to sample per facility might vary by type - for example, 4 maternal health clients and 2 "others." Selecting fewer than 3 clients per facility would not be cost effective, given the consequences of having to travel to each sample facility for only 1 or 2 client interviews. More than 6 clients would, by contrast, likely pose a budget problem in the other direction, that is, the overall sample size for clients could easily exceed the budget, especially when the facility sample itself is large.

⇨ *Example*: If the facility sample size were 600 and 7 client interviews were wanted in each one, the client sample would be 4200.

RECOMMENDATION 4.14

⇨ Select in the range of 3 to 6 clients per sample facility.

The actual number of sampled clients would no doubt vary from the average, perhaps considerably, across facilities. Controlling exactly the total sample size, not to mention the average size per facility, will also be very difficult because of the variability with which clients appear at the facility on a given day. This problem is exacerbated if clients are sampled by type of condition because there may be a conflict between number of clients by type and the frequency of their visits.

The selection method is a crucial design feature in implementing a plan to interview a predetermined expected number of clients in each facility, and that is discussed in the following sub-section.

Sample selection of clients and interview methods

To consider the sample selection methodology for the client sample, it may be prudent to illustrate the procedures with a specific sample size for the number of *client-visits* to be sampled per facility. For that purpose, we will assume that it has been decided to select, on average, 4 clients in each sample facility. Further, it is recommended that exactly one full day of interviewing will be devoted to gathering the data from the selected clients, and that has been assumed in the calculations which follow. Note however, that in facilities which have a very small client workload, the number of clients available in a single day will likely be less than 4.

The procedures below should be carried out separately for urban and rural strata, although they are identical for each. The procedures are easily extendable for values other than 4 for the desired client sample size per facility.

RECOMMENDATION 4.15

⇨ Fix the length of client interview time in each sample facility to exactly one day.

The steps for the client sample may be summarized as follows (with detailed discussion following the brief list):

(1) For each sampled facility, select a random day of the week to conduct the client interviews.
(2) Visit the facility for one full day of client interviewing.
(3) Look up the average daily volume of clients in the study universe.
(4) Select clients who show up at the facility on that day, in accordance with the figures in Table 4.3.

The particular day of the week which is chosen for the interview may present some bias for a given facility if the latter normally has an unequal client volume based on day of the week. This might be the case, for example, in facilities where clients tend to come in greater numbers on the first day after the weekend (Monday in most Asian, African and Latin American countries, Saturday in Muslim nations). The frequency of various types of visits may vary by day of week also - more routine visits on the first day of the week, for example. Since it is not feasible to obtain daily client volume for different days of the week, it is recommended that the day chosen for interviewing clients be other than the first day following the weekend[21].

RECOMMENDATION 4.16

⇨ Choose an interview day other than first day after weekend.

It is extremely important that the interviewing team stay at the facility for one full day, no matter how many clients show up. It is equally important that it be only one day and not more, even if less than 4 clients have been selected. While the interviewer may be tempted to extend the interview period over two or more days for a facility whose client volume is so small that four clients do not show up in a single day, to do so would constitute a biased procedure. Similarly, the interview procedure would be misapplied if an interviewer departs whenever the fourth client is completed; it is the length of *time* spent which determines the proper probability, not the *quota* of four clients. Strict adherence to these instructions is the only way to ensure that the probabilities of selection can be accurately computed - for unbiased estimation.

The information on client volume is found on the facility listing sheet that was obtained during the facility interview. The daily volume may have to be computed, depending upon how volume data were collected. If a facility provided weekly volume, the daily volume is computed as that number divided by 5 (or divided by the number of days during the week that the facility is open to patients/clients). Similarly, an entry in terms of monthly volume would have to be converted to daily volume by dividing by the (average) number of days per month that the facility is open.

The facility listing sheet may contain volume information for all clients and a separate item for clients seeking maternal health services. Depending upon the measurement objective of the survey, the figure to use is the one which corresponds to the in-scope universe.

Sampling of clients must be done by the interviewing team. The sample should be selected systematically as clients come in for consultation. The rate of sampling is given in Table 4.3.

[21]An exception would be in facilities that only offer certain, target services on set days of the week. This should obviously be taken into account before ruling out the first day after the weekend.

Table 4.3. Sampling interval for selecting average of 4 clients

Average daily client volume	Sampling interval
5 or less	1 (take all)
36685	2
36811	3
14-18	4
19-21	5
22-26	6
27-29	7
30-34	8
35-37	9
38-42	10
etc.	*****

Daily client volumes are shown in Table 4.3 as whole numbers. Thus, if the recorded volume for a given facility is fractional, it should be rounded to the nearest whole number to use the table. The variable-width ranges chosen in the left column are those necessary for the integral "take every" of the right column, so that an average of 4 clients per facility is selected. When the recorded volume is 5 or less, all clients will be interviewed, irrespective of the number that happen to show on the day the interviewing is scheduled.

In selecting the sample, the interviewing team must begin with a random start and apply the interval shown in the second column to select the clients for interview.

⇨ *Example*: Suppose the facility has average daily client volume of 16; Table 4.3 shows the sampling interval to be 4. A random number between 1 and 4 must be selected; suppose it is 2. Then the second client who comes in for consultation that day will be sampled and subject to interview and client-staff observation. Every fourth client thereafter will also be sampled and interviewed.

It is very important to keep track, accurately, of the number of clients that come in, so that the sampling interval is applied correctly. Therefore, a listing sheet of clients must be used for the sampling operation. The listing sheet can be very simple in form, containing only the names of the clients in the order in which they appear at the facility (see appendix 2 for illustration of client listing sheet). The sample would then be selected from the list so compiled. The random start and sample selection numbers should also be entered on the blank listing sheet prior to the visit by the interviewing team to the facility, because the sampling interval will be known in advance.

⇨ *Example*: In the preceding example, the random start would be 2 and the sample selection numbers would be 2, 6, 10, 14. These numbers would be indicated on the blank listing sheet and the particular clients that fall into the sample would correspond to those whose names are listed on those lines. The sample selection numbers should be extended beyond the expected 4 sample cases, since the exact number of clients that will show up on a given day will vary. In our example, therefore, the selection numbers might be extended to 18, 22, 26, 30, etc.

Note, the sample must include every k^{th} client on the listing sheet (where k is the sampling interval), no matter how many the list contains, even when the sample results in many more than the expected 4 cases.

From the listing sheet important information should be available for use later in weighting and survey analysis. The total number of clients, as mentioned in the above paragraph, is needed of course to establish the sampling rate. In addition, the number of clients selected and the number participating should also be recorded, so that response rates and, possibly, non-response adjustments for use in the weighting procedures can be calculated.

In facilities with high volume or where many clients show up at once, it may not be feasible to have the selected clients wait in a queue to be interviewed. It would be better if their names and addresses could be obtained so that the survey interview may be undertaken in their residences later the same day or the next day. Client-staff observations, however, probably cannot be delayed without unduly burdening both respondent groups. For that reason the survey team may have to be large enough in big facilities so that all sampled clients can be observed with their staff counterpart at the appointed time. A related issue in high volume facilities is those that have multiple entry points; in such cases it would be necessary to have a team of interviewers that could stake out each entry, in order to ensure unbiased coverage of the sample.

Again, strict adherence to the procedures described in the above paragraphs is crucial, or else the resulting estimate of the total number of clients will be biased. The estimate would be understated, for example, if interviewers have a tendency to end the interview process when 4 clients are sampled, rather than accurately applying the sample interval to all clients on the list.

Considerations of the alternative strategy of sampling clients at a fixed rate over all facilities are the same as those for the staff sample discussed in the preceding section. The sample size would be out of control, as would interviewer workloads.

The procedures described in this section apply for the situation in which all targeted (in-scope) clients are sampled and treated as a single group. The sampling method makes no differentiation by type of client. In country applications where the types of clients are important measurement domains, separate strata may be set up. For example, sick children and antenatal patients may be treated as separate strata. The procedures described in this section would then be carried out separately for the two strata.

Interviewing over a longer period

Another alternative is to conduct the client sample over a longer period - say, 2 days instead of 1 - in order to increase the chance of getting at least 4 clients in the very small facilities. Whatever the number of days that is settled upon, *it must be fixed and used for all facilities,* large as well as small, for the estimation procedure to be unbiased. For an interviewing period longer than one day, the numbers in Table 4.3 do not apply and would have to be re-computed, depending upon the number of days and whether the expected client sample size per facility is 4 or a different number.

Summary of preferred sample design

The preferred sample design for a stand-alone facility survey makes use of a dual-frame, probability sampling methodology. Because of its probability nature, the survey may be linked with population data to analyze facilities and service providers in relation to population outcomes, either for monitoring or impact assessment. In the preferred design a maximum of 100 facilities is selected from a list frame composed of large facilities and several hundred (300-500) area sample units are selected from an area frame comprising the entire geographic universe. All facilities located in the area sample PSUs are sampled and interviewed, giving an overall sample size of several hundred facilities. The selection of facilities from both frames is a one-stage sample, chosen systematically with equal probability in most country applications, although *pps* sampling may sometimes be necessary. Implicit stratification, used in combination with systematic sampling, is employed in both frames, to achieve a proportionate geographic distribution of the sample for major administrative areas and for urban, rural areas separately.

Since list-frame facilities are also represented on the area frame, an important facet of the survey methodology requires that sample facilities on the area frame be screened to ascertain whether they appear on the list frame, whether selected from the list frame or not. All area sample facilities found to be on the list frame are eliminated from interview.

Within each sample facility, the information collected is not only that which is necessary to meet the substantive requirements of the survey but also some basic data on staff (a list of these by type) and client volume - to be used for sampling purposes. A systematic sample of up-to-3 staff members is selected with equal probability from the compiled list. The staff sample size is, thus, about 3 times larger than the facility sample.

For each sample facility a client sample is selected through the technique of having the interview team visit the facility for one full day and conduct an interview with an average of 4 clients that show up for consultation on that day. The sample of clients is selected systematically with equal probability from a list, compiled by the interview team on the day of interview, of all clients that appear. Client-staff observation data are also collected at this time - for each sampled client. The size of the client sample is thus about 4 times the size of the facility sample, though somewhat less because many small facilities do not have as many as 4 clients on a daily basis.

Throughout the selection process, the use of probability methods is strictly adhered to.

Alternative approaches

Sampling completely from a list frame

For program monitoring of a project's own facilities, the recommended approach would be to use a list frame only. This is also the recommended approach for monitoring or evaluating an action program that is confined to a comparatively small geographic area. Such an area might be a district, a metropolitan area, a set of villages, a single city. The operative definition is that the area must be limited enough in geographic size so that it is feasible to (a) compile a complete list of facilities through field canvass and (b) avoid great travel distances between those facilities ultimately selected into the sample. An action program area as large as a state or province would probably not qualify on those grounds in most countries.

Except for the case where a project's own facilities are being monitored, it is assumed that a complete list of facilities does not exist prior to the survey, although a partial list may be available as a starting point. Therefore, a two-phase field operation would be necessary to conduct the survey. In the first phase, any pre-existing list(s) of facilities would have to be supplemented by a field canvass of the entire program area. Its purpose would be to locate and record all facilities so that, when combined with lists already available, the complete universe of facilities is available for sampling. The methods for conducting the field canvass are the same as those previously described in the subsection in this chapter entitled, "Canvassing."

