Survey Methodology, December 1997
Vol. 23, No. 2, pp. 137-146
Statistics Canada

Estimating the Population and Characteristics of Health Facilities and Client Populations Using a Linked Multi-Stage Sample Survey Design

K.K. SINGH, A.O. TSUI, C.M. SUCHINDRAN and G. NARAYANA[1]

ABSTRACT

This paper demonstrates the utility of a multi-stage sample survey design that obtains a total count of health facilities and of the potential client population in an area. The design has been used for a state-level survey conducted in mid-1995 in Uttar Pradesh, India. The design involves a multi-stage, areal cluster sample, wherein the primary sampling unit is either an urban block or rural village. All health service delivery points, either self-standing facilities or distribution agents, in or formally assigned to the primary sampling unit are mapped, listed, and selected. A systematic sample of households is selected, and all resident females meeting predetermined eligibility criteria are interviewed. Sample weights for facilities and individuals are applied. For facilities, the weights are adjusted for multiplicity of secondary sampling units served by selected facilities. For individuals, the weights are adjusted for survey response levels. The survey estimate of the total number of government facilities compares well against the total published counts. Similarly the female client population estimated in the survey compares well with the total enumerated in the 1991 census.

KEY WORDS: Sample survey; Program evaluation; Health services; Developing country.

1. INTRODUCTION

The evaluation of the impact of health programs on population-level health outcomes often requires knowledge of the number and characteristics of facilities and potential clients. Such information is frequently lacking in developing countries where program record keeping and vital registration systems tend to be incomplete and poorly maintained.

To obtain current information on health status, health service use, service performance, and client needs, programs have resorted to occasional sample surveys, often designed and conducted independently and subareally (Aday 1991; Ross and McNamara 1983). Some demographic and health surveys (Macro International 1996), however, do provide a national profile of population-level health outcomes, such as fertility, child mortality, and nutritional well-being. The distinct advantage of a national population sample for planning health programs is its ability to measure the attitudes and behaviors of clients as well as non-clients. Program service statistics are limited to actual clients and may not yield the most current or accurate picture of service use.

In addition to client behaviours, it is useful to monitor the accessibility and quality of services, but this requires a separate review of service provision at health facilities or related outlets. Efforts in developing countries, like the situation analysis studies (Miller, Ndhiovu, Gachara and Fisher 1991), involve probability surveys of health facilities and can provide a national overview of program performance. However, often they are restricted to reviewing public health programs because of incomplete registration of private health providers, such as private clinics or pharmacies. The lack of complete and accurate registration of private-sector service providers prevents probability sample surveys from being used to monitor health care patterns through this sector.

Constraints on available resources to expand and improve the delivery of health care in developing, as well as developed, countries are increasing. This suggests that a more efficient use of resources available for monitoring and evaluation, particularly through surveys, is a consideration for all concerned. Innovative approaches to sample surveys should be developed to provide health planners and managers with a maximum of information at a minimum of precision loss.

We present results from a multi-stage, cluster sample survey designed to estimate the population and characteristics of health facilities and target client populations. The cluster sample for the survey, conducted in the large northern Indian state of Uttar Pradesh, is used as a basis for selecting health facilities and households, with subsequent selection of service staff from the facilities and of married women of childbearing age from the households. The survey was designed to generate independent samples of health facilities, staff, households, and client populations for the health services.

The next section of this paper will describe the survey design, its contents, and fieldwork procedures as applied in

[1] Kaushalendra K. Singh, Carolina Population Center, University of North Carolina at Chapel Hill, CB #8120 University Square, Chapel Hill, NC 27516-3997 and Department of Statistics, Faculty of Science, Banaras Hindu University, Varanasi 221005 India; Amy O. Tsui, Director, Carolina Population Center, University of North Carolina at Chapel Hill, CB #8120 University Square, Chapel Hill, NC 27516-3997 and Department of Maternal and Child Health, School of Public Health, University of North Carolina at Chapel Hill, CB #7400 Rosenau Hall, Chapel Hill, NC 27599-7400; Chirayath M. Suchindran, Carolina Population Center, University of North Carolina at Chapel Hill, CB #8120 University Square, Chapel Hill, NC 27516-3997 and Department of Biostatistics, School of Public Health, University of North Carolina at Chapel Hill, CB #7400 Rosenau Hall, Chapel Hill, NC 27599-7400; Gaade Narayana,The Futures Group International, 1050 17th Street, N.W., Suite 1000, Washington, DC 20036.

Uttar Pradesh. The following section presents the comparative results on health facilities and population, and the last section will discuss lessons learned for survey design from the Uttar Pradesh application. These lessons will be important specifically for this survey's planned replication in two years but generally informative for other countries that may adopt the linked design.

