

THE SURVEY STATISTICIAN



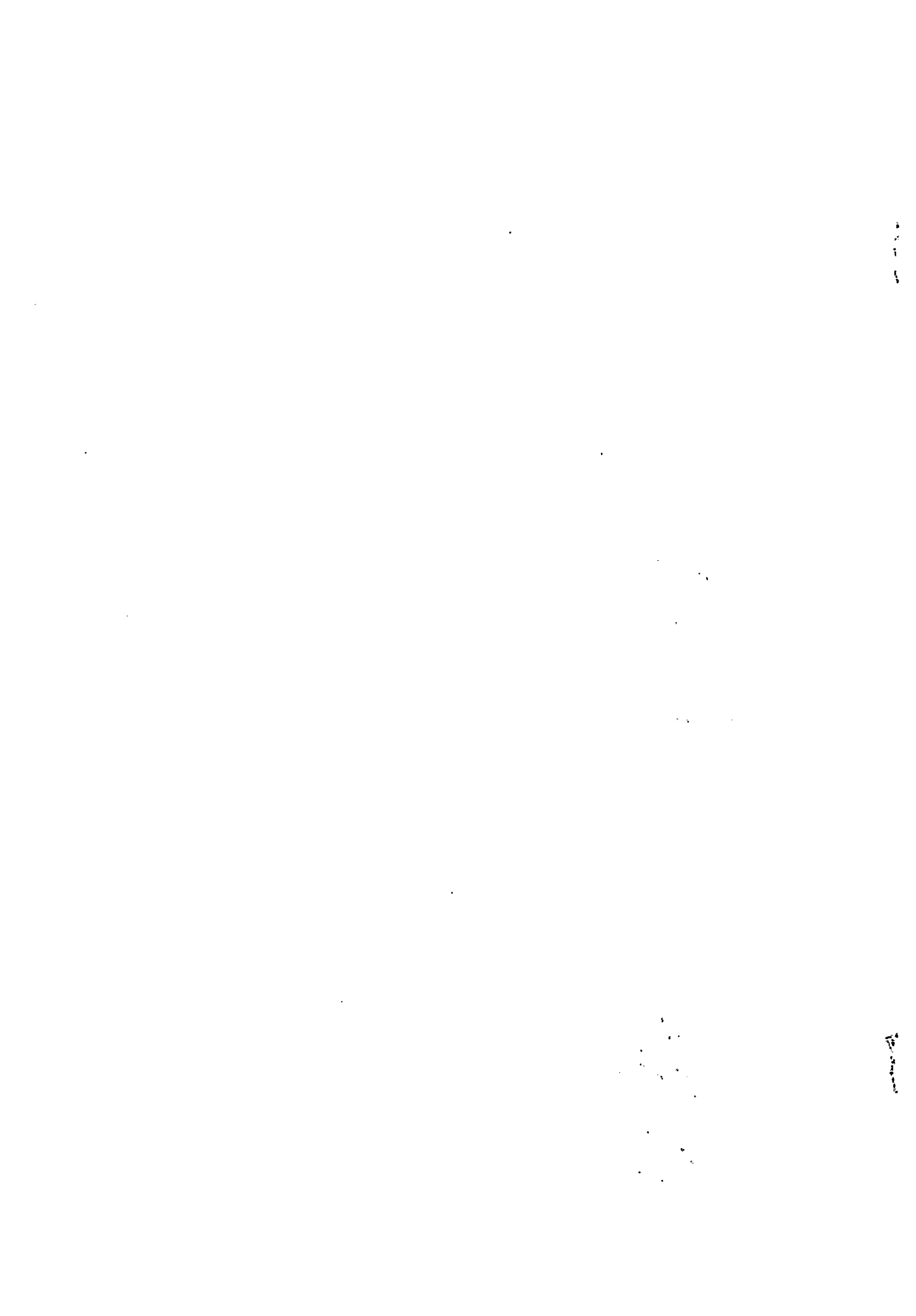


Table of Contents

Letter From the President	1
Special Articles	
"Sampling Practice in the Demographic and Health Surveys" – Thanh Lê	3
"The European Community Household Panel (ECHP)" – Anne Clemenceau	7
"InterCASIC '96 -- Conference of a Lifetime!" – William Connett	11
Question/Answer	13
Country Reports.....	16
Announcements	
IASS Invited Papers Program in Istanbul	24
Symposium '97: New Direction in Surveys and Censuses; Statistics Canada, Ottawa, Ontario, Canada; November 5-7, 1997	26
IASS/IAOS Conference Statistics for Economic and Social Development; Aguascalientes, Mexico, September 1-4, 1998	26
Statistical Terminology	27
In Other Journals	
Journal of Official Statistics – An International Review Published by Statistics Sweden	28
Survey Methodology – A Journal Published by Statistics Canada	28



Letter From the President

Dear IASS Colleague:

This will be my last letter as President. My term is quickly coming to an end, and in August I will be handing over the reins to Nanjamma Chinnapa. By the time you read this letter, the elections for the next IASS Council should be complete. I was very pleased with the slate of candidates from which we had to choose. Xavier Charoy and his committee should be congratulated for their efforts.

At the end of this article, I have summarized our achievements over the last 18 months. Although there is always more that could be achieved, I think we can feel pleased with our efforts given the IASS is a largely voluntary organization. We should not forget that our annual subscriptions are considerably lower than those of the other ISI Sections.

The big forthcoming event is the ISI Session in Istanbul in August. As well as the scientific meetings, the IASS Council will be meeting and there will be a General Assembly of IASS members. It is the best opportunity for you to influence the future directions of the IASS. If you cannot attend yourself but would still like to make a contribution, raise it with one of your colleagues who will be attending.

We are holding three short courses immediately prior to the ISI Session. At the time of writing, we are seeking nominations and there does seem to be a lot of interest so we are confident the courses will be well subscribed. We have been assisted by very generous financial assistance from the UN Statistical Office which will enable a number of participants from developing countries to attend. Unfortunately, this will not be sufficient to allow everyone who has expressed interest to attend so we have used a selection process to choose the lucky nominees.

A major activity at the ISI Session will be the establishment of the IASS program at the 1999 ISI Session. Susan Linacre will be the program chair. She would welcome your suggestions on any aspects of survey statistics with preference to topics of wide and practical interest. Susan's E-mail address is susan.linacre@abs.gov.au or her fax number is 61 6 253 1093.