The second phase of the field work entails conducting the facility interview in a sample of the facilities selected from the universe list. The considerations for stratification, sample size, method of selection (systematic, *epsem* or *pps*), staff and client samples are the same as described in the subsections with those titles.

Otherwise (that is, for much larger geographic areas such as provinces or nationally), it is not recommended to use a list-frame only sample design, because the expense of compiling the universe list and the travel costs associated with, essentially, a randomly distributed sample of facilities would likely be prohibitive.

Sampling completely from an area frame

In some instances, an area-frame only sample may be necessary. This would occur in a country that has no lists at all of health facilities. A key problem with this approach is the difficulty of controlling the sample size. It would be strongly recommended that preliminary work be carried out, in the form of a pretest, to obtain information to help define the geographic size of area sample units - PSUs - and to obtain a rough estimate of the expected number of facilities per PSU. A pretest for this purpose is the same as that described previously in this chapter.

Chapter 5
SAMPLING FOR LINKED FACILITY SURVEYS
(FACILITY AND POPULATION SURVEYS IN SAME SAMPLE AREAS)

This chapter discusses
- # sampling, survey approaches for linking with population survey areas
- # preferred "linked" design
- # survey and sampling methods
- # extra data needed for estimation

Introduction

In this chapter suggested guidelines are given in detail on sampling approaches for a facility survey designed to be conducted in conjunction with a relevant population survey such as DHS. More specifically, the facility survey is to be conducted in the same sample PSUs or clusters as the population survey plus surrounding clusters. In that context, the facility data (as explained in chapter 2 under the section, "Sampling issues"), when analyzed together with population data from a population-based survey, can be used either (1) to evaluate program performance including target outcomes, but without attempting to establish a causal relation or (2) to assess program impact more rigorously by examining the causality dimension. The essential differentiation between these two is in the analytical techniques used, unless the research design is a carefully controlled experiment. Investigating causality, or plausible association, requires multi-level and multi-variate analysis to control, statistically, for various external factors.

The facility data that are needed to accomplish either (1) or (2) above are the same. They just need to be linked with outcome variables for the relevant population. An issue in sample design and analysis is trying to relate the geographic outreach of a given facility (so-named its *catchment* area) with a suitably defined geographic area of community residents who may potentially use it - the latter representing the targeted population for whom information about outcomes is sought. Women and caretakers of children may travel beyond the borders of their residential communities to seek health care, while facilities draw their clients from several communities.

When there is a relational link between the individuals whose health outcomes are being studied and the individuals' health services supply environment (which is provided by a facility survey based on population survey PSUs), the statistical models available for analysis are better estimated. Perhaps equally important, there is strong justification for carrying out a facility survey and population survey in the same geographical areas quite beyond the statistical issue of linking the two data sets for analysis. It is advantageous to plan the two surveys in tandem to coordinate field work and to ensure that the questionnaires complement one another. Considerable cost savings will be achieved if the same sample areas are used for both surveys. Moreover, both survey coordination and cost savings would be increased if the sample designs for the facility survey and DHS (or similar) were, ideally, designed together.

Approaches used

A number of sampling approaches have been used or proposed for collecting facility data in relation to the sample areas that are selected in a relevant household survey. Each has its limitations. The main approaches are as follows:

(1) In the sample PSUs (or clusters) of DHS or a similar population-based survey, ask the respondents about the facilities they use and conduct the facility survey in all or a sample of those mentioned.

(2) Inventory the facilities, through canvassing, located within the boundaries of the population-based sample PSUs and conduct the facility survey, again on a census or sample basis.

(3) Determine the nearest facility, or facilities, to each population-based survey PSU (if not within its boundaries) and conduct the facility survey in all identified.

(4) Construct one or more concentric rings (with variations) around each population-based sample PSU, inventory the facilities contained within the ring boundaries and conduct the facility survey on a census or sample basis.

(5) Identify districts (or other administrative units) to which the population-based sample PSUs belong, inventory the facilities in the whole district and conduct the facility survey in a sample of those identified.

Each of the 5 approaches yields a sample of facilities that is clearly linked to the population survey, and on the surface, they may all appear to be valid methodologies. Statistical, logistical and other implementation problems beset each one, however. The "source" for the facility sample is the population-based sample, whether DHS or another, which is designed for representation of, usually, women of child-bearing age, not facilities. As a result, the distribution of facility types generated from that sample may have higher sampling variances than one designed explicitly for facility measurement. Though the facility estimates may be improved through post-stratification methods making use of independent data on facility totals and types, adequate independent data for that purpose are not likely to be available in most countries. Many of the usual considerations for designing an efficient sample plan are precluded when the population-based PSU determines the facility sample -- considerations such as stratification by type of facility, allocation schemes for domain estimates, determination of sample size to meet pre-specified precision requirements.

There are additional disadvantages with the particular approaches. A key one for (1) above is that under-utilized facilities are likely to be under-represented in the sample, because respondents will tend not to mention them. In approach (2) the sample size is likely to be so small that it would be rendered infeasible in most countries. Under approach (3), sample size is improved by assuring that, at minimum, the number of facilities is no less than the number of population-based clusters. The main disadvantage of the approach is determining the probabilities of selection and the extra information that must be gathered to do so. A complex set of rules for doing this must be followed, which may not be well-understood by most practitioners.

The sampling objective of approach (4) is to build up the facility sample size to an adequate level. This method, too, requires development of procedures to establish the populations of surrounding clusters for use in weighting and estimation. Those procedures could be fraught with inaccuracies in field implementation, unless carefully controlled. A variation of the method, which utilizes an area of fixed radius from the center of the population-based sample PSU, would result in nebulous and non-natural boundaries on the outer ring, so that coverage error (in terms of which facilities to include or exclude) is apt to abound. Still, the variation of approach (4) that utilizes concentric rings around an index cluster defined as the population survey PSU is likely to be the most feasible option, operationally and statistically. From a analytical point of view defining an area around the population survey PSU provides a proxy for the service supply environment to which the individuals in the PSU are exposed.

Approach (5) also builds up sample size. But unless the sample universe is redefined to include only the specified districts, the approach otherwise presents some of the same issues for estimation as approach (4). Redefining the universe may not be a suitable option either, because the ensuing estimate of facilities pertains only to the restricted universe, not nationally. Another disadvantage, though perhaps not a major one, is the assumption that the districts constitute a closed set for service delivery and utilization. A more important logistical disadvantage is the necessity of having to inventory entire districts or other administrative units that will, by definition, be substantially larger than the population-based sample PSU. Depending upon the definition and size of "districts" this might involve most of the territory of the country, since, in some countries, most if not all districts would contain at least one sample PSU. The situation then reduces to the case where a list of facilities has to be compiled to constitute the frame before sampling, and linking the facility survey to the population-based PSUs becomes essentially irrelevant.

The rest of this chapter will focus on approach (4) as the most plausible, and that will form the basis for our recommendations. We will present a preferred design and discuss variations as appropriate. In some instances the particular details of sample implementation will be identical to those already discussed in chapter 4, in which case the reader will be referred to those sections.

Preferred survey and sample design - essential features

This section provides recommendations on sample design for a facility survey intended to be linked directly with the same sample PSUs of a population or household survey on health. The sample will come from facilities either located in sample PSUs chosen for the population-based (household) survey or surrounding PSUs.

The essential features of the preferred design can be summarized in the following 7 steps:

1. In each PSU or cluster selected for the household survey, referred to hereafter as the index cluster, canvass its geographical area to identify all facilities contained within the boundaries.
2. In addition to the index cluster, identify one or more *rings* (but preferably one) of PSUs surrounding it.
3. Canvass the totality of the geographical area in the *ringed space*[22] to identify all facilities.
4. Compile a list of all facilities and service providers so mentioned in steps 1 and 3.
5. Conduct the facility survey interview in all or a sample of the facilities mentioned.
6. In a post-survey office operation, gather population data for use in calculating weights for the estimation phase.
7. In an independent operation, include in the facility survey all large facilities, irrespective of location.

The rest of this section gives advantages and disadvantages of this method, followed by a detailed presentation of the survey methodology and sampling methods. A brief discussion of the reliability of the facility estimates is presented, noting that it is highly dependent on the household survey sample design. The section concludes with the observation that recommended methods for staff and client samples are the same as those in chapter 4.

Advantages and disadvantages

The methodology permits national, unbiased estimates of the universe of facilities and their attributes. To the extent that the sample is large enough, facility estimates by type and geographic domains are also possible.

A major advantage is that many of the usual concerns of sample design are obviated by

---

[22]The term, ringed space, refers to the index cluster plus the concentric ring(s) of clusters that surround it.

the method, and sampling thereby is greatly simplified. Potentially difficult problems of stratification and sample allocation for domains including type of facility or urban-rural and other geographical sub-regions are precluded, because these facets of the design are pre-determined by the population survey sample design. Facility sample size is also pre-determined to some extent, so that budgetary considerations are more likely to guide the sample size than precision requirements. Perhaps the most significant advantage is that, in many countries that apply the method, sampling may not have to be used at all at the facility phase (it will be needed, however, for staff and client surveying). Instead, all facilities identified in the canvassing operation would be covered on a census basis.

Paradoxically, pre-determined limits on the sample size may also pose significant disadvantages, if the maximum sample size available is too small, especially for important domain estimates that may be wanted. The problem of sample size can be ameliorated by using not one, but two concentric rings around the index cluster.

Determining probabilities of selection (and weights) for the methodology necessitates the compilation of population or household counts for non-index cluster PSUs from the original frame that was used to select the index cluster.

A somewhat minor disadvantage is that program monitoring of the type that would be necessary to focus on a small project area would not be well served by the method (although national-level program monitoring would be). This is because the household survey clusters that would fall within the project area are likely to be few if not none. More importantly, the selection of those clusters for the household survey is likely to have been based on a sample plan totally unrelated to the selection of the project area. However, monitoring of a samll project area may need an special sample design or use of list sampling.

Survey methodology

The section title, "Preferred survey and sample design - essential features," mentions both survey and sample design, because the survey procedures to implement this option are inseparable from the sampling requirements. Sampling of facilities *per se* may not even be necessary, as mentioned and further discussed later.

> **RECOMMENDATION 5.1**
>
> ⇨ Plan facility survey *together* with linked household survey

The method is best used when the facility survey and the household survey are planned and developed in tandem. The type of household survey that is to be used should be one designed to measure health variables of women and children, notably DHS, Gulf Family Health Survey (GFHS), Multiple Indicator Cluster Survey (MICS), PAPFAM Survey (Pan Arab Project for Families and Maternal Health) or others with similar content and design. It should be national in scope and probability-based. Coordination and collaboration with the organization responsible for the household survey must be close, continuous and collegial in order for the facility survey to be successful. It is essential for the facility survey manager to obtain the cooperation of the household survey organization well in advance of the fielding of either survey. As alluded earlier, ideally the sample designs for both surveys would be developed in concert.

The first three steps in implementing the facility survey entail mounting a canvas operation in the index cluster and its surrounding ring or rings. Procedures for canvassing to locate and identify facilities are the same as those described in chapter 4 in the subsection, "Canvassing," and are not repeated here. Carefully constructed instructions to the interviewer on what types of facilities should be recorded when mentioned by respondents have to be developed. Normally, facilities of the types listed in chapter 2 are in scope, yet respondents may mention any number of out-of-scope services including dental clinics, de-toxification centers, witch doctors, acupuncturists, faith healers, midwives or other traditional healers that may be out of scope. Alternatively, the interviewer might be instructed to record any mention, and the determination of whether it is in scope would be made back in the office by health workers that may be more qualified to do so than interviewers.