2. THE PERFORM SURVEY IN UTTAR PRADESH

The PERFORM (Project Evaluation Review For Organizational Resource Management) Survey was designed to measure benchmark indicators for a large family planning project called the Innovations in Family Planning Services (IFPS) project sited in Uttar Pradesh and co-funded by the Government of India and the U.S. Agency for International Development. Uttar Pradesh has a population of over 140 million and by itself would rank as the fifth largest developing country.

2.1 Content

Indicator estimates for IFPS are needed at three levels: (1) public and private service delivery points (SDPs), (2) service providers staffing the SDPs or facilities, and (3) client population, represented by women of reproductive age. As IFPS seeks to improve the family planning service environment, it is imperative to obtain measures of indicators at this level but in such a way as to be relatable to the women resident in those environments.

As a result, the PERFORM survey developed seven questionnaires:

1-2) An urban block and village questionnaire to inventory all potential and actual providers of health services in the sampled village or urban block;

3) A fixed service delivery point (FSDP) questionnaire to gather information on the staff, services, equipment, supplies, and education and motivation activities at sampled public and private facilities.

4) A staff questionnaire administered to all FSDP staff involved in family planning services (identified from the FSDP questionnaire) to assess their capabilities and service experiences;

5) An individual service agent (ISA) questionnaire to all individuals working outside of self-standing facilities (FSDPs) who currently or potentially can provide health planning services, such as private doctors, pharmacists, midwives, lay health workers, and retailers;

6) A household questionnaire to be administered to heads of the sampled households to enumerate household members and selected demographic and social characteristics;

7) An individual questionnaire for currently married women between the ages of 13 to 49 (identified from the household questionnaire) to collect information on knowledge of and past, current, and intended use of health services, recent pregnancy and contraceptive behaviors, and additional background characteristics.

2.2 Sampling Design

PERFORM was designed to provide estimates of facility and population characteristics at the state, regional, divisional, and district levels. The district was important since it was the focal point for introducing innovative approaches and additional IFPS inputs. At the time of the survey design, Uttar Pradesh had 14 administrative divisions; two districts were selected from each using probability proportional to size (PPS) procedures. These areal units have administrative-political boundaries and thus public administration utility. The districts were also aggregated into five regional groupings.

In each district, the total number of households to be sampled was fixed at 1,500. A sample of 1,500 households per district was determined to be sufficient to provide estimates for the main population level indicators. An overall target sample size of 1,627 ever-married women aged 13-49 was required to detect a change of 5 per cent point in contraceptive prevalence (with $\alpha = 0.05$ and $1 - \beta = 0.90$) at district level. It is expected that the number of ever-married women aged 13-49 per household would be 1.15 and therefore, by visiting a sample of 1,415 households the required number of ever-married women would be obtained. Allowing for an increase of 5 per cent to accommodate non-response and non-availability, a target sample of 1,725 ever-married women aged 13-49 from the 1,500 households was considered to be sufficient. The schematic diagram of the sample design is given in Figure 1.

The districts were further stratified into rural and urban areas. According to the Census of India, all places with a municipality, a municipal corporation, a cantonment board, a notified area committee, or all other places with a minimum of 5,000 population, with at least 75 percent of the male working population engaged in non-agricultural pursuits and a population density of at least 400 persons per square kilometer, are classified as urban areas. Urban blocks and rural villages served as the secondary sampling units (SSUs). The 1,500 households to be sampled from each district were allocated to the rural and urban areas in proportion to the size of population within the district. However, if the allocated proportion of urban population was less than 20 percent, the allocation of households in the urban area was fixed at 20 percent. This allocation was prescribed to ensure coverage of a sufficient number of health delivery points.

Households within rural areas were selected using a stratified two-stage sampling plan. The villages in the rural areas were first stratified into four strata depending on the size of the of the population as follows:

Stratum	Population size of the village
I	100 - 499
II	500 - 1,999
III	2,000 - 4,999
IV	5,000 and above.