Immediately following the ISI Session, there will be a satellite meeting on longitudinal studies to be held in Jerusalem from 27-31 August. It is intended to cover both the design and analysis of longitudinal surveys. The major focus will be on the special problems posed by the longitudinal nature of the survey. Gad Nathan and his team have done an excellent job organizing the conference and I would strongly recommend your attendance.

The volume of Monograph Papers from the Survey Measurement and Process Quality Conference in Bristol in 1995 has now been published by Wiley. It is a first rate reference on the topic.

There is another Section activity of which you may not be aware. The American Statistical Association is holding a Joint meeting with other associations at their August meetings. The organizers have accepted a proposal from the IASS to hold a panel discussion on "What role should methodologists have in national statistical agencies?" Don Dillman will be the keynote speaker and we have an international panel of experienced and expert speakers. It should be a great meeting.

Our other major conference activity is the joint conference with IAOS to be held on "Statistics for economic and social development" in Aguascalientes, Mexico from 1-4 September 1998. Plans are now well advanced and Geoff Hole of Statistics Canada has kindly agreed to chair the program committee. Geoff would certainly welcome any suggestions concerning the conference program. If you have any suggestions, please send them directly to Geoff. His E-mail address is holegjc@statcan.ca.

Our major recent development has been the setting up of a home page. As I have previously advised, the ISI Permanent Office will be managing the home page for us to save the expense of establishing our own operation. Fred Vogel of the US National Agricultural Statistical Service has agreed to be editor for the home page and any material should be submitted through him. The address for our home page is <http://www.cbs.nl/iass.htm>.

I have also been collaborating with a number of publishers to obtain discounts for IASS members. An arrangement has been made with Marcel

Dekker and a copy of a brochure outlining the discounts is enclosed with this Newsletter.

Finally, as this is my last letter, I would like to review developments over this period. You may recall that in my first letter, I set out my goals for the period. Our successes have been:

1. We have held a vigorous conference program in different parts of the world. (We were involved in conferences in France on privacy issue in survey taking, USA on Computer-Assisted Survey Information Collection, and will be holding conferences in Israel and Mexico as well as being part of the ISI Session program.)
2. The IASS papers from the Beijing Session have been distributed.
3. Another short course program has been developed for the 1997 ISI Session.
4. The Survey Statistician has maintained a high standard of presentation and content and has been published on time, thanks to the efforts of Mike Brick.
5. We have set up a home page.

6. We now have an attractive brochure to assist with the promotion of the IASS.
7. We have negotiated book discounts with one publisher. Unfortunately, other publishers we approached have not been so cooperative.
8. We will soon be releasing an upgraded directory with a higher proportion of E-mail addresses for members.

These are some areas where we have been less successful. Perhaps, the most important has been our failure to increase membership in those countries with relatively few or no members. Cost is a limitation and we have not found a way to overcome that problem. Membership is one of the issues I will address during the remaining months of my Presidency and it will certainly be an important challenge for Nanjamma.

Yours sincerely,

Dennis Trewin
IASS President



**Sampling Practice in the
Demographic and Health Surveys
Thanh Lê**

Introduction

The Demographic and Health Surveys Program (DHS) conducts nationally representative household surveys in Africa, Asia and Latin America. The program has several interrelated objectives: to provide demographic and health data to program managers in the participating countries, to expand the institutional capacity to conduct surveys in countries, and to maintain an archive of data files which are available to international researchers.

Since its inception in 1984, the DHS program has completed 80 standard surveys in 53 countries (a second or third survey has occurred in 19 countries). The surveys collect information from female household members on issues pertaining to maternal, child and reproductive health.

The DHS program is primarily funded by the United States Agency for International Development. Other funding institutions include UNFPA, Unicef and the World Bank. The surveys are conducted by institutions in the participating countries--frequently a national statistical office or research division within the ministry of health. Technical support is provided by a core program staff of 50 professionals (demographers, physicians, statisticians and data processing specialists) from Macro International, Inc. in Calverton, Maryland.

The DHS surveys conform to a relatively standard model although all aspects of a specific survey are adapted to the needs and conditions of a particular country. This is accomplished by modifying the DHS model questionnaires and manuals when host-country and DHS staff collaborate during technical support visits. Variation between DHS surveys is probably greatest in the area of sample design which is dictated by practical considerations such as the availability of a sampling frame and the size of the area units in the frame. Nevertheless, all DHS samples are designed using scientific probability sampling.

This article presents the main sampling features of the 80 DHS surveys completed to date. First, the operational procedures of the surveys are

described, then the sampling characteristics, and lastly some sample results.

DHS Survey Procedures

The DHS surveys are designed as single-round operations. There are at least two distinct survey instruments in all DHS surveys: a household schedule and an individual woman's questionnaire. The primary objective of the household schedule is to provide a list of household members with information on age, sex, residency status and marital status which is used for selecting eligible respondents for the individual questionnaires. Any adult household member can be the respondent for the household schedule which typically requires 10-15 minutes of interviewing time. An individual woman's questionnaire is administered to women of reproductive age in the household--usually all women aged 15 to 49. Female respondents are asked about their background characteristics, reproductive history, knowledge and use of contraception, fertility preferences, knowledge of HIV/AIDS, and questions on the health of each child under age five (in some surveys, under age three). Interviewing time is about one hour per woman but varies depending on the number of births and the number of children under age five that a woman has.

A third survey instrument, the individual man's questionnaire, has been used in 33 DHS surveys. This questionnaire is administered to men usually aged 15-59 in a subsample of households selected for the main survey, usually one-third or one-fourth. Male respondents are asked about their background characteristics, knowledge and use of contraception, fertility preferences, and knowledge of HIV/AIDS. Interviewing time is about 15-20 minutes per man.

Typically, the data collection phase of DHS surveys is preceded by a six-month preparatory period. The sampling activities conducted during this period include sample design, selection of sample areas, mapping and listing of households and selection of households. Then data collection begins and usually continues for a period of 3-4 months. In all DHS surveys, a team approach is used for data collection. Each field team consists of 4-6 interviewers, a field editor and a team supervisor. As female investigators tend to establish greater

rapport with female respondents, interviewers are almost exclusively females, except when a male questionnaire is included. It is usual for field editors to be females so that interviews in progress can be unobtrusively observed. Team supervisors are usually males.