Where health facilities are sparse, there may be a substantial number, perhaps even a majority, of index clusters that yield no facilities in the canvass operation. That is the reason that one or more rings of PSUs surrounding the index cluster will also be canvassed. The procedure requires identifying the index cluster on the original frame that was used to select it in the household survey. The original frame maps must thus be utilized. On the map, all PSUs that are contiguous to the index PSU are marked and identified in a one-ring design, and these constitute the set of PSUs to be canvassed. If the design is two-ring, PSUs contiguous to each PSU in the first ring are also included. Considerations in choosing one or two rings are discussed further below.

A variation of the preferred procedure that might be considered for step 3 is to utilize the index cluster to identify and inventory pharmacies, private doctors, service delivery points, but the ring for all other types of facilities, or only the large ones.

> **RECOMMENDATION 5.2**
>
> ⇨ Identify all facilities in the household sample PSU and a concentric ring of PSUs surrounding it.

Step 4 of the operation calls for organizing the facility mentions into a coherent list of facilities for surveying. Each facility on the list must carry the household survey PSU identification number. This information is needed later for step 6 when further information is gathered in the office to help calculate survey weights (discussed below under the next subsection). The facility survey is conducted in all or a sub-sample of facilities identified in the index cluster and its ring(s) - step 5. In addition, large facilities are also included in the survey, no matter where they are located - step 7 - discussed in the subsection below on the special certainty stratum.

Sampling method - facility sample

The sampling space for the facility survey - in terms of facilities available for sampling - is the list of those identified, as determined through application of the survey procedures discussed in the subsection above. The number, N, of such facilities cannot be known or controlled precisely[23] in advance of the household survey operation. The number can be estimated (described below), and this has a bearing on whether one or two rings of PSUs around the index cluster would be canvassed.

The expense and complexity of the facility survey increase if two rings of PSUs are canvassed, as opposed to a single ring. If we speculate that a typical PSU is surrounded by up to 6 others, then with one ring, the number of PSUs to canvass is approximately 7 times the number of PSUs in the household survey. With two rings, an additional (estimated) 12 PSUs might need to be canvassed for each PSU.

It is recommended therefore that one ring of PSUs be used. However, to ascertain (a) the expected sample size in advance of the survey or (b) determine whether one or two rings is needed to build up sample size, it would be useful to first carry out a small pretest to determine the estimated average N (compare this with the discussion in area sampling of chapter 4 and recommendation 4.9). As with the area sample (chapter 4), the pretest would ascertain the number of in-scope facilities present in each PSU chosen for the pretest. It would not be necessary to canvass the ring or rings to make the estimate. Instead, only index clusters could be covered and the results multiplied by 4 (conservatively, allowing for multiple coverage of PSUs) for one ring or about 15 for two. Again, as with the area sample methodology, the pretest could be confined to 20-30 areas, purposively chosen, but representing, in equal number, both rural and urban communities.

⇨ *Example*: If the pretest found that the index cluster contains 1.3 facilities, that is, the average for the index clusters in the test, then the estimated N for the facility survey could be equal to (1.3 x 4 x number of index clusters) for a one-ring survey, or (1.3 x 15 x number of index clusters).

The number of facilities, N, thus available for the survey will be determined by the household survey index cluster and its ring or rings.

> **RECOMMENDATION 5.3**
>
> ⇨ Survey all facilities identified on a census basis (unless very large).

It is recommended that all facilities be surveyed if N is small enough to fit within the survey budget. Thus, no actual sampling of

[23] In some countries the number of facilities may be known. Generally health facilities are registered but since records may be out-of-date the available number may not be accurate.

facilities is needed. Note that it is unlikely that two rings would be the choice of design if N is very big.

The value of N may, however, be too large for total coverage in the facility survey, even under the choice of a one-ring survey design, in which case sub-sampling of the facilities would have to be applied. When sampling is necessary, it is recommended that the size of the sub-sample be determined, not by precision requirements, but by budgetary considerations and the need for domain estimates. It is recommended also that the sub-sample be done by setting up two strata - the first consisting of hospitals and the second of all others. "Size of facility" would be a better stratification variable, but that would require asking about size from the community respondents - not recommended since the accuracy of response would not be expected to be high. All facilities in the first stratum would be included in the sample and a sub-sample of those in the second stratum would be selected. For the second stratum, the sub-sample should be selected systematically with equal probability after first arranging the list of facilities in geographic order. The procedures for systematic, *epsem* sampling from a list are described in the subsection, "Sample size and stratification" of the "List frame" section, chapter 4.

Special procedures for estimation

Step 6 (in the preceding section) of the survey/sampling operation calls for the compilation of population data needed for weighting and estimation. The compilation does not require any field work *per se* but rather entails additional office work involving the original sampling frame that was used to select the index cluster of the linked household survey. Thus, it is essential that the frame materials be available for this activity.

The probability that a facility is included in the sample is a function of the probability of selecting not only the cluster that contains it but the PSUs surrounding the latter. It is important to note that this statement holds whether the cluster containing the facility is the index cluster itself or one in the ring of the index cluster. In either case it is necessary to ascertain the probabilities of all PSUs (that is, clusters) surrounding the cluster where the facility is located. Thus, when no sub-sampling of facilities within selected rings is used,[24] the probability of a given facility is equal to the sum of the probabilities of all PSUs included in its ringed space.

The specific steps of this operation are as follows:

(1) For each cluster containing at least one sample facility ascertain whether it is the index PSU of the original household sample or a PSU in the surrounding ring of the index cluster.

 (a) If the index cluster, record its first-stage selection probability as given by the household sample, if known, plus the parameters that determined it.

---

[24]When sub-sampling of facilities within the rings is used, the joint probability of both stages of sampling must be determined.

(b) If in the ring, record the parameters that are needed to calculate the probability, that is, the measure of size (number of households or number of persons) from the original household frame.

(2) For each PSU of the index cluster ringed space that contains at least one sample facility, demarcate on the frame map the ring of PSUs surrounding it.

(3) For all PSUs of step 2 not already included from step 1 record the PSU's measure of size (number of households or number of persons) from the original household frame used for the household survey.

(4) The probability of a given sample facility is then determined, partly, by summing the measures of size in its ringed space, but also by taking account of the original sampling parameters from the household survey, discussed in chapter 7.

These procedures are illustrated and described further below in Figure 5.1.

Since the probability of selection of the index cluster is a component of the probability for the facilities located within the entire ring, we must be able to retrieve either the original probability of the household sample PSU from the sample frame materials or information necessary to calculate it (step 1). Often, the PSU probability itself may not be readily available, especially if the household survey is self-weighting (in that case, the overall probability of selection, or its reciprocal - the weight - may be known, but not the first-stage selection probability). In that case, step 1 further states that the parameters needed to calculate the index cluster probability should be gotten from the household survey frame. These parameters are (1) the number of household survey sample PSUs selected in the particular stratum under consideration, (2) the measure of size for the selected PSU and (3) the total measure of size over all PSUs in the stratum.

We will want to record the same measure-of-size data that was used to select the index PSU. Generally, this is number of households, but in some countries it may be number of persons. It must be the measure of size on the original frame, rather than a more current or up-dated measure that may have been developed after the index PSU selections were made.[25] It is essential of course that the household survey frame information be available for this process. Otherwise, probabilities and weights would have to be approximated in some way.

RECOMMENDATION 5.4

⇨ Confine the number of rings in the survey design to 1.

The remaining steps of the procedure are necessary in order to calculate the weights accurately. The next step requires obtaining the same maps, or copies, that were used to

[25]In some household surveys such as DHS the index cluster measure of size may be up-dated in a field operation, especially if the sample design uses a "master" sample, while the surrounding PSUs are not up-dated. An adjustment that might be used in such a situation is to multiply the measures of size for the surrounding PSUs by the ratio of current to original measure of size of the index cluster.

demarcate the index cluster ring. They will be needed to identify the ringed spaces of any cluster containing a sample facility, whether the index cluster or not. Then, by referring again to the original frame materials used in the household survey for selecting the index PSU, one will be able to complete the remaining steps.

Note that in carrying out these procedures, any household sample PSUs which are contiguous to *each other* will share a portion of the same ringed space which will, by definition, overlap (this may be seen in Figure 5.1 - PSUs numbered 1 and 2).

The procedures may be further elucidated by Figure 5.1 and the description which follows it.

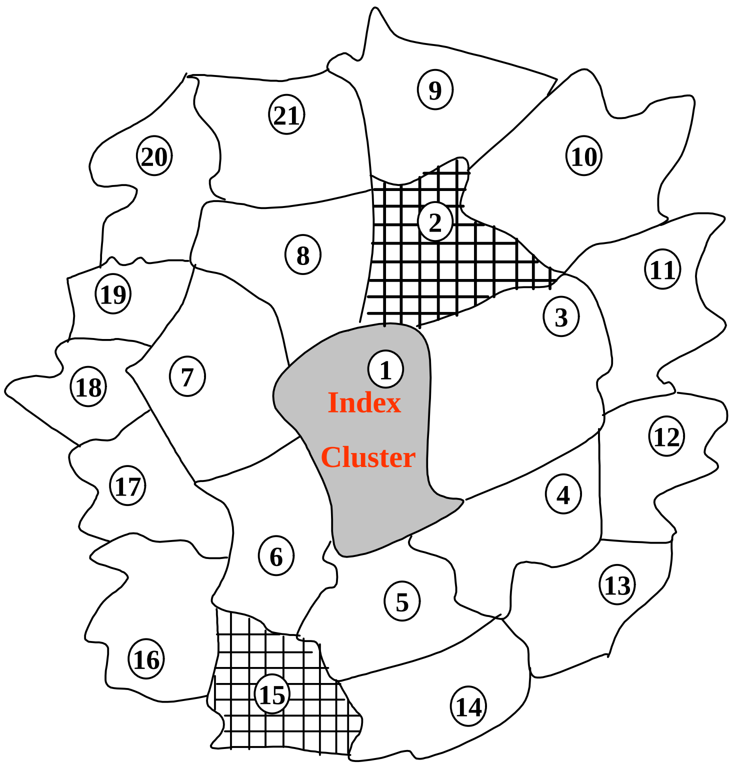

Figure 5.1 Depiction of Ringed Space around Index Cluster and its Contiguous PSUs

Visual inspection of Figure 5.1 shows the index cluster depicted in the center of the diagram with shading and indicated by the number 1. The ringed space around the index cluster consists of the PSUs numbered 2 through 8. Each of these PSUs is contiguous to the index cluster and all of them, 1-8, will be canvassed in the facility survey, under the preferred one-ring design.

Two other PSUs are cross-hatched, number 2 and number 15, also indicating that they are sample PSUs in the household survey. By further inspection of the figure, the ringed space of each of the PSUs that form the index cluster ring may also be seen easily.

⇨ *Examples*: PSU 3 has its own ringed space, which consists of PSUs 1, 2, 4 and 10-12; while the ringed space of PSU 5 consists of PSUs 1, 4, 6, 14 and 15; and so on.