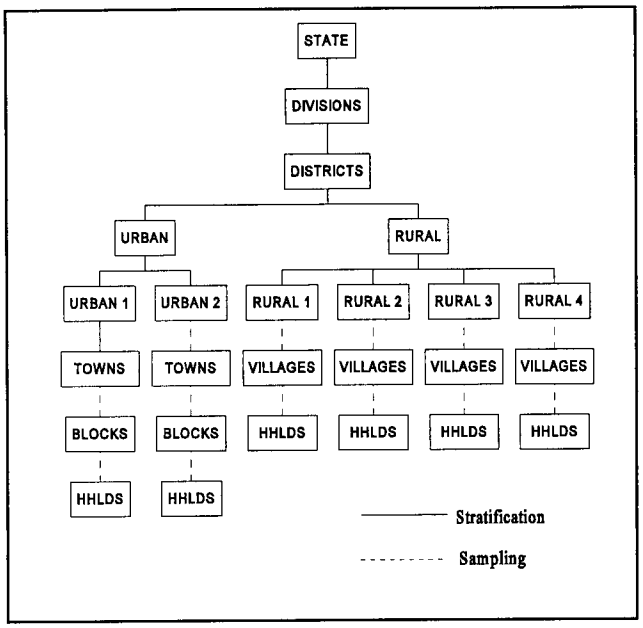

Figure 1. Schematic Diagram of PERFORM Sample Design

Villages with fewer than 100 residents or 20 households were excluded from the list (such villages were rare in the present study). The number of villages to be selected from each district was allocated proportionally to each of the four strata. Villages were selected by first arranging them within the stratum by the female literacy rates and then selecting the required number of villages by a PPS sampling procedure. All households in the selected villages were listed and mapped, and a target number of 20 households was drawn from each selected village using systematic sampling. Villages with more than 500 households or with a population size of 2,500 or more (some in stratum III and all in stratum IV) were segmented into four parts, and two segments were selected for household listing and selection. The required 20 households were selected taking ten households from each segment using systematic random sampling.

Households in urban areas were also selected using a stratified two-stage sampling plan. The towns in the urban areas of a district were stratified into two strata according to population size as follows:

Stratum	Population size of the town
I	100,000 and more
II	Fewer than 100,000.

All towns within stratum I were selected with certainty. Towns in stratum II were arranged according to population size and the required number of towns were selected by PPS. From each sampled town a minimum of two blocks were selected using PPS methods. All households in the selected blocks were listed and mapped, and 15 households were selected from each urban block using systematic random sampling.

2.2.1 District Selection Probability

Let m_k denote the population of the k-th district within a division. Because two districts must be selected from each division, the probability of selecting the k-th district from a division r_k is obtained as

$$r_k = 2 * \frac{m_k}{M}$$

where M is the total population of the division ($M = \sum_{k=1}^{t} m_k$) and t is the total number of districts in the division.

2.2.2 Village and Household Selection Probability

Let n_{ijk} denote the number of households in the i-th village, j-th stratum and k-th district. Then, p_{ijk}, the probability of selecting village i from the j-th stratum and k-th district is obtained as,

$$p_{ijk} = a_{jk} * \frac{n_{ijk}}{N_{jk}} * r_k$$

where a_{jk} and N_{jk} are, respectively, the number of villages selected and the total number of households in the j-th stratum and k-th district.

Let q_{ijk} be the probability of selecting a household from the rural areas of a selected district. Then q_{ijk} may be given as

$$q_{ijk} = p_{ijk} * \frac{20}{n_{ijk}}$$

where 20 is the number of households drawn from the selected village.

The weights for villages and households are then the inverse of their selection probabilities, *i.e.*, $1/p_{ijk}$ and $1/q_{ijk}$, and are denoted as VW_{1ijk} and HW_{1ijk} respectively.

2.2.3 Town, Urban Block and Household Selection Probability

The probability of selecting the j-th town from the k-th district, t_{jk}, is obtained as

$t_{jk} = 1$ if the population of the town is > 100,000

$t_{jk} = c_k \frac{s_{jk}}{S_k}$ if the population of the town is < 100,000

where s_{jk} is the total number of households in the j-th town (with a population < 100,000) in the k-th district, c_k is the number of towns selected in district k, and S_k is the total number of households in towns with less than 100,000 population in district k.

Let u_{ijk} denote the probability of selecting the i-th urban block from the j-th town and k-th district. Then u_{ijk} is obtained as

$$u_{ijk} = b_{jk} * \frac{x_{ijk}}{Y_{jk}} * t_{jk} * r_k$$

where b_{jk} is the number of urban blocks selected and Y_{jk} is the total number of households in the j-th town and k-th district, and x_{ijk} is the number of households in the i-th block, j-th town and k-th district.

The probability of selecting a household from the i-th urban block and the k-th district, denoted as v_{ijk}, is given as,

$$v_{ijk} = u_{ijk} * \frac{15}{x_{ijk}}$$

where 15 is the number of households drawn from the selected urban block.

The weights for urban blocks and households are then the inverse of their selection probabilities, *i.e.*, $1/u_{ijk}$ and $1/v_{ijk}$, and are denoted as UW_{1ijk} and HW_{1ijk} respectively. Since the population-level estimates are based on individuals, all individuals in a selected household received the household weight. No selection procedure was used for eligible respondents within a household.