Sample Characteristics

DHS samples follow a number of general principles. They are all national in scope and are based on probability sampling. They cover the *de facto* population and have, as main respondents, women of child-bearing age. In countries where it is culturally unacceptable to ask unmarried women questions on topics such as contraceptive use, eligibility is restricted to ever-married women.

Of the 80 DHS surveys completed to date, 37 are in Sub-Saharan Africa, 21 are in Asia/Near East/North Africa and 22 are in Latin America/Caribbean. Seventeen of the 21 surveys in Asia/Near East/North Africa cover *ever-married women* while all but one survey in Sub-Saharan Africa and Latin America cover *all women* irrespective of their marital status. Fifty-one surveys have complete coverage of the population. In the remaining 29 surveys, certain areas are excluded due to inaccessibility or disperse population. Coverage falls below 95 percent in 6 surveys and below 90 in only 3 surveys. In all 3 cases, the exclusion is due to security.

Target Sample Size/Reporting Domains

In terms of female respondents, DHS target sample sizes vary from country to country. Practical considerations such as time and budgetary constraints aside, DHS samples are designed to provide sufficient cases for the estimation of key demographic rates at the national level and for major reporting domains with an acceptable level of sampling error as well as for meaningful cross-tabulation of data. The DHS program, originally designed for samples of approximately 4,000 to 6,000 women, allows tabulation of key variables for up to 4 to 6 major reporting domains. Producing estimates for different domains with the same level of precision requires similar sample sizes across domains. Analyses of sampling errors from two earlier survey programs, the World Fertility Survey

(WFS) and the Contraceptive Prevalence Survey (CPS), indicate that a sample of 1,000 women per domain is adequate for most variables.

In DHS surveys, the two main reporting domains are urban and rural areas, a dichotomy which has substantiated marked differentials in many statistics of interest to the survey (e.g., contraceptive knowledge and use, levels of fertility and infant mortality). In most surveys, other reporting domains correspond to geographic or administrative regions which usually coincide with the units used for the purpose of planning national development and the implementation of national health programs. Larger samples are allowed in countries where the need for regional estimates is more pressing. And, as time has passed, the pressure for regional estimates in countries has increased so that DHS sample sizes have gradually increased from an average sample size of 5,800 women in DHS-I (1984-1989) to 7,700 in DHS-II (1988-1993), to 8,800 in DHS-III (1992-present). The largest DHS sample to date (Peru, 1996) covers close to 28,000 female respondents and provides estimates for 24 provinces throughout the country and 5 additional "special" population subgroups.

Sample Allocation/Sampling Domains

Sampling efficiency and simplicity favor the use of self-weighting samples. Self-weighting DHS samples are possible as long as the sample size for each reporting domain satisfies a minimum number. Nevertheless, many DHS samples deviate from self-weighting for two reasons: (i) the urban population is quite small in many of the DHS countries, and (ii) regional reporting domains typically differ greatly in size. In the first case, the urban domain is oversampled. In the second case, the sample size is redistributed so as to obtain roughly equal domains. Weights are then introduced to compensate for this disproportionate sample allocation, but generally at the domain level; within the sampling domains, the sample is self-weighting.

Of the 80 DHS surveys, only 17 are completely self-weighting, and 12 of these are in DHS-I when only urban and rural estimates are of interest. In 63 surveys, weights are used to compensate for disproportionate sample allocation at the domain level; in 13 of these, the urban area was

oversampled; in the other 50 surveys, small domains (such as regions, districts) were oversampled.

Area Sampling Frame

In practice, the availability of a suitable sampling frame is a major determinant of the feasibility of conducting a DHS survey and that issue is addressed in the earliest planning for a survey. Whenever possible, the DHS program prefers to obtain an area sampling frame from a single source. Most often, this is either an existing master sample, the sample of a previously executed survey or the list of enumeration areas (EAs) from the recently completed census. If an existing sample is available and is well-documented, with measures of size or sketch maps of area units, DHS subsamples from that sample. Use of an existing sample has the advantage of allowing comparative analysis of DHS sampling results with that of another survey. Alternatively, a list of census EAs can be a frame provided the EAs are small enough to serve as ultimate area units (UAUs) and maps showing boundaries and relevant characteristics of the EAs are available.

In cases where a nationally representative sampling frame could not be obtained from a single source, DHS has constructed special frames as a patchwork of existing frames. On occasion when it was possible to assemble and to update lists of administrative units in a reasonable time period and at reasonable costs, that was done. In total, DHS has constructed patchwork frames for 11 surveys and these are the most interesting and challenging sample designs. In Kazakhstan and Uzbekistan where rigorous health systems follow the model of the former USSR, district level health offices were visited and data for health blocks were used to construct sampling frames. In Eritrea, DHS constructed the sampling frame by combining data from different sources including voters' registration for one domain. In Port-au-Prince, the capital of Haiti, DHS selected sample points from a map using a grid procedure; for each selected sample point, DHS then listed an approximately fixed number of households around it, measured the surface area that covered the listed households, then selected the number of households required with probability proportional to its population density.

For the remaining 69 of the 80 surveys, DHS used census enumeration area frames and master samples (or pre-existing samples of on-going or recently completed surveys) in almost equal proportion.

Sample Structure

DHS samples are generally selected in two stages. The primary sampling units (PSUs) (which also served as UAUs) are selected from an area frame. A household listing operation follows in the selected PSUs, then the required number of households is selected from the household frame. In a number of cases, an intermediate area stage is necessary, (i) to further cluster the ultimate area sampling units (thus reducing travel cost), and (ii) to reduce the size of the UAUs (thus lessening work load). In most of these cases, however, this stage is not considered "effective" since only one UAU is selected from each PSU.

Stratification/Area Selection

Stratification of the area frame in the DHS is mostly explicit, with type of area (urban/rural) and location (geographic or administrative) being the stratification variables. Within each stratum, the area units are usually ordered by location, thus providing implicit stratification. Occasionally, size groups are used for implicit stratification within the explicit stratum, especially if the frame is a list of administrative units (such as villages that vary widely in size) instead of the census enumeration areas which are more homogeneous in size. In the context of the DHS, it would be desirable to stratify by more directly relevant characteristics such as level of female literacy or presence of health and family planning facilities in the areas. However, typically such data are not available in the sampling frame.