The totality of the secondary ringed space is simply all PSUs around the primary space which touch at least one PSU in that space. This fact is useful to keep in mind when performing step 2 above, because to establish the secondary ringed space only requires demarcating the second ring around the index cluster, rather than demarcating, separately, the ringed space around each of the PSUs in the primary ring.

Recall that the probability of a facility is a function of the probability of the cluster containing it plus those clusters in its ring. Thus, the probability of any facility located in cluster 1 is partly determined by summing the measures of size, as determined from the household sample, of clusters 1 through 8. Observe also that the survey procedure requires canvassing all the clusters in the ringed space of the index cluster. Therefore, for example, if a facility is located and hence sampled in cluster 4, then its probability of selection is a function of the sum of the first-stage selection probabilities of clusters 1, 3-5 and 12-14, even though clusters 12-14 are not canvassed since they are not in the ringed space of the index cluster . An important aspect to recognize here is that the probabilities of selection for facilities differs according to which one of the clusters, index plus ring, it happens to be located in.

It is quite possible that the index cluster itself may not contain any facilities, as mentioned previously. In that situation of course the sample values for the index cluster would be zero, so the probability and weight for the index cluster are irrelevant. In fact, weighting must be done only for those clusters that contain sample facilities.

See chapter 7 for the mathematical procedures to calculate the probabilities and weights.

It should be noted that in applying the ringed methodology for the facility survey, there will no doubt arise situations in which the ring of an index cluster crosses stratum boundaries. This would occur when a PSU in the ring is, for example, in the urban stratum - as defined by the parent household survey frame and design - while the index cluster is in the rural stratum. More precise estimates for facilities could be obtained if such cases are identified in advance of the survey. That procedure would require examining each ringed PSU to determine if it is outside the parent stratum of the index cluster, and, if so, eliminating it from survey coverage. This would, however, further complicate field implementation and may not be feasible. Without such a "solution," however, the weights calculated for ringed spaces that cross stratum boundaries would be inexact. The overall effect may be negligible, however, if there are only a few such instances.

As mentioned, there are no field tasks associated with the above-described procedures, except in the case where the cross-boundary problem just described is addressed in advance of the survey. The office work, however, is considerable. Nevertheless, it is thought to be manageable, so long as the number of rings in the design is only one. When it is two or more, the procedures become more complex to carry out.

Reliability of estimates

The sample design of the facility sample under this methodology is highly dependent, by intent, upon the household survey. The sample size, as mentioned above, cannot be controlled exactly but is dependent on the household survey PSU and the surrounding ring, and is, in fact, pre-determined on that account. As already mentioned, stratification and/or sample allocation for urban-rural, types of facilities or other important estimation domains is largely precluded, since the facility survey team must, in effect, take whatever design parameters are generated from the household survey sample space. In addition, design effects from the clustering component cannot be altered from whatever is "given" by the household sample starting point, though these would not be expected to be serious for the facility survey.

As a result, the survey estimates for the facility sample cannot be improved beyond what is possible from the household survey parent sample, unless independent data are available for post-stratification (see further discussion below). Consequently, it would be expected that estimates of facilities by type, for example, would be somewhat less efficient under this option than those obtained under the stand-alone facility sample described in chapter 4. Urban-rural estimates and other geographical domains may also be similarly less reliable than a stand-alone survey sample. Still, the concerns about geographic domains should not be overly worrisome with regard to reliability of the estimates, because the household survey sample itself will have been designed, in all likelihood in most countries, with careful attention to appropriate stratification for urban-rural and other geographic areas. The survey estimates for the facility survey, however, are still unbiased.

RECOMMENDATION 5.5

⇨ Use auxiliary data, if available and reliable, to improve facility estimates.

If information is available from an external source on the number and or distribution of facilities by type, it can be used in a post-stratification scheme to improve the estimates. The sample of facilities, after applying the design weights associated with the probabilities of selection (and perhaps adjustment for non-response), would be further weighted by application of a factor to make the distribution of facilities by type conform to the distribution from the external source. It is presumed that the external source is the "standard," and for that reason it must be, in effect, either a universe count of facilities by type from a complete register or a recent estimate from an independent, national, probability survey. Where suitable independent data exists, it is strongly recommended that they be used. Not only will the estimates be improved but the quality of the facility sample can be evaluated by comparing the magnitudes of the post-stratified weighting factors applied to correct the sample distributions.

Special certainty stratum and un-duplication with household survey frame

RECOMMENDATION 5.6

⇨ Establish a certainty stratum of large facilities.

⇨ (Use sampling if number exceeds 100.)

One way to improve the facility estimates and make the correlation between residence and facility stronger, for linking data, is to give greater chance of inclusion to those facilities that are likely to be heavily visited irrespective of their location (especially for urban residents though it applies to rural residents as well).

It is recommended, therefore, that a special stratum of certainty facilities be established for the facility survey, and this is step 7 of the list provided at the beginning of this chapter. All large, public and private hospitals - general hospitals, health centers, etc. - should be included in the survey with certainty, or probability 1.0. This certainty stratum would be constructed irrespective of the locations of the facilities - that is, they are not to be confined to the sample PSUs and their rings generated by the household survey. The stratum requires obtaining a complete, accurate and current list of such hospitals of course (see discussion of this aspect under the section, *List sampling*, of chapter 4). The definition of a large hospital would be country-specific. In some large countries, the number of large hospitals could be too numerous to cover on a census basis, in which case the sampling procedures described in chapter 4 for the list frame should be applied. As in chapter 4, it would be recommended that sampling be used if the number of facilities exceeds 100.

Any facility identified from the household survey PSU-based sample that happens to be on the list of large hospitals would have to be unduplicated in accordance with the procedures described in chapter 4.

Staff and client samples

The sampling methodologies, including sample size considerations, that should be used for this sample design to collect information from staff and clients is identical to that discussed in chapter 4. The reader is referred to those sub-sections, but a summary is provided below.

RECOMMENDATION 5.7

⇨ Use same sampling procedures for clients and staff as in chapter 4.

Within each sample facility, the information collected in the interview includes both the substantive data to meet the requirements of the survey and some basic data on staff (a list of these by type) and client volume - to be used for sampling purposes. A systematic sample of up-to-3 staff members is selected with equal probability from the compiled list. Hence, the staff sample size is approximately 3 times larger than the facility sample size.

For each sample facility a client sample is also selected. This is effectuated by having the interview team visit the facility for one full day or more to conduct interviews, and observe client-staff interactions, with an average of 4 clients that show up for consultation on that day. The sample of clients is selected systematically with equal probability from a list, compiled by the interview team on the day of interview, of all clients that appear. The size of the client sample is a little less than 4 times the size of the facility sample, because many small facilities do not have as many as 4 clients on a daily basis.

Summary of survey and sample design for facilities

The recommended sample design for a facility survey intended to be linked with household survey data is essentially determined by the sample design for the latter. Stratification and allocation considerations are restricted to what the household survey design renders, though sample size is built up by coverage of area sample space which is considerably larger than the household survey sample PSUs. The facility survey is conducted, ideally, in *all* facilities located within the boundaries of the sample PSUs (the index clusters) used in the household survey plus the surrounding ring of clusters. Sub-sampling of facilities may be necessary in some countries. A special stratum of large facilities such as general hospitals is also included with certainty.

The sample size of the facility survey is determined by three factors: (a) the number of index clusters as determined by the household survey, (b) the number of facilities in those clusters and the surrounding concentric ring, or rings, of PSUs and (c) the number of large facilities, up to 100, on the certainty stratum list.

It is worth noting that sampling as such may not be involved in the facility sample, although it will be used to select the staff and client samples. The *survey* procedures, however, to implement the facility survey are very exacting. A key procedure that will affect the accuracy of facility estimates greatly is the methodology required to compile the population data necessary from the original household survey frame to enable weighting to be carried out correctly.

The staff and client samples, briefly summarized above, are as described in chapter 4.

Chapter 6
SAMPLING TO ESTIMATE CHANGE, TRENDS

This chapter presents
- # sampling options when estimated change is the objective
- # sample size considerations

In this chapter we discuss sampling issues related to obtaining estimates of change from one period to the next. The discussion is centered primarily on sample size and whether to use the same or different sample units. The discussion applies, more or less, whether a stand-alone facility survey is being contemplated or one linked with a population-based sample survey. Where there are implications for trend measurement that are different between stand-alone and linked survey approaches, they are pointed out.

Sample design to estimate change

Many countries that conduct facility surveys are interested in monitoring change over time. Change estimates that are of interest are of two types. The first is change in the health characteristics of clients served by facilities. The second is change in various performance measures of the facilities themselves - measures that are needed for program monitoring. The latter is of particular interest to project personnel who may be monitoring their own program interventions. In relation to program monitoring, the change statistics that are often wanted are for two points in time only - (a) prior to the beginning of a program or project and (b) after it has been implemented for a period of time, a *pre-post survey* scheme. In other instances there may be situations in which a country wishes to monitor changes repeatedly, over a period of years, to examine trends.

The need to estimate change, whether once or repeatedly, has implications for survey operations and sample design of a facility survey. Three methodologies[26] may be considered, for which there are trade-offs, as follows:

(a) use of the same sample of facilities on each occasion,

[26]The description of the methods draws heavily upon Anthony Turner's chapter on sampling that appears in the World Bank's 1999 *CWIQ Survey Manual*.

(b) use of rotating or replacement panels of facilities, or
(c) use of new, different samples each time.

Proceeding from (a) to (c), sampling error on estimates of change increases while non-sampling error decreases. Sampling error is least when the same sample facilities and clients are used on each occasion, because the correlation between observations is highest (discussed in greater detail below). By contrast, use of the same sample units introduces two negative respondent effects - more non-response and conditioned responses - which combine to increase non-sampling bias. The respondent effects with repeated use of the same sample are cumulative and increase the more the sample is used - twice, three times, and so on. It should be noted, however, that the problems non-response and conditioned responses are probably not as important for facility surveying as for household surveying.

The opposite pertains if a new sample is used each time. Sampling error to measure change is highest, but the non-sampling biases mentioned above are least - respondent conditioning is absent and the non-response rate on each occasion is that which is associated with first-time interviews.

Method (b) above is a good compromise in balancing sampling error and non-sampling bias. If part of the sample is retained on each occasion, sampling error is improved over (c) and non-sampling error is improved over (a).

For countries that plan to conduct the facility survey on only two occasions, as opposed to repeat monitoring periodically, method (a) is a plausible choice. The respondent effects in a facility survey are likely to be negligible if a sample is used only twice. Annual monitoring - three times or more - would be better served by method (b). A convenient strategy is to replace a portion of the sample on each occasion. The sampling error on estimates of change, under that scheme, would be greater than method (a) by the factor

$$\sqrt{\frac{1-kr}{1-r}} \text{ , where} \qquad [6.1]$$

r is the correlation between the same sample observations on two different occasions, and

k represents the proportion of the sample retained.

⇨ *Example*: If the correlation between observations is .5 and 50 percent of the sample is retained, the sampling error for the estimate of change between the two occasions would be greater than method (a) by the factor

$$\sqrt{\frac{1-0.25}{1-.5}}\text{, or 1.22.}$$

The replacement panel design would give sampling errors on change statistics about 22 percent greater than a design that uses the same facilities on both occasions.

