Large scale sample surveys of human population are among the most complex and expensive undertakings of applied statistics. Often the scope of field operations and data management is so vast that practitioners and participants lose sight of the project's goals and objectives. Subsequently, in evaluating these efforts there is a tendency to focus on apparent failures in executing initial plans precisely. Alternatively, when survey operations are compared to some ideal standard, they appear defective. Evaluations of these types tend to overlook the fact that given the realities of field conditions, ad hoc compromises must be made.

Most writers about complex surveys recognize they are a series of compromises. One way to study the impact of such compromises is to compare two (or more) surveys focusing on the compromises and seeing how alternative decisions lead to different designs. In what follows the Indian National Sample Survey (NSS) and the U.S. National Health Interview Survey (HIS) are compared. The focus is on four areas: (1) Sample selection design; (2) Field operations; (3) Analytic strategy; and, (4) Dissemination of survey results.

1.0 Background and overview

1.1 National Sample Survey

P.C. Mahalanobis, with the active support of the government, initiated the organization of the National Sample Survey (NSS) in 1950. The survey objective was to collect data to augment the estimation of national accounts, primarily in rural areas which at that time contained ninety percent of the nation's population. Urban coverage began with the Third Round of NSS, August - November 1951. Virtually complete national coverage was achieved by the Ninth Round, May - September 1955 (Murthy and Roy 1975). The survey is administered by the National Sample Survey Organization (NSSO), Department of Statistics, Ministry of Planning, Government of India (NSSO 1983b).

Today, as from its beginning, the NSS is an integrated multi-subject survey used for obtaining statistical data on socio-economic topics. This discussion focuses on the Socio-economic survey. This survey regularly covers five groups of topics: (1) Employment, rural labor, and consumer expenditure; (2) Unorganized enterprises in the non-agricultural sector; (3) Population, births, deaths, disability, morbidity, fertility, maternity and child care, and family planning; (4) Livestock and land holdings; and (5) Debt, investment, and capital formation. The program of interviewing is to cover topic areas 1. and 2. quinquennially and areas 3–5 decennially with three years of the decade reserved for special topics (NSSO 1981).

Generally the survey is conducted in independent rounds. From 1950 to 1957-58 the rounds were 3 to 9 months long. Since that time the rounds have been annual covering the agricultural year, July to June. Exceptions occurred in rounds 16, 17, 18, 27, and 28 and 36 each of which were only 6 months.

The NSS is among the world's largest household surveys. For example, the 36-th round (July - December 1981) was approximately half the standard survey round. It covered 5,409 villages and 3,652 blocks (FSU's) 81,858 rural households and 56,452 urban households (SSU's). This sample led to 449,802 rural and 296,852 urban individuals being enumerated (NSSO 1983a). The immense size of the NSS permits precise estimates at the National and State level. When supplemented by State surveys of comparable size, estimates at the regional (sub-state level) are also feasible.

India is a vast and diverse country with 14 major languages and hundreds of minor languages. For this reason interview schedules with topic coverage rather than fixed questionnaires are utilized. Interviewers are centrally trained according to a standard procedure at one of 5 zonal Training Centers located across India. They are then able to conduct interviews in the appropriate regional language or dialect.

Overall the NSS has good coverage and very high response rates. Generally 95 percent of all FSU's allocated are surveyed (NSSO 1983a Table 1.1 and 1.2) and 99 percent of sampled SSU's are interviewed.

1.2 Health Interview Survey

The United States National Health Survey is a collection of surveys established by Congress in 1956 to estimate the health status of the American population. The first of these surveys to be fielded was the U.S. National Health Interview Survey (HIS). Its objective is to describe the social, demographic, and economic aspects of illness, disability, and the use of medical service (NCHS 1972 p2). HIS is an annual survey of the civilian non-institutionalized population of the United States. The survey is designed and analyzed by the National Center for Health Statistics (NCHS), Office of the Assistant Secretary for Health, Department of Health and Human Services, United States Government. It is assisted in this work by the Bureau of the Census, Department of Commerce which is also responsible for the actual interview operations.