The number of UAUs which are selected in each domain depends on the number of women to be interviewed per UAU. In DHS surveys, the average "take" is about 30 women per UAU. This figure is considered optimum in terms of sampling efficiency and cost considerations, based on analytical studies carried out by the WFS. However, in urban areas a smaller "take" is usually the rule because

the cost advantages of a larger "take" are less pertinent than in rural areas.

The average number of UAUs selected per survey differs by region: 230 in Sub-Saharan Africa, 450 in Asia/Near East/North Africa, and 550 in Latin America/Caribbean. The smaller number of UAU in Sub-Saharan Africa is due to the smaller sample sizes in surveys in this region. The average sample size in Sub-Saharan Africa is 5,700 women compared to 9,500 in Asia/Near East/North Africa and 8,800 in Latin America. In Latin America where the average number of UAUs is the largest, a high proportion of sample points are in urban areas where the average "take" is small.

Mapping and Household Listing

It is DHS policy to conduct a mapping and household listing operation in all selected sample areas two or three months prior to data collection. This operation is costly and time consuming; however, it is considered to be worthwhile because a complete list of households allows greater control over sample size, and accurate household information facilitates data collection. Segmentation of large sample areas into more manageable segments followed by selection of one segment for coverage is sometimes implemented in the field prior to listing households.

A mapping and household listing operation was implemented in all but 10 DHS surveys. In these 10 surveys, DHS used a recent list of households which was constructed for the survey from which DHS subsampled.

Household Selection

Once household listing is completed, the household lists are sent to the central survey office for household selection. Centralization of household selection has the major advantage that the completeness of the listing operation can be assessed by experienced survey staff. Discrepancies between the "expected" and the "listed" number of households can be evaluated and corrective action taken, when necessary.

Some Sample Results

Sample Shortfalls

Sample shortfall is defined as the percent difference between the target and achieved samples of female respondents. A sample shortfall of 5 percent or more is observed in one-third of the DHS surveys and of 10 percent or more in 20 percent of the surveys. However, it is worth noting that a sample shortfall of 10 percent or more was particularly a problem in DHS-I (12 out of 28 surveys) and is less evident in DHS-II (2 out of 22 surveys) and in DHS-III (2 out of 30 surveys).

The shortfall in DHS-I samples could be due to (i) sample implementation errors, especially errors in the mapping and household listing operation and (ii) conscious omission of eligible women by the interviewers, for example by pushing them out of the age limits. Other problems that should not be disregarded are inaccuracy of parameters used for sample design (e.g., the estimated number of eligible women per household and the assumed non-response rate). After the experience of DHS-I, more intensive training and supervision are devoted to the mapping and household listing. Also, the procedure for selecting eligible women has changed; namely, the interviewers are instructed to interview all respondents between the ages of 15 and 49, regardless of whether or not they slept in the household the night before the interviewer's visit (in DHS-I and the earlier DHS-II surveys, only women who slept the night before the visit were eligible for interview). By requiring interviewing of all women, the incentive for misrecording residency status during the night prior to the survey is eliminated. However, the *de facto* character of DHS surveys is maintained at the data analysis stage.

Response Rates

In general, response rates are quite high in DHS surveys, although somewhat higher for surveys in Sub-Saharan Africa and Asia/Near East/North Africa than for surveys in Latin America. The average overall response rates (products of household response rates and women's response rates) are 94 percent for Sub-Saharan Africa, 95 percent for Asia/Near East/North Africa, and 85 percent for Latin America. Response rates for the

male surveys are generally lower than for the female surveys: 87 percent for Sub-Saharan Africa, 76 percent for Asia/Near East/North Africa and 73 percent for Latin America. There are no differences in response rates for the different phases of DHS.

Sampling Errors

For every survey, DHS calculates sampling errors for a large number of variables (background characteristics, reproductive history, family planning practice, current fertility intentions, reproductive health, child health, infant and child mortality, and nutritional status of children) for the total sample, urban-rural domains, and geographical or administrative domains. Sampling errors have decreased with the increase of sample size. For example, in Indonesia, the sampling error for the infant mortality rate has decreased from 7 percent of the estimate (1987, 11844 respondents), to 5 percent (1991, 22909 respondents), to 4 percent (1994, 28168 respondents). The overall design effect averaged over variables and countries is around 1.5, which means that, due to multi-stage cluster sampling, variance is increased by a factor of 2.25 (deft²) over that of an equivalent simple random sample.

DHS data can be downloaded via www.macrint.com/dhs/. DHS basic documentation can be obtained by writing to Macro International/DHS, fax (301) 572-0999.

For more information, contact Thanh Lê at Macro International/DHS, 11785 Beltsville Drive, Calverton, Maryland 20705, or by E-mail at le@macrint.com.

References

- Macro International Inc. 1996. *Sampling Manual*. DHS-III Basic Documentation No.6. Calverton, Maryland.
- Verma, V., and T. Lê. 1996. An analysis of sampling errors for the demographic and health survey. *International Statistical Review*, 64, 3, 265-294.
- Verma V. 1980. *Sampling for National Fertility Surveys*, Proceedings, World Fertility