It is useful also to compare method (b) with method (c). For (c), the value of r is zero because there is no overlap in the sample between occasions. Mathematically, the denominator in the expression above then disappears (it refers to method (c)), but the numerator which refers to method (b) does not. Again, if we let the correlation be 0.5 for that part of the sample which is common, we have

$$\sqrt{1-(.5)(.5)}\text{, or approximately 0.87.}$$

The replacement design would yield sampling errors about 87 percent as large as a different sample on each occasion.

Operationally, there are advantages and disadvantages in using the same sample on more than one occasion. It is cheaper to use the same sample, as sample selection need not be done after the first occasion. The exception would be to update a list frame periodically to account for new facilities and then to select an appropriate sample from the new ones. If the survey is conducted only twice and within a year or so apart, it may not be necessary to update the sample. In a rotating design, however, new samples would also have to be selected for the replacement panels.

With respect to client samples within facilities, it is probably not feasible to consider sampling the same ones periodically. This is because the procedure for client sampling entails selecting them in connection with their visits to the facilities. Trying to accommodate this feature on a subsequent survey occasion presents logistical difficulties that are insurmountable, with respect to the same sample of clients. Follow-up surveys for *longitudinal* data, however, may be considered with the same sample of clients. This could be done by recording their addresses or housing locations on the first occasion, so that visits a year or so later could be scheduled for follow-up.

Sample size to estimate change

Estimating change or trends requires a larger sample size than estimating levels, which was discussed in chapter 4. There are various approaches to figuring the appropriate sample size including one given in the aforementioned *FANta Sampling Manual* that takes direct account of the magnitude of the change.

We can also approach it from a different but equivalent perspective. When estimates from two independent (that is, option c above) samples are compared for change, each one has an error variance and they have to be added. To consider the implications for sample size, it is necessary to find out the standard error of the difference - the estimated change. When the two samples have the same sample size, it is given[27] by:

$$(SE)_d = \sqrt{[2p(1-p)][deff]/n}, \text{ where} \qquad [6.2]$$

$(SE)_d$ is the standard error of the difference between the estimate for the first occasion compared to the second,

p is the estimated proportion, assumed to be virtually the same for each year (the values will be different of course or else no measure of change would exist, but the slightly different values between the two surveys will have little effect on the standard error and can be ignored)[28],

$deff$ is the sample design effect, and

n is the sample size.

The number, 2, is necessary because each survey contributes approximately the same sampling variance. Clustering, or design effects, symbolized by *deff*, may often be ignored because they are generally very small when estimating differences, so long as the sample designs for the two surveys are similar. Thus, the sampling error to estimate change, when design effects are negligible, are roughly 41 percent, that is $\sqrt{2}$, greater than the sampling error to estimate level. Note that the formula above without the number 2 would be the standard error for level, but the design effects could not be ignored.

One significant interpretation of this result is that the sample size needs to be increased by a like amount if the estimate of annual change takes precedence over the estimate of annual level. Consequently, small changes from one period to the next cannot be measured reliably without large sample sizes.

⇨ *Example*: If the sample size needed to estimate the level were calculated to be 1000 cases

---

[27]See Scott, C. "Ghana Survey Programme: Some Options and their Sample Design Implications," Report for the World Bank, July 1996.

[28]If the change is large, say 10 percent or greater, the formula is $(SE)_d = \sqrt{[[p_1(1-p_1)]deff / n_1] + [[p_2(1-p_2)]deff / n_2]}$ where the subscripts for the p and n values denote the 2 occasions.

then 2000 cases would be the required sample size to estimate a small change, using an independent sample (option c).

The sampling error of the difference, and likewise the sample size, can be reduced, however, if the *same* sample is repeated from the first to the second survey. In that case, the standard error of the difference is multiplied by the factor, $\sqrt{1-r}$, where r is the correlation between the two sets of results, as discussed in the subsection above. The sample size would be reduced by the square of this factor, or $1-r$. For many facility survey variables, r may take on values of 0.5 or higher. If sample rotation is used, the correlation is reduced by multiplying it by the fraction of the sample that is retained (also discussed in the previous subsection).

⇨ *Examples*: Using the same sample on both occasions, the sample size necessary for the change estimate would be reduced by $1-r$. If r is 0.5, the sample size, from the preceding example, would become (2000) x (1-.5), or 1000 cases. With sample rotation of, say, 25 percent, the retained sample would be 75 percent and the sample size needed would be (2000) x [1-(.75)(.5)], or 1250 cases.

Chapter 7
WEIGHTING AND ESTIMATION

This chapter presents
- # probabilities and weights for sample designs of:
- # stand-alone facility survey
- # facility survey linked to population survey PSUs

Introduction

The probabilities and weights discussed in this chapter pertain strictly to the sample designs discussed in the manual - the sample design for a stand-alone facility survey discussed in chapter 4 and the design for one which is linked to household survey PSUs discussed in chapter 5. It is crucial to recognize that even minor deviations in sample designs from the two recommended ones could change the probabilities from those described below, and hence, the weights and weighting procedures.

The sample designs that are presented in this manual are not self-weighting, that is, different probabilities of selection are present for the sampled units, whether facilities, staff or clients. The stand-alone facility survey design of chapter 4 is selected from two frames - list and area - with quite different procedures (of necessity) and, of course, different probabilities of selection. The probabilities of selection for the population-linked design of chapter 5 are determined by (1) those of the first-stage selection units (PSUs) of the household survey and by (2) the particular cluster of the ringed space in which a facility happens to be located, and both these factors are virtually always differential. The client and staff sampling methods, which are the same for the stand-alone facility survey and the population-linked one, utilize different rates of selection depending upon the size of the facility - the technique necessary to fix an expected number of client and staff samples per facility. These differential rates, alone, depart from a self-weighting design for the client and staff samples. Their overall probabilities of selection, which must take account of the probability of selecting the PSU at the first stage, further ensures a non-self-weighting sample.

Weighting, therefore, will be necessary to produce any of the estimates needed from the facility survey - whether totals or percentage distributions. Obtaining the weighting factors necessary for estimation requires calculating the probabilities of selection. The *design* weights are the reciprocals of the probabilities. In addition, further weighting adjustments will likely be necessary to account for non-response and, possibly, to bring the sample estimates into conformance with independent facility estimates, where the latter may be available.

Probabilities of selection for stand-alone facility survey

This section provides probabilities and weights for the sample design as recommended in chapter 4.

Facility estimates

Selection of facilities is done through a one-stage sample but from two different frames - list and area. From either frame, the recommended sampling method is *epsem*, though in some countries *pps* may be used. We consider the list and area frames separately as follows:

<u>List frame</u>

Two strata have been designated for the list frame - a stratum of large facilities (stratum 1) and all other (stratum 2). The probability of selection for stratum 1 facilities is 1.0, because every facility is included with certainty. In notation, it is simply

$$P_{(L1)} = 1.0, \text{ where} \qquad [7.1]$$

$P_{(L1)}$ refers to the probability of selecting a facility from the first stratum of the list frame.

For the second stratum, the probability of selection is expressed differently for the *epsem* and *pps* designs, as follows:

$$P_{(L2)s} = \frac{n_{(L2)s}}{N_{(L2)s}}, \text{ where} \qquad [7.2]$$

$P_{(L2)s}$ is the probability of selecting a facility with *epsem* from the second stratum of the list frame in sub-stratum s (say, urban or rural),

$n_{(L2)s}$ is the number of facilities in sample (second stratum and sub-stratum), and

$N_{(L2)s}$ is the total number of such facilities in the stratum universe.

(Note that the s subscript is left out of $P_{(L1)}$ above because the probability is 1.0 over all substrata.)

In unusual instances where *pps* sampling is used:

$$P_{i(L2)s} = \left(n_{(L2)s}\right)\frac{b_{i(L2)s}}{\sum_i b_{i(L2)s}}, \text{ where} \qquad [7.3]$$

$P_{i(L2)s}$ is the probability of selecting the i^th^ facility with *pps* from the second stratum of the list frame in sub-stratum s, and

$b_{i(L2)s}$ is the measure of size (i.e., number of beds or staff members) for the i^th^ facility of stratum 2 of the list frame in sub-stratum s.

Weights for list frame sample facilities are reciprocals of the three expressions above.

<u>*Area frame*</u>

The recommended design for the area frame is to select areas with *epsem*, though *pps* may be called for in some applications. In general, all facilities located in the selected areas would be covered on a census basis. Thus, the area frame will yield a one-stage selection of facilities. The probabilities are as follows, for *epsem* and then *pps*:

$$P_{(A)s} = \frac{m_s}{M_s}, \text{ where} \qquad [7.4]$$

$P_{(A)s}$ refers to the probability of selecting a facility from the s^th^ stratum of the area frame under the *epsem* option,

m_s is the number of areas selected from the s^th^ stratum of the area frame, and

M_s is the total number of areas in the universe of the s^th^ stratum.

Under the *pps* option,

$$P_{i(A)s} = a_s\frac{b_{is}}{\sum_i b_{is}}, \text{ where} \qquad [7.5]$$

$P_{i(A)s}$ is the probability of the selecting a facility in the i^th^ area of the s^th^ stratum under *pps*,

a_s is the number of areas selected from the s^th^ stratum,

b_{is} is the measure of size (see text, chapter 5) of the i^th^ area of the s^th^ stratum.

Again, weights are the reciprocal values.

Staff estimates

The staff sample is a sub-sample of workers selected from the sample facilities. Thus, it is a two-stage sample design.

The probability of selection for a given worker is equal to the probability of selecting the facility times the probability of selecting the worker, expressed as follows:

$P_{(w)} = P_{(L,A)i}\ (r_j\ )$, where [7.6]

$P_{(w)}$ is the probability of selecting a given worker,

$P_{(L,A)i}$ is the probability of selecting the i^{th} facility from either the list or area frame (and as prior determined above), and

r_j is the rate of selection to choose worker j in facility i.

The value of r_j varies according to the size of the facility. The rates, r_j, are given in Table 4.2 (for the specific case where an average of 3 workers per facility is to be sampled). Note that these rates apply irrespective of whether the facility itself was sampled from the area frame or the list frame. Also, in very small facilities, all workers will be selected for the sample, in which case r_j is 1.0 and the overall probability for those workers reduces to $P_{(L,A)i}$.

The weight for each sampled worker is the reciprocal of the value calculated in [7.6].

Client estimates

There are two estimates of interest regarding clients. The first - percentage of clients with a given attribute - is straightforward and no different in treatment than similar estimates for facility or staff. The second estimate must be a relevant and meaningful denominator for that percentage. The recommended sample design, in this manual, assumes that the denominator is defined and estimated as the *number of daily client-visits*. The probability of selection associated with the design that will estimate that denominator, and any relevant numerator, is

$P_{(c)} = P_{(L,A)i}\ (\ r_k\ )$. [7.7]

This probability is exactly analogous to that for workers above - $P_{(w)}$. The subscripts, c and k, refer to clients, but otherwise the calculations required are identical. The rates, r_k , for clients are given in Table 4.3, for the specific case where an expected 4 clients per facility are to be sampled. Again, for small facilities, the rate is 1.0 and the overall probability for those clients reduces to $P_{(L,A)i}$, and the weight is the reciprocal of expression [7.7].

Probabilities of selection for facility survey linked to household survey PSUs

This section provides probabilities and weights for the sample design as recommended in chapter 5.