The data is gathered in household interviews conducted by trained interviewers using a fixed format questionnaire. The data included acute conditions, days of disability, activity limitation, and hospitalizations, as well as aspects of medical and dental care. In addition subjects are queried about their social, demographic and economic characteristics.

The personal interview is conducted with any responsible adult usually residing in the sampled...
household. Generally health status is elicited on the basis of the 'person approach' which focuses on health related actions consequent to illness and injury. Chronic conditions are grouped by specific body system and surveyed in rotation. The questionnaire has a basic fixed format of annual inquiries which is supplemented by special questionnaires used every 2 to 5 years with occasional single time questionnaires. Interviews average one hour for each sample household.

The survey is quite large. In 1983, the sampling process yielded approximately 41,000 households containing about 120,000 individuals. The response rates for the survey generally exceed 95 percent.

2.0 Sample selection design

2.1 National Sample Survey

NSS is a two stage stratified random sample of households. Separate samples are drawn from urban and rural areas. The first stage units (FSU's) are blocks and villages which are selected at random according to a stratified selection design. After listing the selected block or village, interviewers select and interview households, second stage units (SSU's). The head of household provides information for completion of the interview schedules to the interviewer.

The sampling frames for the first stage are: (1) census villages; (2) the listings of the Urban Frame Survey (UFS) blocks; and, (3) census blocks in the absence of UFS blocks. When villages are too large to list conveniently, 1200 or more households, they are sub-divided into 2 or more hamlets by the investigators (NSSO 1983b).

The decennial census provides a nearly complete list of villages in rural areas. On average, a village contains 700 households. The villages are sorted into tehsils of 200 villages each. Tehsils are grouped to form districts which sub-divide the States and Union Territories of India. The districts within larger States are grouped into 2 or more regions. There are a total of 76 regions and 384 rural districts (NSSO 1981).

The second frame, used in urban areas, is based on the Urban Frame Surveys (UFS). Beginning in 1959-60 (15-th Round) NSS conducted a special survey to form area units with more or less permanent boundaries. These have 150-200 households each. The frame has been revised every 5 years beginning in 1972. In addition, the construction of blocks for the decennial census follows the same principles as used in the UFS. Hence since 1971 the census and UFS blocks have formed the same frame.

After identification of the sample village, hamlet group or urban block the interviewer lists the households of the area following the procedures of the preceding decennial census (NSSO 1983b). When required by a particular survey, he may also collect additional household data for improving estimates by either ratio estimation or over-sampling. After listing, the interviewer selects the sample households according to a fixed probability based procedure and proceeds with interviewing.

The structure of the available frame dictates the possibilities for the selection design. Because the frame is grouped according to districts and tehsils within each state and Union territory it is possible to use these as strata. In addition for planning purposes there is great interest in district level and state level estimates. For this reason samples of at least 360 villages from each state are required in order to achieve reasonable precision (Bhattacharyya 1981). After this, the objective is to form strata and allocate the sample so as to maximize precision for fixed cost, keeping in mind the complexities of subsequent tabulations. The evolution of the design is discussed by Murthy and Roy (1975) and Bhattacharyya (1981).

Currently rural India is divided into 384 districts. Each of these form a stratum. Districts of 1.5 million or more population are subdivided according to contiguous tehsils to produce 'ultimate strata' of approximately 1 million population. An exception occurs in the states of Gujarat and Arunachal Pradesh where several districts of less than 1.5 million are also subdivided into 2 or more strata. The ultimate strata boundaries are delineated so as to be homogeneous with respect to rural population density and crop pattern (NSSO 1983b). There are 530 rural strata in the all India design (NSSO 1981 p.4). Because of inaccessibility only 378 and 379 strata were covered in rounds 33 and 34.

For the urban areas, the districts again form the basis of stratification. In general all urban areas within a district are treated as a single stratum. Within a district each city of 100,000 + population forms a sub-stratum and then 3 additional sub-strata are formed for towns of sizes (i) < 20,000, (ii) 20,000 - < 50,000, (iii) 50,000 - < 100,000. This procedure yielded 375 urban strata for the 33rd and 34th rounds. Note that some districts and states have no urban areas, hence no urban strata.