Conference. The Hague: International Statistical Institute.

~~~~~

### The European Community Household Panel (ECHP) Anne Clemenceau

#### 1. The ECHP: A 14-nation Community Survey

The European Community Household Panel (ECHP) is a standardized survey conducted in Member States of the European Union under auspices of the Statistical Office of the European Communities (Eurostat). The survey involves annual interviewing of a representative panel of households and individuals in each country, covering a wide range of topics. It was launched in response to the increasing demand in the European Union of *comparable information* across the Member States on income, work and employment, poverty and social exclusion, and many other diverse social indicators concerning living conditions of private households and persons.

A major aim of the survey is to provide up-to-date and comparable data on *personal income*. Information on incomes of households and persons is indispensable for policy-makers at the national and European levels, particularly in monitoring changes arising from the establishment of the Single Market. The survey provides detailed information at the individual and household levels on a variety of income sources: wage incomes, rent subsidies, unemployment and sickness benefits, social assistance benefits, occupational and private pensions, and so on. Hence the strength of the ECHP is its focus on the income received from all sources by every member of the household.

ECHP forms the most closely coordinated component of the EU system of social surveys of households and persons. Indeed, from its very conception and in design and implementation, it is an *EU-wide undertaking*. Its specified characteristics and arrangements derived from its *European perspective*.

Following a two-wave pilot during 1993 in the then 12 Member States, the full-scale survey began in 1994 and is planned for a total duration of six years.

The second and third rounds were conducted during 1995 and 1996 respectively, following-up the original sample in each country. Two of the new Member States, namely Austria and Finland, subsequently joined the survey - Sweden remaining the only exception for the present.

Given that data collection takes place in 12 different countries - 14 from 1996 - a degree of flexibility has been allowed so as to permit each country to adapt common procedures to its national situation.

The organizational structure of the ECHP is composed of the three following parts:

1. **National Data Collection Units (NDUs)** are responsible for selecting the national sample and for adapting the questionnaire to the national standards, for carrying out the fieldwork, basic data processing and editing at the national level;
2. Micro-data are then transmitted by NDUs to a **central technical team and consultants at Eurostat**, who implement and manage the EU micro-database and produce publications; in addition, Eurostat tasks include the definition of a standardized design and common technical and implementation procedures, with centralized support and coordination of the national surveys;
3. In order to ensure harmonized technical implementation, the associated **Working Group**, including representatives from EFTA countries, is scheduled to meet yearly.

## 2. Outline of the Design

Three characteristics make the ECHP a unique source of information. These are (i) its coverage of a wide range of topics simultaneously; (ii) standardized methodology and procedures yielding comparable information across countries; and (iii) a longitudinal or panel design in which information on the same set of households and persons is gathered to study changes over time at the micro level.

## Multi-dimensional Coverage

In each country, the survey begins with a nationally representative sample of a few thousand households, interviewing around 60,000 households and 130,000 adults in the first wave in 1994. Within each sample household, a listing of its members along with their demographic particulars, relationship, activity, etc. is obtained. This is followed by a detailed household interview. This covers information on migration status of the household, tenure of accommodation, housing amenities and costs, possession of durable goods, major sources of income, and above all, diverse indicators of the household's financial situation. Finally, all household members aged 16 and over are subject to a detailed personal interview. Two major areas covered in considerable detail in this interview concern the economic activity and personal income of the individuals concerned. In addition a wide range of other topics are covered, such as individuals' social relations and responsibilities, health, pensions and insurance, degree of satisfaction with various aspects of work and life, education and training, and biographic information. Hence compared to other social surveys in the EU, the ECHP has a much broader and integrative character: it aims to provide comparable and inter-related information on, for instance, earnings *and* social protection benefits *and* employment *and* working conditions *and* housing *and* family structures *and* social relations *and* attitudes. Information on some of these topics may be less detailed or less precise than that in national and/or single-topic sources, but in ECHP it forms a part of a single micro-data source on the basis of which inter-relationships between different fields and the relevance of specific factors for the individuals' living conditions can be analyzed.

## Cross-national Comparability

Furthermore, these inter-relationships can be studied and compared across countries. Comparability is achieved through a standardized design and common technical and implementation procedures, with centralized support and coordination of the national surveys by Eurostat. The ECHP design has a number of features to enhance cross-national comparability:

- Common survey structure and procedures: annual interviewing of a representative panel using specified followup rules etc.
- Common standards, and where possible common arrangements as well for data processing and statistical analysis, including editing, variable construction, weighting, data adjustment, variance computation etc.
- Common sampling requirements and standards (concerning sample size, probability selection procedures, respondent and call-back rules etc.) coupled with flexibility in the actual designs to suit national conditions.
- Common frameworks for analysis through a collaborative network of researchers.
- *A central feature of the project is the use of a common 'blue-print' questionnaire which is to serve as the point of departure for all national surveys.*

### **Longitudinal or Panel Design**

The truly unique feature of ECHP is of course its panel design. Within each country, the original sample of persons is followed over time at annual intervals. In addition to providing longitudinal data, ECHP is also designed to provide representative cross-sectional pictures over time by constant renewal of the sample through appropriate followup rules. Persons who move or otherwise form or join new households are followed up at their new location all over the EU. Children in the original sample become eligible for the detailed personal interview as they achieve the age of 16, and children born to sample women are automatically included as a part of the survey population. In this manner the sample reflects demographic changes in the population and continues to remain representative of the population over time, except for losses due to sample attrition and households formed purely of new immigrants into the population. Furthermore, at any time the detailed survey covers all persons living with any of the original sample persons in the same household. This is so that the sample persons can be studied in the context of their total household.

In each wave, the sample data are edited, weighted and imputed as required to obtain a representative picture of the population. Similarly, across waves, the data are linked, edited, imputed and weighted as required to construct a micro (household and personal) level database. The database is standardized across the countries, and is archived and disseminated by Eurostat to facilitate comparative analysis.