Facility estimates

Facilities are selected from two frames, as in the design of chapter 4. A certainty stratum of large facilities is selected from a list frame. The probabilities of selection for these facilities is given in expressions [7.1, 7.2 and 7.3] above, depending upon (a) whether sub-sampling is used in countries where the number of large hospitals is too numerous to include all of them on a census basis and (b) if so, whether *epsem* or *pps* sub-sampling is applied.

The second "stratum" comprises the pre-specified set of sample PSUs that were selected in the concomitant household survey (DHS or other), plus its surrounding ring of PSUs. Recall that the preferred design calls for including in the sample all facilities that are located within the sample space defined in this way, that is, index cluster plus ring. In some cases it may be necessary to sub-sample facilities, however. Under the preferred scenario, the probability of selecting a facility is thus related to the probability of selecting the household survey PSU itself, though it must be modified to account for the surrounding ring. It is useful, therefore, to look first at the probability of selecting the household survey PSU.

In most household surveys the latter is based on a *pps* selection at the first stage in which the sampled cluster is in effect a primary sampling unit (PSU). In some countries, however, the clusters may be selected through two or more stages of selection. In the one-stage design the household survey cluster probability is as follows:

$$P_{is} = \frac{a_s m_{is}}{\sum_{i}^{M_s} m_{is}} \text{, where} \qquad [7.8]$$

P_{is} refers to the probability of selecting the i[th] household survey cluster or PSU, in the s[th] stratum (urban or rural for example),

a_s is the number of clusters selected in the s[th] stratum,

m_{is} is the measure of size - usually number of households or persons in the frame - for the i[th] sample PSU of the s[th] stratum.

The above is the probability of selecting the index cluster, but it would also be the probability of selecting a facility if the facility survey were confined to the index cluster. It is necessary, however, to modify [7.8] to take account of the actual locations of the sample facilities, whether inside the index cluster or in one of its contiguous clusters. The modification must account for the measures of size for other PSUs in the ringed space of any PSU containing at least one sample facility.

It is important to note that the probability is determined by the PSUs surrounding the cluster that contains the facility rather than the PSUs surrounding the index cluster. While the latter is the sampling methodology used to identify and sample the facilities, it is the former that determines their *chance* of selection.

Thus the probability of selecting a given sample facility is equal to the probability of selecting its own ringed space as follows:

$$P_{(r)is} = \frac{(a_s)\left(\sum m_{(r)is}\right)}{\sum_{i}^{M_s} m_{is}} \qquad [7.9]$$

$P_{(r)is}$ refers to the probability of selecting the ringed space of the r^th^ cluster containing a facility, and hence any facility contained within, that is associated with the i^th^ household survey PSU in the s^th^ stratum,

a_s is the number of household survey clusters selected in the s^th^ stratum,

$\sum m_{(r)is}$ refers to the summed measures of size for all PSUs in the ringed space of the r^th^ cluster containing a facility associated with the i^th^ household survey PSU in the s^th^ stratum.

It is to be noted that to ascertain these probabilities it is necessary that the requisite information from the household survey sample clusters be made available, that is, the summed measures of size for all PSUs in a given ringed space. Since the household survey may have been conducted some time in the past, the population/household data may not be readily available for all the PSUs. In that case, $\sum m_{(r)is}$ may be estimated for the cluster or PSU containing the sample facility by multiplying the population sum of the clusters for which the data are available by the ratio of clusters with data to those without, illustrated as follows:

⇨ *Example*: Suppose sample facility y in is in ringed space 001, which contains 7 clusters, 6 of which have the measure of size data available, which is in terms of number of households. The sum of total households in the 6 clusters is 6308. Multiply the latter by 7/6 to obtain the estimate for the ringed space - (7/6 x 6308, or 7359).

The more clusters for which data are available the more accurate the estimate will be of course. In general it would be recommended to use the estimation method only when the population/household data are available for 75 percent or more of the clusters in the particular ring. Note that since each sample facility may, conceivably, be in a different ring, the 75 percent threshold must be figured for each affected ring, not overall. In rings where it is impossible to obtain the data for as much as 75 percent of the clusters, the estimation technique may be used anyway, but it would be prudent to include a cautionary note in the survey report warning users that weighting "imputations" of this kind were made.

It was mentioned previously in chapter 5 that there will no doubt be cases in which the ring of an index cluster may cross stratum boundaries. For example, a PSU in the ring may be in an urban stratum, while the index cluster is in the rural stratum. In that case, the procedures for calculating the weights are inexact and are only approximations. A solution that would yield more precise estimates would entail identifying in advance of the survey those ringed PSUs which are outside the parent stratum of the index cluster and eliminate them from coverage. This would, however, complicate the field implementation to a degree that is not likely to be feasible.

As mentioned, the discussion above for [7.9] applies when all facilities within the ringed space of an index cluster are included with certainty in the facility survey. In some countries, it may be necessary to sub-sample facilities, in which case $P_{(r)is}$ must be multiplied by an additional factor representing the sub-sample rate of selection.

Weights are the reciprocals of the probabilities.

Staff and client estimates

The probabilities of selection for staff, P_w, and client, P_c, samples are analogous to [7.6] and [7.7]. They are, respectively:

$$P_w = P_F (r_j) \text{ and} \qquad [7.10]$$

$$P_C = P_F (r_k), \text{ where} \qquad [7.11]$$

P_F is the probability of selecting the facility, as determined in [7.9] above, and the other terms are defined as before.

Weights, as usual, are the reciprocal values.

Adjustments for non-response and population controls

Non-response will occur to some extent in every facility survey that is undertaken. This will happen in varying degrees with each sampled population - facilities, staff and clients. In producing the survey estimates it may be necessary to adjust the *design* weights discussed above to account for non-response.

Non-response is likely to vary for important subgroups, such as geographic areas, types of facilities. Non-response rates should be examined carefully when the survey results are available, in order to establish categories of non-response to use for weighting adjustment. The non-response categories do not have to conform to the frames or strata that were used in sample selection. For example, in the stand-alone survey it is not necessary that the categories conform to the list and area frames. An example of 4 non-response categories that might be set up would be urban-rural by large and small facilities. Application of the adjustment in this example would require that each sampled facility be assigned the non-response adjustment factor associated with whichever of the 4 categories that the facility belongs to.

The facility sample non-response adjustment factor is given by:

$$(NI_f)_a = \frac{n_a}{I_a}, \text{ where} \qquad [7.12]$$

$(NI_f)_a$ is the non-response adjustment factor to be applied to facilities in the a^{th} non-response category,

n_a is the number of facilities in sample in the a^{th} non-response category,

I_a is the number of interviewed facilities in the a^{th} non-response category.

For the staff or client sample, non-response adjustment would be cumulative. That is, the non-response adjustment factor for the parent facility would be carried over, and multiplied by a factor for the staff sample, similarly calculated, and/or the client sample. For data processing it would be convenient to define the non-response categories for client and staff samples the same as that for facilities. In that case, the non-response adjustment for the client sample would be as follows:

$$(NI_c)_a = (NI_f)_a \left(\frac{n_{c,a}}{I_{c,a}} \right), \text{ where} \qquad [7.13]$$

$(NI_c)_a$ is the non-response adjustment factor to be applied to clients in the a^{th} non-response category,

$n_{c,a}$ is the number of clients in sample in the a^{th} non-response category,

$I_{c,a}$ is the number of interviewed clients in the a^{th} non-response category.

The staff sample non-response adjustment would be similarly computed.

Note that the non-response adjustment factors are *additional weights* and they are applied by multiplying them times the design weights - the reciprocals of the probabilities of selection described in this chapter.

As discussed in chapter 5, when information is available from an external source on the number and or distribution of facilities by type, it can be used in a post-stratification scheme to improve the estimates. After applying the design weights and non-response adjustment factors, a further factor would be applied to force the distribution of facilities by type to conform to the distribution from the external source.

Nevertheless, where suitable independent data exists, it is recommended they be used for post-stratified weighting in the sample design that relates the facilities to the same PSUs as the household survey. Most importantly, each type of facility is re-weighted so that the distribution of facilities matches that of the independent source. In this case, unlike the stand-alone survey, post-strata should reflect the sample stratification (of the household survey) as much as possible.

V. Verma[29] describes the procedure as follows:

For each type of facility compute the distribution of the number of facilities (p_i) in the sample, already weighted for the design weights and non-response adjustments. Index i refers to classes in the distribution, ideally the strata used in the selection of PSUs.

Let P_i be the same distribution form the external source.

The post-stratified weights equal $P_i/(p_i)$.

These weights would be multiplied with the design and non-response weights.

[29]See letter from Vijay Verma to Martin Vaessen, February3, 1995 (unpublished) - copy available at DHS headquarters at Macro International, Inc. in Calverton, Maryland.

Chapter 8
ESTIMATING SAMPLING ERRORS

This chapter discusses
- # sampling error estimation
- # software packages
- # Taylor linearization technique
- # jackknife technique

Need for sampling errors

It is important to obtain estimates of the sampling errors associated with the survey estimates, in order to evaluate their reliability and to conduct the analysis of the results properly. Confidence intervals around the survey estimates cannot be constructed without the accompanying standard errors, or sampling errors. The sampling errors are also needed in order to make between-group comparisons, through the use of standard *t*-tests, to establish whether statistical significance holds - that is, whether an observed difference can be claimed beyond what is attributable to its sampling error.

With computer power available today, it is possible to prepare standard error estimates for every variable (survey estimate) that is tabulated. Realistically, however, it is only necessary to identify the main estimates of concern for standard error estimation - those that comprise the important analyses under consideration. Generally, selecting 50-100 key estimates for standard error computation would suffice, whether using existing software or using a dedicated computer program.

Software

Development of an appropriate estimation method to calculate standard errors (and sample variances) can be a complicated undertaking, since it must reflect the particular sample design that is used in a given application. As a result, many software packages have been developed that use generalized variance estimation procedures. Some of them available for personal computers are:

- G CENVAR - U.S. Bureau of the Census
- G Stata - Stata Corporation, College Station, Texas
- G SUDAAN - Research Triangle Institute, North Carolina
- G WesVarPC - Westat, Inc., Rockville, Maryland.

- CLUSTERS - University of Essex
- Generalized Estimation System - Statistics Canada
- PCCARP - Iowa State University

Some of these packages are available without charge and can be downloaded from the Internet, while others are commercially sold. There is a comprehensive review[30] of them which may be accessed on the web site, www.fas.harvard.edu/~stats/survey-soft/iass.html.

Variance estimation methods

The particular sample design for the facility survey will differ in each country. The main difference will be whether the country utilizes a stand-alone facility survey methodology or one which is linked to the sample PSUs of DHS or some other household survey. In either case, the survey design will be country-specific even within those two broad types. The sample design will be a complex sample using two or more stages of selection - a minimum of two stages for client and staff samples - with stratification and clustering.

Standard textbook formulas for simple random samples are inappropriate for estimating the standard errors for the facility survey and would result in under-estimates. It is necessary to use a variance estimation technique that reflects the complex design used in the survey. It was mentioned above that software packages already exist for various generalized variance estimators. Two generalized estimators that are appropriate for consideration with the facility survey are the Taylor linearization method and the jackknife method. Either may be used to estimate variances for simple means or proportions. For more complex statistics the jackknife method should be used.