2.2.4 Adjustment for Household Questionnaire for Non-response and Over-sampling of Urban Blocks

The adjustment of the household weight for non-response is done under the assumption of random non-response within the village (or urban block) and is carried out as follows:

Let n_1 be the number of households selected and n_2 be the number of households where interviews are completed. Then the adjusted weight for households due to non-response is defined as

$$HW_{2ijk} = HW_{1ijk} * \frac{n_1}{n_2}.$$

The final household weight also includes an adjustment of proportion of urban population in the district, where an over-sampling of urban blocks has occurred (districts with less than 20 percent of urban population).

Let n_3 be the actual proportion of urban population in a district and n_4 the proportion of urban population in the sample. Then the adjusted weight for households due to non-response and over-sampling of urban blocks is defined as

$$HW_{3ijk} = HW_{2ijk} * \frac{n_3}{n_4}.$$

2.2.5 Selection of Service Delivery Points in Sample Districts

To obtain a probability sample of service delivery points, FSDPs and ISAs were selected in relation to the SSUs, *i.e.*, the villages or urban blocks, as follows:

1) All private and public sector health institutions in selected rural and urban SSUs;
2) All sub-centres, primary health centres, community health centres, post-partum centers providing services to the population in the selected rural SSUs;
3) All private hospitals with 10 or more beds in the nearest town (with fewer than 100,000 population) within 30 kms of selected rural SSUs;
4) All municipal hospitals, district hospitals, and medical college hospitals;
5) All clinics and hospitals runs by voluntary agencies, the organized sector, and cooperatives; and
6) All ISAs in selected villages and urban blocks.

It is probably helpful first to describe the organized delivery of health care through the government sector. Residents of all villages are entitled to obtain health care from a government sub-centre (SC), a primary health centre (PHC), and a community health centre (CHC). Villages with 5,500 population or more often have an SC located within their boundaries. Approximately six SCs will report to one PHC, and PHCs in turn are linked to a CHC. At times the PHC is integrated with the CHC; as a result, our estimation must be of CHCs and PHCs combined, while SCs are estimated separately. (Population growth has led to the establishment of "additional PHCs" and redistricting of the original PHC catchment areas. These additional PHCs have been included in the estimation of the number of PHCs.) All SCs assigned to a sampled village were visited, as were their affiliated PHCs and CHCs.

At the time of listing and mapping households in each urban block and village, the FSDPs and ISAs were also listed and mapped. In addition, key informants in each SSU were interviewed regarding health outlets not visibly obvious. The selection of service delivery points – FSDPs and ISAs – within the SSU boundaries, or affiliated with the government's health subcentre, involved a full census. The one exception to this was for municipal hospitals, district hospitals and medical colleges, which were self-selected and thus had a weight of unity. The selection probabilities of the other FSDPs and ISAs are then a function of the probability of selecting the SSU, and the inverse of the latter serves as the weight of the FSDP or ISA unit. Weights for CHCs, PHCs, and SCs were calculated with the procedure below after determining some fieldwork "failure" in selecting these types of facilities correctly. (This failure is discussed later.)

Since CHCs and PHCs are associated with more than one SSU, we have assumed that one PHC exists per 30,000 population (which is approximately the actual average for Uttar Pradesh) and that one SC serves approximately 5,500 (actual district averages range from 4,000 to 6,500). Under this assumption, the CHC/PHC weight for each selected SSU is then

$$W_{\text{CHC/PHC}} = \frac{\text{Total population in selected SSU}}{30{,}000} * VW_{lijk} \text{ (or } UW_{1ijk})$$

and the SC weight for each selected SSU is

$$W_{SC} = \frac{\text{Total population in selected SSU}}{5{,}500} * VW_{1ijk} \text{ (or } UW_{1ijk}).$$

All weights for FSDPs that were not self-selected had to be adjusted for multiplicity, *i.e.*, when an FSDP was selected into the sample on the basis of more than one SSU. For example, a CHC/PHC might be selected because of two sampled SSUs. In this case, the weight for the CHC/PHC was the sum of the weights of the two selected SSU, *i.e.*, $W_{CHC/PHC}$, associated with its selection.