In general the sample is allocated across strata so as to be proportional to some measure of size which is appropriate for a given round. In this way, the sample is self-weighting for appropriate inquiries. This is modified so as to also balance interviewer work loads. In the 28-th through 35-th rounds random selection with probability proportional to a measure size was used. The total number of first stage units has varied from 920 to 11,245 villages and from 406 to 9,756 blocks depending on the overall objectives of the particular round. In the 35-th round there were 8008 villages and 4500 blocks. The number of second stage units is determined by the overall optimum selection design. Examples are found in Bhattacharyya (1981). The number has varied from as few as 4 SSU's to as many as 25.

Three additional points are important. First, from 1950 to 1972, all rounds were selected as two completely interpenetrating samples. That is, two sub-samples with identical designs were selected. Separate interviewers and data proces-
sing centers were used for each of the replicates. Since 1972 only the sample selection design has been in replicates (Das 1984).

The second point is that the survey is conducted in independent subrounds. These are of 2 or 3 month duration. In the latter case quarterly estimates of various characteristics are feasible.

Third, separate independently selected state samples are selected using the same design so the combined sample size is approximately double the central sample.

2.2 Health Interview Survey

HIS is also a multi-stage stratified design. It has the special feature of being designed as 52 samples so that each week of the year provides a 'representative sample' of the civilian non-institutionalized population of the United States. These samples are additive over time so that relatively rare characteristics and events may be studied.

The first stage of selection is a sample of 376 primary sampling units (PSU's). Each PSU is a county, group of contiguous counties, county equivalent, or metropolitan statistical area (MSA). The PSU's are selected according to the following stratified design.

Altogether there are approximately 1900 PSU's which exhaustively and exclusively cover the 50 states and the District of Columbia. These are divided into 376 strata, of which 112 contain one PSU and are termed self-representing (NCHS 1973a). The remaining 264 strata contain 2 or more PSU's one of which is selected with probability proportional to population size. The strata are constructed so as to be geographically compact and demographically homogeneous.

Several subsequent sampling stages are used to select clusters of households known as segments. Two types of segments have been used: compact and dispersed segments. For the former all households in the segment are eligible for interview, for the latter a sub-sample is used. Currently compact segments of four adjacent households are selected at the final stage. A discussion of past segment structures is found in NCHS (1975) and a study of some alternatives is found in NCHS (1973a). The process of segment selection yields approximately 41,000 occupied housing units containing 120,000 individuals, on average.

2.3 Comparison

Both samples are multi-stage samples of very large size. NSS is about 3 times the size of HIS, yet retains an appealingly straightforward 2-stage design. Moreover, NSS stratification is used to permit precise estimates not only at the rural urban level but also for states and districts. By contrast HIS can only provide estimates for several states (Regions).

Both samples rely heavily on census data for the respective frames. NSS uses interviewer listing and selection of households. HIS relies on updated frames and selection of households by central offices.

HIS has the facility for generating estimates for weekly periods whereas for NSS the shortest period is a quarter. Earlier versions of NSS permitted bimonthly estimates.

NSS has two special features of note. First, until 1972 the sample was selected and processed as two interpenetrating replicates. Subsequently interpenetration applies only to the selection of PSU's. Second, the States of India also interview a sample matched to the central sample. If the two samples are combined the sample size is effectively doubled and has 4 interpenetrating replicates. This has never been done.

3. Field Operations

3.1 Field Operations—NSS

The Field Operations Division of NSSO is divided into three wings: 1) Socio-economic Surveys 2) Industrial Statistics 3) Agricultural Statistics (Directorate 1969). It is responsible for data collection, preparation of data sheets, the Urban Frame Survey, improvement of crop statistics, technical assistance to states in crop estimation surveys, and collection of industrial statistics. The staffing has had two general periods, before and after the 1972 reorganization of the NSS and the formation of the NSSO.