Taylor linearization method

The Taylor linearization method is recommended for a facility survey design linked to the population sample PSUs. Most applications of that design are likely to take place in conjunction, specifically, with a DHS survey, for which the Taylor linearization method is typically used and technicians are quite familiar with it.

The technique operates by organizing the first-stage sampling units (PSUs or clusters) into groups for variance estimation, termed implicit strata, with each group containing at least two clusters. Homogeneity of implicit strata is achieved by pairing adjacent clusters in the same order in which they were originally selected.

---

[30]Lepkowski J. and J. Bowles, (University of Michigan) “Sampling Error Software for Personal Computers.” The article appears in *The Survey Statistician*, No. 35, December 1996, 10-17 - the newsletter of the International Association of Survey Statisticians.

Consider a ratio, r = y/x, of two sample totals y and x, where y is the weighted sample total for variable y and x the weighted sample total for the sub-group forming the denominator of the indicator. The variance of r is given by:

$$\text{var}(r) = (se)^2 = \frac{1-f}{x^2}\sum_{h=1}^{H}\left[\frac{m_h}{m_h-1}\left(\sum_{i=1}^{m_h} z_{hi}^2 - z_h^2 / m_h\right)\right] \text{, where} \qquad [8.1]$$

(se) is the standard error,

f is the overall sampling fraction, usually ignored unless greater than .05,

x is the weighted sample total number of cases,

m_h is the number of sample index clusters in the h^{th} implicit stratum,

H is the total number of implicit strata,

z_{hi} equals $y_{hi} - rx_{hi}$, where

y_{hi} and x_{hi} are the weighted sums of the y and x variables in the ringed space of cluster i of stratum h, and

z_h equals $y_h - rx_h$..

The weights used to calculate the weighted sums and totals mentioned above are those that are described in chapter 7.

Note that formula [8.1] applies only to that part of the survey estimate that comes from the sample clusters. The certainty stratum of large facilities (see chapter 5) has no sampling error associated with it because the facilities are included with probability 1.0. In (rare) instances where the facilities from the list frame may be sub-sampled, an estimate of their variance contribution to the total variance could be approximated by assuming simple random sample formulas and adding the result to that obtained from [8.1].

Formula [8.1] may be used for each study population - facilities, clients or staff - but its application differs. For the client or staff samples the “cluster” referred to in defining the various variables becomes, instead, the facility. This pertains also for client and staff samples selected from the certainty facility stratum.

Jackknife method

The jackknife method of variance estimation may be considered for the stand-alone facility survey, although the Taylor linearization method could be applied there as well. To estimate variances using the jackknife method requires forming replications from the full sample by randomly eliminating, one at a time, one sample cluster from a stratum (or estimation domain). A pseudo-estimate is formed from the retained clusters and these are re-weighted to represent the eliminated unit. For a particular stratum containing c clusters, c replicated estimates are formed by eliminating one of these, at a time, and increasing the weight of the remaining (c - 1) clusters by a factor of c/(c - 1). The process is repeated for each cluster.

For a given stratum or domain, the estimate of the variance of a rate, r, is given by:

$$\text{var(r)} = (\text{se})^2 = \frac{1}{c(c-1)}\sum_{i=1}^{c}(r_i - r)^2 \text{, where} \qquad [8.2]$$

c is the number of clusters in the stratum or estimation domain,

r is the weighted estimate calculated from the entire sample of clusters in the stratum,

r_i is equal to cr - (c-1) $r_{(i)}$, where

$r_{(i)}$ is the re-weighted estimate calculated from the reduced sample of c-1 clusters.

For an estimate of the variance at a higher level, such as national, the process is repeated over all clusters at the higher level, but with c redefined to refer to the total number of clusters at that level.

Application of formula [8.2] pertains to the following:

1. Facility, client or staff samples selected from the area frame.
2. Client or staff samples selected from the list frame, but with “cluster” above redefined to facility.
3. Facility sample selected from the list frame, *when sampling is used*.

For countries where a facility certainty stratum from the list frame is used, there is no sampling error. Where *epsem* sampling of facilities is used, the variance estimates may also be approximated from simple random sample formulas, as an alternative to the jackknife, for the list frame sample.

Appendix 1

GLOSSARY

General Terminology

Action program: an organized set of activities undertaken by a government or private agency with the purpose of achieving a goal or social reform (e.g., improving maternal, child, and infant health; reducing unwanted fertility; increasing literacy)

Catchment area: surrounding a service delivery site that contains a high percentage of the facility's clients or potential clients

Client visit: refers to the care-receiving experience of clients who visit a facility. Often used as the unit of analysis in facility surveys, and as such, refers to visits (or sampled visits) occurring during the time the survey is being conducted.

Client volume: the average number of clients who seek care (services and information) at a service delivery point during a given interval of time (e.g., one day, one week, one year). Client volume may be expressed as the a sum of the total number of clients who seek care at a facility or it can be disaggregated by type of service (e.g., family planning, immunization, prenatal) or client (e.g., new contraceptive user, continuing contraceptive user).

Community-level variables: refer to variables, which are measured at the community or aggregate level. Community-level variables describe the social and economic context in which individuals reside and in which programs operate. Such variables are often defined for community infrastructure, including schools, health services, transportation, communication, business and industry, and public health (sanitation and water supply).

Evaluation: the application of social science research procedures to judge and improve ways in which social policies and programs are designed and conducted, from the earliest stages of defining and designing programs through their development and implementation (Rossi and Freeman, 1993).

Fixed service delivery point (FSDP): refers to static health service facilities and excludes individual service agents (ISAs) providing home-based care, health camps, or itinerant providers. FSDPs, both public and private, comprise hospitals, health centers, health posts, private clinics and sanatoriums. Physicians' private practices are excluded from this category.

Geographic market of health services: refers to the full set of health service resources in a defined geographic area that define the choices for source of care for a resident of that area seeking care.

Health facility or health care facility: a building or structural space within a building where health services are provided to clients (e.g., hospital, clinic, health center or health post)

Health information system: centrally (and possibly regionally) based system for collecting, analyzing, and managing information related to health infrastructure, health personnel, and program-based outputs. In general, health information systems are computer based, at least at the central level, and are dedicated to managing information from and for public-sector health services, though private facilities may also be included.

Health provider (**service provider**)**:** personnel or individual who provides health or family planning services, including those with medical training and qualifications, as well as non-physician and auxiliary health staff.

Impact evaluation (impact assessment): a study to examine the influence of a program or intervention on health, behavioral, or other population-level outcomes that the program is intended to influence. Impact evaluation requires plausible evidence than an observed change in outcome indicators is attributable to the program or intervention (cause and effect). Assessing impact in its strictest sense, therefore, requires a causal model, as well as analytical and statistical methods that allow for causal inference.

Indicators: variables that measure the different aspects of a given program: the inputs, processes, outputs, and outcomes (Bertrand, Magnani, and Rutenberg, 1996:29).

Individual service agent (ISA): refers to individuals who provide health services or supplies, usually operating as independent agents, and who may not be associated with a fixed service delivery point (FSDP). ISAs include allopathic physicians in private practice, trained practitioners of indigenous medical systems (e.g., Indian System of Medicine), pharmacists/pharmacies and other commercial outlets that distribute health and contraceptive supplies, traditional or untrained practitioners of health services (e.g., traditional healers, quacks), trained and untrained traditional birth attendants, depot holders, and trained field workers.

Managed care: refers to a system of health care delivery that aims to control costs by assigning set fees for services, monitoring the need for such procedures as tests and surgical operations, and stressing preventive care. Managed health care systems include health maintenance organizations (HMOs), preferred provider organizations (PPOs), networks of doctors and hospitals that adhere to given guidelines and fees in return for receiving a certain number of patients; and point of service (POS) plans, which are similar to PPOs but allow patients to go outside the network for treatment, usually at a higher cost (Source: http://www.encyclopedia.com/, accessed 5/28/2000).

Monitoring (program monitoring): the process of tracking the status of program implementation and changes in the program's performance, including the use of financial, commodities, personnel and other resources (inputs), progress in implementation and service provision (process), program output(s), and short and long-term outcomes (at both the program- and population-levels). Further, monitoring involves the systematic documentation of aspects of program performance that are indicative of whether the program is functioning as intended or according to some appropriate standard. Monitoring generally involves program performance related to program process, program outcomes, or both (Rossi, Freeman, and Lipsey, 1999).

Monitoring, global (general purpose): refers to program monitoring, which is conducted on a routine basis to track the status of program implementation, including the use of resources, implementation of service delivery and other program activities, outputs, and outcomes.

Monitoring, special purpose: refers to program monitoring of a specific dimension or activity of a larger program. Special-purpose monitoring requires defining and collecting (or tracking) performance measures that are specific to the activity, output, outcome, or population of interest (e.g., low-income population).

Performance measures or indicators: refer to variables that measure the performance of different aspects of a given program. Performance measures can be defined for inputs, processes, outputs, and outcomes.

Service Availability Module (SAM): the term used for a module developed and used as a component of the demographic and health surveys (DHS) in the 1980s. The objective of the SAM was to describe the family planning and MCH environment, and to generate data for analysing the relationship between service availability and the use and quality of family planning/health services. Community informants were used to identify the "closest" facility by type (hospital, clinic, .health center, pharmacy, private doctor) within a 30 kilometer radius of the (household) survey cluster. Interviewers then visited the identified facilities to collect data on the facility's geographic accessibility, staffing, hours of operation, service offerings, length of time providing services in the community, contraceptive method availability, family planning service hours, family planning staff, and contraceptive costs. The SAM is to be replaced by the SPA (Service Provision Assessment) currently under development.

Service delivery point (SDP): a site at which health services are provided to clients, including fixed service delivery points (FSDPs) and individual service agents (ISAs). SDPs include fixed health facilities (e.g., clinics, hospitals, health centers and posts), individual health practitioners (e.g., allopathic, non-traditional) who may provide services to clients in informal environments, such as a home or a small medical office, and other sites (e.g., pharmacies, other commercial outlets, depot holders, and fieldworkers).

Service Provision Assessment (SPA): an 11-module questionnaire designed as an instrument for linked facility surveys, and intended for administration in fixed service delivery points. The SPA modules cover various areas of health facility operations: inventory of equipment and supplies; health worker interview questionnaire; observation of client-provider interactions in the delivery of family planning (new clients), sick child, antenatal, and sexually transmitted infection (STI) care; and exit interview modules for family planning clients, STI clients, and for caretakers of sick children. The final module consists of a community questionnaire to assess the market of health services in the survey cluster.

Situation Analysis (SA): a methodology originally developed by the Population Council as a tool for the in-depth assessment of contraceptive service delivery conditions at fixed service delivery points. SA data are collected directly from program managers, service providers, and program clients. The SA methodology was the first to use direct observation of the client-provider interaction. SA can be used to evaluate a program as wells as to conduct operations research. The SA methodology was first applied in Kenya in 1989. As originally developed, the SA was designed as a stand-alone or unlinked facility survey.

Target population: refers to the population to which a health program or facility, survey, or other intervention is directed. Multiple populations may be targeted by a survey. As an example, a facility survey may target facility staff, facility clients, as well as the target population of the facilities surveyed.

World Fertility Surveys (WFS): Program sponsored by the International Statistical Institute and funded by UNFPA, USAID, France, Canada, United Kingdom, and Japan to collect information on fertility trends and contraceptive practices. Between 1972 and 1984 the WFS program conducted surveys in 41 countries.