2.3 Survey Implementation

Fieldwork for the PERFORM Survey was conducted from June to September 1995 in Uttar Pradesh. The survey was executed by four organizations contracted following a competitive procurement process. One organization that had tested the PERFORM survey design in one district a year earlier served as the nodal or coordinating organization. Master training to survey project coordinators and supervisors was provided, including a field pretest. The actual fieldwork for PERFORM was carried out in six-member teams composed of 1 male supervisor, 1 female editor, 1 male interviewer and 4 female interviewers. Each fieldwork organization on average engaged 3 teams to cover one district, or a total of 18 field staff for data collection per district (or 21 teams for a total of 126 field staff to cover 7 districts). Overall field supervision was the responsibility of a specially–appointed four–member team, one assigned to each consulting fieldwork organization. Following field editing, the questionnaires were transported to the home offices of the survey organizations for data entry and cleaning. One type of staff person, the auxiliary nurse–midwife who is stationed at a subcentre, was difficult to reach, even after the standard three attempts.

3. RESULTS

Table 1 gives the sample coverage for the PERFORM survey, in terms of the number of units selected of each type, the number successfully interviewed, and the completion rate. The completion rates are very high for ample units requiring personal contact – ranging from 94.3 for eligible women to 96.7 percent for households. Interview completion rates were 95 percent for facilities and agents. Only for fixed facility staff was the rate somewhat lower at 90 percent, a respectable although not an outstanding level. (One type of staff person, the auxiliary nurse-midwife who is stationed at a subcentre, was difficult to reach, even after the standard three attempts.)

3.1 Population Size and Characteristics

We compare first population–level measures on selected demographic indicators obtained from other sources with those from the PERFORM survey, as shown in Table 2. The figures indicate that PERFORM results compare favorably with census measures as well as these from the recent National Family Health Survey (NFHS) conducted in Uttar Pradesh in late 1992 and early 1993, with a sample size of 11,438 ever-married women aged 13 to 49. The enumerated population shows a growth of almost 10.5 million persons since the 1991 census, and the percentage of households in urban areas is close across all three sources. The ratio of women to men is slightly lower in PERFORM (891) than in the NFHS (917). The percentage of the population in the two age groups (0 to 14 and 65 and over) compares well, as does the percentage of households belonging to the scheduled castes. The percentage of households belonging to scheduled tribes is 3.1, higher than the 1.1 observed in the NFHS. This may reflect an actual growth in such households with increased in-migration to large towns and cities by scheduled tribe members. The proportions literate show small gains since the NFHS but compare well overall. The total fertility rate and the level of modern contraceptive use also are similar and change in a consistent direction between the dates of the two Uttar Pradesh surveys. Results in Table 2 suggest that PERFORM's sample design, based on traditional multistage cluster sample designs used for demographic surveys, was executed properly to produce state-level results comparable to the census and earlier NFHS survey. The standard error and design effect of the estimates were also given in the Table

Table 1
Coverage of Sample Units of PERFORM Survey: Uttar Pradesh, 1995

Sample Coverage	Sample Units						
	Villages	Urban Blocks	Households	Eligible Women	Fixed SDPs	FSDP Staff	Individual Agents
Number Sampled	1,539	738	42,006	48,009	2,549	7,026	23,364
Number Interviewed	1,539	738	40,633	45,277	2,428	6,320	22,335
Percent completed	100.00	100.00	96.7	94.3	95.3	89.9	95.6

Notes: Villages and urban blocks served as the primary sampling units; eligibility criteria for women were currently married and between ages 13 to 49 years; SDP = service delivery point.

Table 2
Basic Demographic Indicators for Uttar Pradesh, India

Index	Uttar Pradesh				
	Census (1991)	NFHS (1992-93)	PERFORM (1995)	Standard Error	Design Effect
Population	139,112,287	*u*	149,758,641	1,542,952	–
Percent urban	19.8	22.6[a]	21.6[a]	0.6553	12.6095
Sex ratio[b]	879	917	891	34.1010	0.9727
Percentage 0-14 years old	39.1	41.8	40.2	0.1306	1.9049
Percent 65+ years old	3.8	4.8	4.7	0.0513	1.5789
Percentage scheduled	21.0	18.0[a]	20.0[a]	0.3790	3.6536
Percentage scheduled tribe	0.2	1.1[a]	3.1[a]	0.1818	4.4694
Percent Literate[c]					
Male	55.7	65.3	67.6	0.3352	6.4634
Female	25.3	31.4	37.4	0.3824	8.6821
Total	41.6	49.9	53.3	0.3352	12.2385
Total fertility rate	5.1	4.8	4.5	–	–
Modern contraceptive	*u*	18.5[d]	22.0[d]	0.3499	3.4111

u = Unavailable
[a] Based on number of households
[b] Number of females per thousand males
[c] Based on population aged 7 and above for the census and population aged 6 and above for NFHS and PERFORM
[d] Percentage of currently married women aged 15 to 49 using modern contraceptive method.