Both before and after reorganization the primary worker in the Socio-economic Division is the Investigator (interviewer). He is a carefully trained college graduate with responsibility for all field work and submitting completed schedules according to the current round time schedule. The time allocation has been fairly constant since the inception of NSS: 10 percent travel, 20 percent listing, 30 percent camp setting and administration, 40 percent inquiry. A total of 758 investigators surveyed 8400 villages and 4632 blocks in the 23rd round.

The next level prior to 1972 was Inspectors. They were required to organize and supervise an average of 4 Investigators. They also supervised the submission of completed work. This was done by a series of detailed work programs. In addition, Inspectors were to handle queries about fieldwork by Investigators. The 23rd round was supervised by 192 Inspectors (Field Directorate 1969).

Again prior to 1972, the third level was filled by Assistant Superintendents. These were of two types: Field and Scrutiny. Assistant Superintendents (Field) had basically the same function as Inspectors but supervised the Inspectors. In addition they had responsibility for observing field operations to ensure the proper execution of field-work. An Assistant Superintendent (Scrutiny) functioned as an office based editor of schedules as they were returned from the field. There were 5 Assistant Superintendents, 3 field, 2 scrutiny, for a team of 32 Investigati-
tors and 8 Inspectors. For the 23rd round a total of 112 Assistant Superintendents were required.

After 1972 the position of Inspector was largely abolished. In addition the three wings were administratively combined. Subsequently, Assistant Superintendents are also responsible for agricultural and industrial surveys. The total staff at this level in 1976 was 1209 of whom approximately half were allocated to the Socio-economic Survey. While the number of Investigators devoted to the Socio-economic Survey has been roughly constant since 1970 at around 800, the number of immediate supervisors has doubled.

The next level of supervision is the Superintendent. In 1969 he supervised the standard team of 32 Investigators. A total of 23 were required for the 23rd round. During round 35, about half of the staff of 217 Superintendents was assigned to the NSS Socio-economic Survey, a roughly 5 fold increase.

3.2 Field Operations - HIS (NCHS 1973b)

Interviewing at the household level is conducted by a permanent staff who are employees of the U.S. Bureau of the Census, Department of Commerce. Procedures for interviewing and quality control have been developed jointly by the Bureau and NCHS.

The sample PSU's are divided into 12 interviewing regions. There is one supervisor and 8-9 interviewers per region for a total staff of approximately 100 interviewers, half of whom are working in any given week. While supervisor's offices are located in a major city of each region, much of their time is spent visiting each of the PSU's of the region to observe interviewing operations and to handle administrative work. Similarly, since interviewers cover 3-4 PSU's they also spend substantial time travelling.

A breakdown of field costs shows that while over 60 percent of the expense is for interviewing, less than half of this amount is directly attributable to interviewer salaries. The majority of the remaining expenses are for travel, training and field quality controls (NCHS 1973b).

Quality control in HIS is process rather than product oriented. This means few interviewer questionnaires are returned to the field, rather the focus is on insuring that the questionnaires are completely and accurately filled out during the initial interview.

The best method of obtaining high quality data is to employ competent interviewers on a permanent basis and the first step is selecting top quality trainees. All potential interviewers are tested for reading, arithmetic, and map reading skills (NCHS 1973b).

The training of interviewers occurs both when they are initially employed and on a continuing basis. The initial training includes a general introduction by self-study, classroom training, and field practice. During field practice and through the initial 70 household interviews, the trainee's questionnaires receive detailed scrutiny by supervisors.

The second level of training is on a continuing basis. Semi-annually interviewers are into their respective regional offices for group meetings and special training on new questionnaires. Quarterly, interviewers are paid to complete a 3 hour home study program. Finally, every quarter supervisors undertake a detailed edit of one week's questionnaires. The results of this edit are discussed with the interviewer with the aim of strengthening her weak points. The key to all of this training is the formal and informal interchange between supervisor and interviewer (1973b).