Statistical and Sampling Terminology

Accuracy, also validity: refers to the degree of unbiasedness of a survey method or estimate and related, generally, to non-sampling aspects of survey methodology such as questionnaire design, response quality and interviewer effects

Area frame sampling: sampling of geographically defined areas

Cluster sampling: sampling of groups of facilities (or other sampling elements) in next-to-last stage of a sample

Cluster size: number of sampling units sampled per cluster

Coefficient of variation (cv): synonymous with *relative standard error* and equal to the standard error divided by the survey estimate

Confidence interval: in a survey estimate, the interval described by the estimate plus and minus the standard error (sampling error)

Deff (design effect): ratio of the sampling variance from a particular survey to the variance of a simple random sample of the same size, and thus a measure of how much more unreliable the survey sample is compared to a random sample

Design weight: the inflation factor for "blowing up" the raw survey data, reflecting the probability of selection of the sample unit equal to the reciprocal, or inverse, of the probability

Domains, also estimation domains: sub-groups (sub-national areas or population sub-classes) of the universe population for which separate estimates are wanted and, thus, for which the sample plan may have to be appropriately modified in terms of stratification and sample size

Dual frame sampling: in the preferred design of the stand-alone facility survey (chapter 4), the use of two frames simultaneously for sample selection - area frame and list frame

Epsem: equal probability selection method

Implicit stratification: a method of stratification in which the areas from the area frame are arranged in geographic order and then sampled systematically

Index cluster*:* in the preferred design of the population-linked facility survey (see chapter 5), the household survey sample PSU

Intraclass correlation: degree to which a sample element (facility, person or household) in same cluster has same characteristic compared to another selected at random in the whole population (also referred to as "degree of clustering")

List sampling*:* sampling of facilities from a list

Mean square error: refers to the concept of total survey error; mathematically - variance of the survey estimate plus the square of the bias

Measure of size*:* in surveys, the count of units on a sampling frame that is used to select, usually first-stage sample units, with probability proportionate to size

Multi-stage sampling: sampling method whereby the ultimate units are sampled through a series of stages (e.g., first-stage selection of facilities, second-stage selection of clients)

Non-response: failure to obtain a response from a sampled facility, staff member or client, due to refusal or not available

Non-sampling error: the totality of survey errors apart from sampling error, and comprising response error, non-response, interviewer error, data entry and coding errors, errors of concept and questionnaire design and wording

Probability proportionate to size, pps: sampling scheme by which units are selected in accordance with their sizes, so that large units are more likely to be selected than smaller ones

Primary sampling units, PSUs: the units selected at the first stage of a multi-stage sample, usually defined in terms of geographic areas

Population-linked facility survey*:* survey in which the sample facilities are selected from the PSUs, and usually others nearby, that were chosen for a population or household survey such as DHS

Probability sampling: sampling methodology by which every element in a defined target population has a known, non-zero chance of being selected

Relative standard error (or relative sampling error, or relative error): the standard error of an estimate divided by the estimate - also known as coefficient of variation

Relative variance, also rel-variance*:* the variance of an estimate divided by the estimate, equal to the square of the coefficient of variation

Reliability, also precision: refers to the degree of sampling error present in a survey due to the fact that a sample rather than the entire population was interviewed; determined by the sample size and sample design

Ringed space: in the preferred design of the population-linked facility survey (chapter 5), the cluster containing a sample facility plus any cluster, or PSU, contiguous to it geographically

Sampling bias: refers to a class of survey errors related to the sampling process such as faulty frames or misapplication of the sampling procedure in the field (other survey-related biases occur that are apart from the sample implementation, such as response errors)

Sampling error (also standard error): the square root of the sampling variance (see latter), used to construct the confidence interval around the survey estimate

Sampling frame: the set of materials from which a sample is selected, usually involving both lists (of areas and/or facilities) and areal units

Sampling variance: refers to random error that occurs in a sample survey because only one of all possible samples that could have been selected is actually selected; mathematically equal to the square of the standard error or sampling error

Self-weighting sample: sample plan that results when every element is selected with same overall probability

Simple random sample: sample method by which every unit in the population is given the same chance of being selected

Stand-alone facility survey: facility survey that is designed without regard to the sample PSUs that are chosen for a population or household survey

Stratified sampling: sampling method by which homogeneous categories, called strata, are constructed and an appropriate sample procedure is applied to each (e.g., urban and rural areas)

Systematic sampling: sample method involving the selection of units in sequence from a list, beginning with a random start and selecting every k^{th} unit thereafter

Standardized weight*:* relative survey weight that is used to reflect differential weights for a non-self-weighting sample, and calculated in such a way that the sum of the standard weights equals the sample size

Target population: the population which a survey is intended to cover; surveys often have more than one target population, such as the facility survey with its target populations of facilities, facility staff and clients

Appendix 2

Data Collection Forms for Sampling

This appendix provides two simple, illustrative data collection forms of information necessary to carry out the sampling operations.

Form 1. Facility Information

<table>
<tr><td>Name, address and/or location of facility ➜</td><td colspan="4"></td></tr>
<tr><td>Type of facility (hospital, health clinic, pharmacy, etc.) ➜</td><td colspan="4"></td></tr>
<tr><td rowspan="3">Estimated client volume - Indicate whether daily, weekly, monthly or annual ➜</td><td colspan="4">Average number of clients</td></tr>
<tr><td rowspan="2">Circle whether
Daily
Weekly
Monthly
Annual</td><td>Total</td><td>Maternal health</td><td>Children</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td rowspan="4">Staff size ➜</td><td colspan="4">Number of</td></tr>
<tr><td>Doctors</td><td>Other health staff</td><td colspan="2">Support staff</td></tr>
<tr><td></td><td></td><td colspan="2"></td></tr>
<tr><td colspan="4">Obtain or compile a complete list of staff, by type</td></tr>
<tr><td>For hospitals only - number of beds ➜</td><td colspan="4"></td></tr>
<tr><td rowspan="2">Days and hours of operation ➜</td><td colspan="2">Days of the week</td><td colspan="2">Hours open for clients</td></tr>
<tr><td colspan="2"></td><td colspan="2"></td></tr>
</table>

The information above would most likely be collected in the substantive interview for a facility anyway. However, it is information which is needed for sampling also. In that regard, it ought to be obtained for each facility (1) whenever a list frame is being constructed or (2) whenever facilities are being inventoried through canvassing either in an area frame or in conjunction with the ringed space of a household survey sample PSU. When information on client volume and/or staff size is not available at the time of frame construction or canvassing, it will have to be collected when the facility *sample* is interviewed.

Note that it is important to obtain a list of the staff, rather than merely a count, since a sub-sample of staff will be selected in large facilities. If a list cannot be provided by the facility, it must be constructed by the interview team.

Form 1 calls also for obtaining average client volume by type. This would be done either when the frame is being constructed - if available of course - or at the time of interviewing sample facilities. The average for the total only would suffice if that is all which can be gotten. Moreover, depending upon the measurement objective of the survey, the figure to use is the one which corresponds to the in-scope universe. As explained in the text (chapter 4), the daily volume may have to be computed, depending upon how volume data are collected. If a facility provides weekly volume, the daily volume is computed as that number divided by 5 (or divided by the number of days during the week that the facility is open to patients/clients). Similarly, an entry in terms of monthly volume would have to be converted to daily volume by dividing by the (average) number of days per month that the facility is open.

Appendix 2 *continued*

Form 2. Client Listing Sheet

Facility Information: Name, Address, Location ➜			
Number	**Name of Client**	**Sample Selection**	
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
Date of Listing ➜			
Date of Interview ➜			

Total Number of Clients	**Total Selected in Sample**	**Total Responding**

The form, as shown, only provides for a maximum of 15 clients. It may of course be designed for more than 15, or, alternatively, more than one form could be used when listing in a given facility. Clients would be listed in the order in which they show up for consultation at the facility. No other information than the client's name need be recorded on the listing sheet, as it is presumed that interviewing of sample clients will take place at once or within an hour or so after his/her arrival at the facility. Other demographic information as well as the substantive data that are needed would be collected in the substantive interview of sample clients.

It is very important to keep track, accurately, of the number of clients that come in, so that the sampling interval is applied correctly. The sampling interval will have already been applied to the listing sheet. In Form 2 above, we have illustrated application of the sampling interval, by showing the symbol, , beside the names of clients that would be selected. Refer to the text of chapter 4, where the procedure is described for choosing the random start and sample selection numbers to be entered on the blank listing sheet prior to the visit by the interviewing since the sampling interval will be known in advance.

In the example illustrated by Form 2, the random start would be 2 and the sample selection numbers would be 2, 6, 10 and 14. These numbers would be indicated on the blank listing sheet and the particular clients that fall into the sample would correspond to those whose names are listed on those lines. As explained in the text, the sample selection numbers should be extended beyond the expected number of sample cases, since the exact number of clients that will show up on a given day will vary. In the example the selection numbers might be extended to 18, 22, 26, 30, etc.

Note that in small facilities - usually those in which 5 or fewer client-visits occur per day - all the clients who show up would be sampled and interviewed. In that case, Form 2 would show the symbol, , beside the name of every client.

The information at the bottom of Form 2 will be needed for quality control and for use in adjusting for non-response in those cases where total response from the sample clients is not obtained.

REFERENCES

Bertrand, J., Magnani, R. and Rutenberg, N. (1996). "Evaluating Family Programs - with Adaptations for Reproductive Health," The EVALUATION Project, Carolina Population Center, Chapel Hill.

Casterline, J. (1987). "The Collection and Analysis of Community Data," *The World Fertility Survey - An Assessment*, J. Cleland and C. Scott (eds.), London: Oxford University Press.

Food and Nutrition Technical Assistance (1999). *FANta Sampling Guide*, Washington.

Hansen, M.., Hurwitz, W. and Madow, W. (1953). *Sample Survey Methods and Theory*. Wiley, New York.

Hansen, M.., Hurwitz, W. and Jabine, T. (1963). "The Use of Imperfect Lists for Probability Sampling at the U.S. Bureau of the Census," Invited paper, 34th Session, International Statistical Institute, Ottawa.

Kalton, G. (1883). *Introduction to Survey Sampling*. Sage, Beverly Hills.

Kish, L., (1965). *Survey Sampling*. Wiley, New York.

Macro International Inc. (1996). *Sampling Manual*. DHS-III Basic Documentation No.6 Calverton, Maryland.

Miller, R., Fisher, A., Miller, K., et al. (1997). *The Situation Analysis Approach to Assessing Family Planning and Reproductive Health Services*, New York: Population Council.

Singh, K., Tsui, A., Suchindran, C. and G. Narayana (1997). "Estimating the Population and Characteristics of Health Facilities and Client Populations Using a Linked Multi-Stage Sample Survey Design," *Survey Methodology*, 23, 2, Ottawa: Statistics Canada.

Scott, C. (1985) "Sampling for Monitoring and Evaluation," A technical supplement to *Monitoring and Evaluation of Agriculture and Rural Development Projects*, Washington: The World Bank.

UNICEF (1999) *Multiple Indicator Cluster Survey Handbook*, New York.

World Bank (1999) *CWIQ (Core Welfare Indicators Questionnaire) Survey Manual*, Washington.