In Table 3 we compare the age and sex distributions for Uttar Pradesh obtained from the NFHS and PERFORM, as well as from the Sample Registration System, operated by the Office of the Registrar General. The sex ratios for the two surveys are also given. The age-sex distributions are again comparable across the three sources. However, there is a markedly lower sex ratio for the age group 30-49 years (820) in PERFORM and a slightly higher one for ages 50-64 (993) than those in the NFHS (941 and 960 respectively). We suspect some of this difference is due to a "push" of females out of the end of childbearing ages by field investigators of *one* survey organization to avoid completion of the pregnancy calendar and history portions of the questionnaire. (Upon further investigation, we found the sex ratios for women aged 50-64 to be uniformly higher in the seven districts under one organization's responsibility than those of others.) As a result, there are somewhat more women aged 50-64 enumerated in the PERFORM Survey than may actually be the case. This also may mean that births to women who were actually under age 50 were under-enumerated. Because this is not a high-fertility age group, the bias is not likely to be large.

3.2 Facility Size and Characteristics

By visiting and interviewing the facilities selected through the SSUs or cluster, we are able to generate an independent sample of health facilities and service providers. (These include those who currently, as well as potentially can, provide family planning services, *i.e.*, not all the estimated number of retail outlets (general merchant, kirana and pan shops) shown presently dispense contraceptives.) The weighted counts of these outlets is shown in Table 4. Our ability to validate the estimates of independent agents is weakened by the fact that many of them are not registered, particularly the "unqualified" (or quack) doctors. Narayana, Cross and Brown (1994: Table 8) report a 1991 total number of 112,568 villages in Uttar Pradesh, which would suggest almost one traditional birth attendant per village and 1 anganwandi worker for every 4.5 villages on average. These ratios appear reasonable given known circumstances regarding access to such types of care. The figures are quite close and provide evidence of the utility of the linked cluster sample design.

Table 3
Percent Distribution of the De Jure Population by Age and Sex, Based on SRS, NFHS, and PERFORM Sources for 1991-95

Age	SRS (1991)		NFHS (1992-93)			PERFORM (1995)		
	Male	Female	Male	Female	Sex Ratio	Male	Female	Sex Ratio
0-4	14.4	14.4	14.6	14.6	917	13.8	14	909
5-14	24.9	24.4	27.5	26.0	868	27.2	26.3	861
15-29	28.4	26.8	25.1	26.4	967	25.4	27.7	972
30-49	20.7	21.9	19.2	19.7	941	19.8	18.3	820
50-64	8.2	8.5	8.4	8.8	960	8.6	9.6	993
65+	3.6	4.0	5.2	4.4	718	5.2	4.1	702
Total	100.0	100.0	100.0	100.0		100.0	100.0	

Source for sample Registration System (SRS): Office of the Registrar India (1993a)
Source for NFHS: National Family Health Survey, Uttar Pradesh (1992-93)

Table 4
Total Number of Estimated Public and Private Sector Delivery Points by Type in Uttar Pradesh, India: 1995

Fixed service delivery points	Number	Individual service agents	Number
Total	31,400	Total	1,099,825
Hospitals		Physicians	
Government allopathic	968	Private resident allopathic	32,182
Government ISM	688	Private visiting allopathic	9,011
Municipal allopathic	57	Private resident (unqualified)	62,880
Municipal ISM	23	Private resident ISM	42,343
Private	5,212	Private visiting ISM	9,138
Private voluntary	130	Anganwadi workers	25,994
Private ISM	35	Village health workers	65,532
Industrial	61	Traditional birth attendants	110,546
Medical colleges	9	Medical shops	40,979
CHC/PHC/Additional PHC	3,948	General merchants	133,517
Subcentres	20,151	Kirana shops	376,679
Other	137	Pan shops	136,353
		Depot holders	5,818
		Other	48,855

3.3 Estimation Approaches

The estimated number of CHC/PHCs and SCs in Table 4 is based on the assumption that each such facility serves a fixed population size, *i.e.*, 30,000 and 5,500 respectively – the figures used by the government for planning health service delivery. The precision of the estimation would have been improved if the actual size of the local catchment population were known. In the absence of this information, we have used a constant population estimate for these two facility types.

Alternate estimation approaches were used prior to arriving at the above procedure. The first is illustrated in Table 5, which presents the actual and weighted counts of CHC/PHCs and SCs in each of the 28 survey districts. These figures are based on weighting the selected facilities by the SSU size only and without adjusting for multiplicity. The PERFORM sample selected in a total of 633 CHC/PHCs or 34.8 percent of the total (1818) and 1,267 subcenters or 13.3 percent to the total (9,491) in the 28 districts. These can be compared against the actual numbers

of CHC/PHCs and SCs in 1995 obtained from the Uttar Pradesh Department of Health and Family Welfare. It is evident that this weighting approach substantially over-estimates the number of CHC/PHCs (3,472 compared to 1,818) but yields a nearly identical number of SCs (9,495 compared to 9,491). Using the villages and urban blocks as SSUs is reasonable as they are the public administration units (and population sizes) used to determine the location of subcenters.