Observation of interviewing is a crucial part of the quality control process. There are three types of observation: initial, systematic, and special needs. Initial observation occurs during field practice sessions of new interviewers. Each experienced interviewer is observed every six months. This continuing observation permits strengthening of specific weaknesses; scheduled interchange between each interviewer and her supervisor; and evaluation of the overall interview quality. Special needs observation provides an opportunity for intensive training after an interviewer's work has been rejected as the result of a reinterview, the interviewer's productivity has declined below acceptable levels, or editing has shown an unacceptable error rate.

The supervisory reinterview program provides a controlled random check of the interview quality. Basically 1 to 2 percent of all households are reinterviewed by supervisors. Sample households are selected at random with probability proportional to the productivity of the interviewers. For 80 percent of the reinterviews the original interview is reconciled with the reinterview by the supervisor and respondent. For 20 percent of the reinterviews the comparison and reconciliation is undertaken by the central office as a further check. Between July 1, 1962 and June 30, 1967 a sample of 1554 reinterviews led to 115 rejected interviews. Of these 6 percent of the interviewers affected either resigned or were dismissed (NCHS 1973b).

The final step of the quality control process is editing. At the regional offices clerks are used to check all questionnaires for completeness and errors in the control items. Supervisors edit a sample of all questionnaires for consistency and adequacy for feedback to interviewers. Coding and office editing is undertaken by NCHS after questionnaires are returned from the field. In addition extensive computer edits are also utilized (NCHS 1973b).

3.3 Comparison

Field operations and the associated quality control measures of NSS and HIS are generally similar. Both surveys utilize a process rather than product orientation in quality control. Both surveys rely heavily on selection, training and supervision to insure high quality data collection. Both also back up the data collection and
supervision with extensive editing in the central office.

The most striking difference between the two surveys is in the allocation of responsibility and manpower. The NSS interviewer (Investigator) is expected to not only conduct the interviews but also to list and draw the sample from the first stage units (villages or blocks), to conduct interviews in the local language or dialect using English or Hindi schedules, to perform limited schedule tabulations, and to code and copy the completed schedule. In addition, in rural areas he may also be required to work out of campsites under difficult conditions. The standard workload (35th round) is 10 villages, 6 blocks, and one town, or 13 blocks and 2 town (NSSO 1983). This partly justifies the current practice of 1 immediate supervisor for every 2 Investigators. This is a shift from the 1 to 4 ratio utilized prior to 1972.

HIS interviewers are expected to interview and perform limited listing. They do not code or tabulate. The standard workload is 4 segments per week or about 50 segments per year (assuming half time work). As a result the ratio of supervisors to interviewers averages 1 to 8.

4. Analysis

The analysis of survey data generally has three major components: estimation of population values, estimation of precision, and cross-classification of summary statistics with appropriate narrative text. The population values of greatest interest are means, totals, and percentages. Precision is usually estimated in terms of the standard errors and biases of the corresponding estimated population values. The cross-tabulations permit comparison of various sub-populations of interest.

The elegant simplicity of the NSS selection design means that simple inflation estimators are readily available. Basically the households are inflated to the village or block and these are inflated to the appropriate stratum. Finally, stratum totals are weighted and summed to produce district, state, and all India estimates.

The use of replicates means simple estimators are available for the standard errors (Murthy and Roy 1975). These estimators reflect not only sampling error but measurement error as well (Mahalanobis 1946). The practice until 1970 was to show estimates based on each replicate and a combined estimate. Subsequently this practice has been dropped. Because of the simplicity of the selection design and the choice of estimators, conventional estimates of sampling variance could also be computed. At present no estimation of precision is undertaken.

The estimates are generally presented for all India, each state, and all districts. In addition rural-urban is a standard cross-tabulation variable. Depending on the subject matter, cross-tabulation by single variables such as age, sex and so forth are often produced. These tabulations are usually accompanied by a brief narrative text pointing out major findings.