### **3. Sample Size and Response Rates**

The sample size achieved in countries which took part in waves 1 and 2 can be seen in Table 1 below. A sample of some 61,000 households (approximately 127,000 adults) was achieved in the EU 12 in wave 1.

The wave 2 sample falls short of wave 1 sample due to noncontact, nonresponse, and household ceasing to exist, but is compensated to some extent by including new (split-off) households coming into the sample as a result of the movement of sample persons. Overall, wave 2 sample is as large as 95 percent of the wave 1 sample. In most cases, this ratio is 90-100 percent, with one exception (UK 79%), as a consequence of the followup procedures used there. The importance of sustaining good response rates in a panel survey such as the ECHP cannot be overemphasized.

#### **Wave Response Rates**

The overall wave 1 response rate exceeded 70 percent, though there was a large variation among the countries - from a high of 85-90 percent in a number of countries including Greece, Italy and Portugal, to a low of under 50 percent in Germany and Luxembourg. In the Netherlands and Belgium, the apparent response rates are high, but this is because 'ECHP Wave 1' for these countries corresponded in fact to later waves of the national panels incorporated into the ECHP. The original wave 1 response rates for these two countries were also low, around 50 percent. Wave 2 response rate averages 88 percent, varying across countries from 82 percent in Ireland to nearly 94 percent in Luxembourg.

Table 1. Sample sizes per country

| Country     | Achieved wave 1 sample size (households) | Achieved wave 1 sample size (individuals) | Achieved wave 2 sample size (households) | Achieved wave 2 sample size (individuals) |
|-------------|------------------------------------------|-------------------------------------------|------------------------------------------|-------------------------------------------|
| Belgium     | 4,189                                    | 8,118                                     | 3,988                                    |                                           |
| Denmark     | 3,482                                    | 5,903                                     | 3,223                                    | 5,510                                     |
| Germany     | 5,054                                    |                                           | 4,755                                    |                                           |
| Spain       | 7,206                                    | 17,908                                    | 6,522                                    | 16,276                                    |
| France      | 7,344                                    | 14,333                                    | 6,723                                    | 12,476                                    |
| Greece      | 5,523                                    | 12,492                                    | 5,218                                    | 7,874                                     |
| Ireland     | 4,048                                    | 9,904                                     | 3,587                                    | 8,094                                     |
| Italy       | 7,115                                    | 17,729                                    | 7,128                                    |                                           |
| Luxembourg  | 1,011                                    | 2,046                                     | 962                                      | 1,968                                     |
| Netherlands | 5,187                                    | 9,407                                     | 5,110                                    |                                           |
| Portugal    | 4,881                                    | 11,622                                    | 4,916                                    | 11,345                                    |
| UK          | 5,779                                    | 10,517                                    | 4,548                                    | 8,391                                     |
| Austria     | -                                        | -                                         | 3,943                                    | 7,441                                     |
| Total       | 60,819                                   |                                           | 60,623 (56,680 without Austria)          |                                           |

#### Longitudinal Attrition

The attrition rate is defined to measure the degree of success in interviewing the same set of units over time, providing longitudinal data. Longitudinal exploitation of the data requires keeping attrition rates low. The wave 1 to wave 2 rates vary from the low 4-7 percent in Portugal, Italy, Luxembourg and Germany to 10-12 percent in Belgium, Denmark and Spain, but to as high as almost 25 percent in the United Kingdom.

#### 4. Using the ECHP Data

##### Through Tables

It is possible for external users to receive tabulations extracted from the ECHP database. They will have to be requested in writing according to the specifications of the Variable List (PAN 15 and supplementary documentation). Appropriate charges will be levied for the production of new tables except for users within the institutions of the EU.

#### Coming to the Eurostat to use the Micro-data

Alternatively, bona fide researchers can come to Eurostat at a charge, to access the micro-database directly, in the secure area.

#### Receiving the Micro-data

Micro-data files for all countries except Germany will be made available free of charge to National Data Collection Units (NDUs) except Germany. Researchers contracted by Eurostat will generally be charged for this transmission. It is not yet possible for the other categories of users to receive the ECHP micro-data files.

#### 5. Further Information

A number of publications have been produced by Eurostat providing further information on ECHP procedures, progress and results. This includes a periodic ECHP Newsletter, short publications giving pertinent results called 'Statistics in Focus', and regular Eurostat publications on methodology and results. In addition, technical details have been

documented in a large number of 'PAN' papers available from Eurostat on request. Of course, NDUs and researchers also publish their own results. For an up-to-date and comprehensive review of ECHP design and procedure, see Eurostat (1996), "The European Community Household Panel (ECHP): Survey methodology and implementation" - Volume 1, Theme 3, Series E, OPOCE, Luxembourg and Eurostat (1996) "The European Community Household Panel (ECHP): Survey questionnaires waves 1-3" - Volume 1, Theme 3, Series E, OPOCE, Luxembourg.

For further information please contact A. Clemenceau, FAX: (352) 4301 34415, E-mail: anne.clemenceau@eurostat.cec.be.



**InterCASIC '96 -- Conference of a Lifetime!**  
**William Connett**

InterCASIC '96, the International Conference on Computer-Assisted Survey Information Collection (CASIC) convened on December 11 for four days in lovely San Antonio, Texas.

The conference was developed under the scientific sponsorship of the American Association for Public Opinion Research (AAPOR), the International Association of Survey Statisticians (IASS) and the Survey Research Methods Section of the American Statistical Association (ASA). The conference was organized by an international committee of experts in automated social science data collection, led by Dr. Mick Couper of the University of Michigan's Institute for Social Research.

The conference was held at a watershed time in the development of CASIC methods. It gave researchers and practitioners the opportunity to look back over past accomplishments in computer assisted data collection, and to look forward to the challenges and opportunities that lie ahead for survey research. Judging from the comments of attendees, the conference certainly met these goals.

The conference began with two well-attended half-day workshops. The first was presented by Dr. Paul Lavrakas of Ohio State University on "Total Survey Error Considerations in CASIC." The second workshop, titled "A Practical Approach to