They, however, do not offer an adequate stratification basis for the larger health facilities. Precision is lost because we weight with the inverse of the SSU's population and when CHC/PHCs are selected in for very small SSUs, the associated weight is disproportionately inflated. This results in a higher-than-actual count of such facilities, a situation most problematic in two districts – Allahabad and Sultanpur. If these two districts are eliminated, the over-estimation is 22.5 (± 0.8) percent instead of 91 percent. (Under-estimation of CHC/PHCs results where the reverse occurs, as in Bareilly district. Because of PPS, large stratum IV villages have small weights, and in fact most selected FSDPs in this district have been sampled in the SSUs of this size.)

A second estimation approach used was to calculate the expected number of CHC/PHCs and SCs based on *a priori* knowledge that such facilities were located in SSUs of minimum size 30,000 or 5,500, respectively. With 1991 census information on the SSU population, we reconstructed the distribution of each district's population by stratum size and divided each stratum by the CHC/PHC or SC catchment size (30,000 or 5,500 respectively). This provides the expected number of CHC/PHCs and SCs for each district. We can compare this with the observed number of such facilities, obtained at the time of fieldwork where local community informants were asked whether there was a CHC/PHC and/or SC located within the SSU. This comparison is shown in Table 6, which also includes a fieldwork organization code (I to IV) in the event any pattern of survey error is evident. This approach overestimates the number of subcenters by 19.6 percent and under-estimates the number of CHC/PHCs by 26.5 percent. Excluding the two districts with a high number of stratum I SSUs (Allahabad and Sultanpur) reduces the CHC/PHC underestimation to 10.2 percent. Tabulation of estimation bias by fieldwork organization shows no systematic bias.

The results from the two weighting approaches suggest that the SSU offers an appropriate measure of size (MoS) for the selection of subcenters, since its average population size may approximate the SC's catchment size of 5,500. A larger MoS may have served the selection of CHC/PHCs better, since this facility's catchment size covers those for five to six subcenters. Because SSU size is the basis for the weight for CHC/PHCs, when the selected SSU is small, the bias in estimated counts can be large. A future design to consider is to use a cluster of SSUs that are contiguous to the selected SSU and have an MoS similar to the catchment size of CHC/PHCs. The probability of such a facility being present within the boundaries of the SSU cluster will then be higher and the weight, constructed on the basis of the total population in the SSU cluster, more reliable. In other words, our estimation is limited by not knowing how many SSUs are served by one CHC/PHC.

Table 5
Total Actual and Estimated Total Number of Community Health Centres, Primary Health Centers,[a] and Subcentres by District in Uttar Pradesh, India: 1995

District	CHC/PHC		Sub-centre	
	Actual	Estimated	Actual	Estimated
Aligarh	77	69	399	369
Azamgarh	103	69	475	949
Almora	44	104	254	468
Allahabad	112	981	594	677
Ballia	73	93	357	485
Banda	89	101	322	302
Bareilly	71	42	355	162
Dehradun	24	41	139	60
Etawah	69	84	323	364
Fatehpur	57	73	309	327
Firozabad	33	34	234	236
Gonda	107	183	528	461
Gorakhpur	59	84	470	460
Jhansi	51	77	251	157
Kanpur Nagar	12	13	81	74
Maharajgang	30	39	195	180
Meerut	76	187	410	119
Mirzapur	64	69	309	302
Moradabad	92	81	485	248
Nainital	53	79	287	344
Rampur	37	19	170	139
Saharanpur	60	49	293	388
Shahjahanpur	52	59	301	298
Sultanpur	70	487	394	649
Tehri Garhwal	31	5	159	63
Unnao	63	162	344	106
Sitapur	87	44	437	450
Varanasi	122	144	616	658
Total	1818	3472(±21)	9491	9495(±15)
Total[b]	1636	2004(±13)		

[a] Includes additional primary health centres
[b] Excludes Allahabad and Sultanpur districts
Source for 1995 actual figures from Government of Uttar Pradesh Department of Medical and Family Welfare.