The use of several stages requires that HIS employ more complex estimators. Usually, the estimates reflect four weights: (1) sample inflation for each stage; (2) inflation for non-response at the segment level; (3) estimates are ratio adjusted to reflect the race-residence population of the U.S.; (4) estimates are post-stratified to reflect the age-sex-race population distribution of the U.S. at the mid-point of the survey quarter.

The use of complex estimators and no replication means that conventional estimates of sampling error are not feasible. As an alternative, an indirect method known as Balanced Repeated Replication (BRR) due to McCarthy (1969, NCHS 1966, 1969) is used to estimate sampling error. This procedure is undertaken for selected variables and summary averages of the sampling errors are reported in charts as an appendix to all reports.

Since the HIS questionnaires are stable over time fairly standard tabulations by two or more demographic variables are usually reported. However, since the stratification does not take account of state boundaries the lowest political level of tabulation is for 10 or more states. As with NSS, tabulations are usually accompanied by a brief narrative text. Since the mid-seventies all statistical statements in the text must be supported with either confidence intervals small enough to be meaningful or appropriate statistical tests.

The two surveys are strikingly different in their potential and realization for analysis. Both use generally similar inflation estimators but HIS then undertakes an extensive adjustment of the estimator with the aim of improving its precision. NSS is sufficiently simple in design to permit direct estimation of sampling error. Alternatively its use of replication would permit a simple estimator of overall precision to be employed. Yet neither tactic is used and only the population estimates are reported. This practice has been strongly criticized by a number of authors including Chaudhuri (1981). In contrast, HIS estimators are sufficiently complex to make even indirect estimation of sampling error formidable task. Yet, NCHS has set this as a major priority for all publications of survey results.

5.0 Media of Dissemination

Both surveys are plagued by difficulties in producing survey results. NSS has published results less than 2 years after completing fieldwork, 38th round (July - December 1981) appeared in July - October 1983 issue of Sarvekshana. Typically, the time to publication is 4-5 years. Adhikari and Roy (1981) have estimated that with computerization of the data processing, especially data entry, the time from completion of fieldwork to publication could be standardized at 12-18 months. The program for computerization is to be implemented in 1984-85.
The publication of HIS results occurs in two phases. In the first phase, standard publications are produced annually or less frequently. The first of these is Current Estimates which typically appears within one year of completion of fieldwork. Other publications can take 2 years or more although the norm is 18 months. A major factor is the rapidity of publication is the standardized format utilized. The second phase, non-standard publications, can entail considerable delay.

The second factor affecting dissemination of results is the accessibility of the survey organization. NSS is administered by the National Sample Survey Organization (NSSO), one of three divisions in the Department of Statistics, Ministry of Planning. Moreover the Governing Council of NSSO includes representatives of two states, five ministries and two representatives of the Indian Statistical Institute. All survey publications are approved by the Governing Council. Thus ultimate consumers of the survey results are directly involved in setting NSSO policy and planning survey operations.

The administration and planning of HIS is done by the National Center for Health Statistics (NCHS). It is currently located in the Office of the Assistant Secretary for Health with no direct access to cabinet level officials. NCHS does have a policy board, US National Committee on Vital and Health Statistics. The board has no formal authority over HIS.

The third factor affecting dissemination of survey results is media of publication. Generally NSS results appear in the government sponsored journal Sarvekshana, which is distributed by NSSO. In addition, district level reports and preliminary reports are distributed to appropriate state statistics bureaus. The circulation of Sarvekshana and the reports is somewhat limited. HIS results are published in Series 10 of the government publication Vital and Health Statistics. Prior to 1980 these were available at no charge to all interested individuals. Subsequently, a charge has been applied for private subscribers. In addition, preliminary results are available in brief Advance Data reports. The journals are generally available in public and university libraries, as well as the libraries of Schools of Medicine and Public Health.

The biggest difference in publication strategies of NSSO and NCHS is the release of individual level data. NCHS through its micro-data tape program makes available the individual level data after all identifiers have been removed. This means consumers can produce specialized reports and studies with HIS data. In contrast, individual level NSS data is generally unavailable to public or private consumers without special permission of the NSS Governing Council.

References