Usability Testing," was presented by Dr. Ginny Redish of Redish and Associates, Inc.

The keynote speaker at the conference was Dr. Ben Shneiderman, Professor of Computer Science at the University of Maryland, College Park. Shneiderman is an internationally recognized expert on the field of human computer interaction research, and author of the book "Designing the User Interface." Shneiderman presented an entertaining and provocative overview of interface issues and the effect that interface design may have on automated survey data collection.

Over three hundred and sixty attendees from the U.S. and over 20 other countries enjoyed a program ranging from demonstrations of cutting-edge data collection hardware and software to discussions of how to measure the sentiments of large groups. All sectors of the survey industry (government, academia, commercial) were represented, both as participants and attendees at the conference, leading to fruitful and lively discussion of new developments and issues in CASIC. In all, there were 48 sessions containing 131 papers and demonstrations. A number of vendors displayed their products in an exhibit area that generated a lot of interest.

By all accounts the conference was an outstanding success. The Marriott Rivercenter hotel provided excellent facilities and meals. The weather cooperated fully with balmy and dry days and evenings for strolls along the riverwalk and virtually every conference session was well presented and attended.

Anecdotal reports from attendees indicated a general overall satisfaction with the conference. The words of one attendee pretty well summed things up. He said, "I've been waiting all my life for a conference like this." Further questioning exposed the feeling that it was a rare opportunity to be able to rub shoulders with peers from all over the world who had the common interest of automated survey data collection.

Members of the committee assisting Mick Couper in the organization of the conference were Reg Baker, Market Strategies, Inc.; Jelke Bethlehem, Statistics Netherlands; Cynthia Z.F. Clark, U.S. Bureau of the Census; William Connett, University of Michigan; Lee Decker, American Statistical Association; Tony

J. Manners, Office of National Statistics, U.K.; Jean Martin, Office of National Statistics, U.K.; William L. Nicholls II, U.S. Census Bureau; James M. O'Reilly, Research Triangle Institute; and Alan Tupek of the National Science Foundation. Numerous other people volunteered their time and effort to make this conference a success. In addition to the scientific sponsors, the conference benefitted greatly from the generosity of a number of corporate sponsors.

The InterCASIC conference was the sixth in a series of conferences sponsored by the American Statistical Association on the subject of survey data collection issues and methods. Previous

conference topics have included panel surveys, telephone survey methods, measurement errors, establishment surveys, and survey measurement and process quality. As with the other conferences in the series, a conference monograph from InterCASIC '96 is in preparation and is expected to be available from John Wiley and Sons in early 1998.

For further information contact William E. Connett, Ph.D. at Survey Research Center, University of Michigan, E-mail [bconnett@isr.umich.edu](mailto:bconnett@isr.umich.edu).



## Question/Answer

These two questions were received and the answers supplied by Prof. Leslie Kish, ISR, the University of Michigan, Ann Arbor, MI 48106 USA. He adds that these general remarks must take diverse forms for specific applications, and these will require technical details that are missing here. Please send questions and observations on this column to: Professor Vijay Verma, University of Essex, Wivenhoe Park Colchester CO4 3SQ, United Kingdom, Fax: 44 1206 873151, E-mail [vjverma@essex.ac.uk](mailto:vjverma@essex.ac.uk).

**Q36.1.** Our country needs annual sample statistics. What are "rolling samples" and how would they help?

I begin with a hypothetical country that wants a national sample to yield data for monthly (or quarterly) reports, and also annual reports. Perhaps, the sample might also be used for weekly reports for health data. Perhaps decennial cumulations of the annual reports could replace the censuses or supplement and augment them. The monthly surveys of  $n = 100,000$  can perhaps be collected as weekly surveys of  $n = 23,000$ , approximately, because the week's boundaries should be flexible for responses. Those periodic surveys commonly serve for monthly (and quarterly) comparisons (differences), but they can also be combined for the much-needed and lacking annual sample surveys. These annual statistics, based on  $n = 1,200,000$ , serve as useful averages of seasonal variations. Then, the annual statistics could be combined into a rolling census of  $n = 6,000,000$  quinquennially, or  $12,000,000$  for a decennial census. Samples of this magnitude can yield the area details that serve as the chief justification for decennial censuses.

Rolling samples are proposed to combine data from periodic surveys, which are collected chiefly for periodic comparisons, into annual sample data, and perhaps even into decennial cumulations that could replace censuses. For efficient cumulation, the sample has to be properly designed and efficient designs for cumulation differ somewhat from efficient designs for comparisons. Comparisons can benefit from overlapping units but combinations are better without overlaps.

Fortunately, some aspects of designs will be common to both comparisons and combinations. It

is imperative for both to maintain the utmost similarity possible in all survey aspects, including the definitions of concepts, variables and populations, as well as the design and methods of measurement. The sampling aspects of design, selection methods and sample sizes could vary theoretically; but practical considerations will tend to keep them similar also.

Extensions of the methods of rolling samples are possible in space and time. The periods may be merely the days of one month of an experiment, or the duration of an epidemic. On the other hand, the 120 months of a decennial period seem to require more profound thoughts about the meaning of variations of populations over time. Such profound thinking seems to be missing from the acceptance of decennial censuses.

The spatial extensions of rolling samples are also useful in both directions. First, it is useful and common to recognize that national samples are comprised of regional (or provincial etc.) domains, which are diverse and are also recognized in analysis. Even for the sampling of a single city, distinct domains are commonly recognized in both the design and the analysis of the sample. Second, today even national samples are increasingly viewed as parts of multinational samples. Statistics on education, wealth, unemployment, crime, and illiteracy take on more meaning in multinational comparisons and combinations. Nowadays we witness the designs of increasing numbers of multinational surveys.

Also consider two kinds of compromises from rolling samples. Asymmetrical cumulations refer to proposals that temporal detail, such as monthly reports, are possible for the national (or other) totals based on the entire sample. On the other hand, temporal cumulations, such as yearly totals, may be preferable for spatial details, such as districts and counties. Similarly, there may be substantive asymmetry with the more frequent reporting providing the main, more frequently required and/or changing results, and the cumulative reporting providing estimates requiring more detailed cross-classifications or otherwise larger sample sizes. Technical details are needed to balance biases and variances. However, the panel component can also be more difficult to implement and maintain than a series of independent cross-sectional samples.

Rolling samples are better designed without overlaps of sampling units. If the positive correlations from using the same units over periods are needed, I propose instead split panel designs (SPD). These would be separate samples that are present in all periods. These have major advantages over the partial overlaps now in use: a) They yield the correlations between any pair of periods, not just those arbitrarily designated by foresight; and b) They could be based on true panels and thus yield micro-, not only macro-changes.