Table 6
Observed and Expected Sampled Number of CHCs/PHC[a] and Subcentres Within the Rural Village (Urban Block) by District in Uttar Pradesh, India: 1995

District	CHC/PHC		Sub-Centre		Field Work Company
	Actual	Estimated	Actual	Estimated	
Aligarh	6	5	10	17	II
Azamgarh	3	5	24	15	III
Almora	5	2	14	9	I
Allahabad	19	4	17	18	III
Ballia	9	7	34	27	III
Banda	8	9	19	27	III
Bareilly	5	3	10	16	II
Dehradun	5	7	10	21	I
Etawah	8	7	17	20	II
Fatehpur	9	7	22	25	IV
Firozabad	6	6	28	30	II
Gonda	8	5	15	18	IV
Gorakhpur	5	4	16	20	IV
Jhansi	7	6	16	24	II
Kanpur Nagar	2	2	6	8	II
Maharajgang	4	4	9	13	IV
Meerut	12	8	12	34	II
Mirzapur	7	7	22	22	III
Moradabad	5	5	9	19	I
Nainital	6	4	19	19	I
Rampur	2	5	14	16	I
Saharanpur	6	6	25	21	I
Shahjahanpur	5	3	14	15	II
Sultanpur	16	6	21	15	IV
Tehri Garhwal	1	3	3	10	I
Unnao	3	6	17	17	IV
Sitapur	10	6	9	24	IV
Varanasi	6	5	18	18	III
Total	186	147	450	538	
Total[b]	151	137			

[a] Includes additional primary health centres
[b] Excludes Allahabad and Sultanpur districts.

4. DISCUSSION

The cluster-based sample design for generating independent samples of facilities and households, which can be analyzed individually or jointly, does warrant more extensive consideration in data collection efforts for health program research and evaluation in developing countries. Careful design and fieldwork sampling and execution can yield high-quality and acceptably precise survey estimates, as our results show. The weighted totals, rather than sample totals, themselves are numbers useful to program planners who decide the flow of personnel, material, and financial resources to and among various facility sites and area locations. The linkage of facility to individual records offers further important analytic opportunities to assess the relative importance of personal background and service supply factors on health outcomes of interest (*e.g.*, Boyd and Iversion 1979).

At the same time, our application of this design reveals several lessons. First there is an obvious need to monitor the survey fieldwork closely with increased on-site data entry so that the apparent "push" of eligible women out of the older age ranges can be prevented. This is difficult to detect through individual questionnaire spot checks but can

be observed in aggregate tabulations produced, say, weekly on completed questionnaires. Second, the excess count of CHCs/PHCs in two districts, where the survey fieldwork involved two *different* organizations suggests that stratum I villages might have been disproportionately selected or that some of the CHCs/PHCs reported to be within the SSU boundaries were in fact not. The former may have occurred as a sampling error since each fieldwork organization was provided with a list of sampled SSUs. Third, the listing and mapping of SSUs for facilities, individual health care providers, and households are an important stage of the fieldwork. Careful execution of this task allows the sampled units to be re-located for future follow-up. This will be an essential measurement effort for evaluating the IFPS project.

Certainly for a survey as complex as PERFORM, scaled to capture the levels of and differentials in the patterns of health service delivery and client use in an area as populous as Uttar Pradesh, the fact that the quality of the data meets most standards of precision evidences an important fieldwork achievement as well as design innovation.

ACKNOWLEDGMENTS

Partial support for this study has been provided by The EVALUATION Project, USAID Contract #DPE-3060-C-00-1054-00. The views contained herein are solely those of the authors and not the sponsoring agency. The authors acknowledge with appreciation earlier assistance on the sample design from Daniel Horowitz and T.K. Roy. We thank Lynn Moody Igoe of Carolina Population Center for editing the paper. Authors are also thankful to the anonymous referees for their useful comments and suggestions.

REFERENCES

ADAY, L.A. (1991). *Designing and Conducting Health Surveys: A Comprehensive*. San Francisco: Jossey-Bass Publishers.

BOYD, L.H., Jr., and IVERSION, G.R. (1979). *Contextual Analysis: Concepts and Statistical Techniques*. Belmont, CA: Wadsworth.

MACRO INTERNATIONAL, INC. (1996). *Demographic and Health Surveys Newsletter*, 8, 1-12.

MILLER, R.A., NDHIOVU, L., GACHARA, M.M., and FISHER, A.A. (1991). The situation analysis study of the family planning program in Kenya. *Studies in Family Planning*, 22, 131-143.

NARAYANA, G., CROSS, H.E., and BROWN, J.W. (1994). Family planning programs in Uttar Pradesh issues for strategy development: tables. Centre for Population and Development Studies, Hyderabad, India.

ROSS, J.A., and McNAMARA, R. (Eds.) (1983). *Survey Analysis for the Guidance of Family Planning Programs*. Liege, Belgium: Ordina Editions.