**Q 36.2.** We are designing a national sample survey for diverse social and economic data. We prefer an equal probability of selection, but we have unusual problems. We have 23 provinces that are very unequal in size: the smallest is 60 times smaller than the largest. We would like to have also provincial data with adequate precisions for the small provinces also. What do you advise?

Your problems are not "unusual," and I have seen them in every country I dealt with in 50 years. The number of major domains (provinces, states, districts) may be less than 10 or more than 100; but the range from smallest to largest (in persons or households) is typically 1:50 to 1:100; these great ranges are due mostly to population density more than to differences in areas. Whereas the literature emphasizes the designs for the combined national population, sampling statisticians are often faced also with demands for good domain statistics, just as you are. Furthermore, even if either the combined national or the separate domains are made to appear more important initially, the emphasis of the "decision makers" may shift to the other later in the analysis, and we samplers should anticipate those shifts.

Some argue that the combined national total needs less emphasis because its larger size results in smaller errors than the errors for the provinces; in your case smaller by  $\sqrt{23}$ , or 5 roughly. But then we, the samplers, must point out that most subclasses are "crossclasses" like age, occupation, education, etc.; any increase of variances for the combined sample will also affect similarly the crossclasses, which may be critical.

These conflicts between designing for domains (provinces) and for the combined totals are most

common. We cannot avoid or neglect these conflicts, but must face them. Perhaps we should even enjoy them, because they enliven our profession, especially because we can bring good professional solutions to the conflict, as we shall see. Incidentally, similar conflicts and similar solutions also arise in designing for other populations, such as establishments, hospitals, universities, etc.

The conflict may be stated simply and boldly, but adequately, by beginning with the combined mean  $\bar{y}_w = \Sigma W_h \bar{y}_h$  of the domain means  $\bar{y}_h$ , where  $W_h = N_h / N$  are the relative sizes of the domains ( $h$ ). The optimal allocations for the combined mean calls for proportionate allocation,  $n_h = nW_h$ ; a simplified version of the Neyman allocation  $n_h = nD_h S_h / \sqrt{C_h}$ , based on the assumption that element variances, costs and design effects are constant among the domains. These factors often have negligible practical effects, compared to the  $W_h$ , but if needed, they can also be handled easily with similar treatments of fuller, more general formulations (Kish 1976; Kish 1988). The relative variance of  $\bar{y}_w$  can be stated with the loss function  $1 + L_w = n \Sigma W_h^2 / n_h$ , which is minimal at 1 for proportionate (epsem) sampling  $n_h = nW_h$  with allocations proportionate to domain sizes  $W_h$ . The relative loss  $L_w$  shows the proportionate increase of the variance due to the departures of the allocations  $n_h$  from the optimal  $nW_h$ . (See references.)

On the other hand, for equal precisions for the  $H$  domains (and for domain comparisons) the best allocation would be  $n_h = n/H$ , equal sample sizes for each domain  $h$ . For the unweighted mean  $\bar{y}_u = \Sigma y_h / H$ , the loss function is  $1 + L_u = n \Sigma (n_h H)^{-1}$ ; this is minimal at 1 when the sample sizes  $n_h$  are all equal at  $n_h = n / H$  for all domains, or "equal allocation."

Such equal allocation  $n/H$ , optimal for the separate means results in great losses  $L_u$  for the combined mean; but proportionate allocation  $nW_h$  results in even greater losses  $L_u$  for the separate means  $\bar{y}_h$ . But compromises happily can yield dramatic improvements with only small losses  $L_w$  and  $L_u$  for both kinds of estimates. Many compromises are possible, and a well-known one is "square-root

allocation," where  $n_h \propto \sqrt{W_h}$ , shown in the table. Almost any compromise is better than the extremes (as in politics?), but there is an optimal allocation, when  $n_h = n\sqrt{(W_h^2 + H^{-2})}$ , which yields dramatically improved results in relative variances for both means,  $\Sigma W_h \bar{y}_h$  and  $\Sigma \bar{y}_h / H$ . The derivations are justified and several numerical results are given in the references.

Here, we examine the empirical case of the  $H = 10$  provinces of Canada where the relative sizes  $W_h$  have a range of 70:1, as shown on the first line with 1,000  $W_h$ , which would be the sizes  $n_h$  for a proportionate sample of  $n = 1,000$ . The other allocations follow, with  $n_h = n/H$ ,  $n_h \propto n\sqrt{W_h}$ , and  $n_h \propto n\sqrt{(W_h^2 + H^{-2})}$ , all standardized to total  $n = 1,000$ , to make for fair comparisons; and these relative comparisons hold for any  $n$ . The equal allocation  $n_h = n/10 = 100$  would increase the variance of  $\Sigma W_h \bar{y}_h$  by a factor of  $1 + L_w = 2.30$ . The proportionate allocation of  $n_h = nW_h$  would increase the variance of the separate means  $\Sigma \bar{y}_h / H$  by an average of  $1 + L_u = 3.93$ . Optimal allocation for the combined mean is terrible for the other, leading to a drastic conflict. But the optimal compromise yields  $1 + L_w = 1.26$  and  $1 + L_u = 1.22!$  Very similar dramatic improvements will be found in most countries, including yours, I wager; and you can make your own calculations with your own  $W_h$ . Let me know.

The optimal allocation with  $n_h / n \propto \sqrt{W_h^2 + H^{-2}}$  has theoretical justification and also makes practical

sense. It puts a flat floor under the smaller domains where the  $H^{-2}$  term dominates the small  $W_h^2$  terms. In Canada's case, a flat minimal  $n_h = 67$  for the smaller 6 provinces would do very well. The "common sense" solution of  $n_h / n \propto \sqrt{W_h}$  would get  $1 + L_w = 1.16$  and  $1 + L_u = 1.35$  in Canada, which is not bad, though not as good as the optimal. It can do less well in some countries with more small provinces and other *ad hoc* solutions even worse.

Furthermore the theoretical optimal allocation can easily accommodate modifications. For example, different variance and cost factors can be introduced; also different relative weights for the two statistical aims (e.g. 0.8 and 0.2 instead of equal weights 0.5 and 0.5); also more than two principal aims. These are easily handled and make little difference (see references).

#### References

- Kish L. (1976) Optima and Proxima in Linear Sample Design, *J. Royal Stat. Soc. (A)*, 139, p. 80-95.
- Kish L. (1988) Multipurpose Sample Design, *Survey Methodology*, 14, p. 19-32.
- Kish L. (1990) Rolling samples and censuses, *Survey Methodology*, 16, 63-79.
- Kish L. (1994) Multipopulation survey, design, *Int. Stat. Rev.*, 62, 167-86.
- Kish L. (1997) Periodic and rolling samples and censuses, in Spencer BD, *Statistics and Public Policy*, Oxford Univ. Press.

Table 1. Sample sizes  $n_h$  for four allocation

| Allocation                             | Domain (h) |     |     |     |     |     |     |     |     |     | Total |
|----------------------------------------|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-------|
|                                        | 1          | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |       |
| $n_h = nW_h$                           | 370        | 253 | 121 | 94  | 40  | 36  | 33  | 27  | 21  | 5   | 1,000 |
| $n_h = n/10$                           | 100        | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 1,000 |
| $n_h \propto n\sqrt{W_h}$              | 223        | 185 | 128 | 113 | 74  | 70  | 67  | 61  | 53  | 26  | 1,000 |
| $n_h \propto n\sqrt{(W_h^2 + H^{-2})}$ | 243        | 173 | 100 | 87  | 68  | 67  | 67  | 66  | 65  | 64  | 1,000 |



## Country Reports

### AUSTRALIA (from Susan Linacre)

The **Business Growth and Performance survey** is a new longitudinal survey of businesses; its prime aim is to assess which factors have a significant impact on firm growth and performance, particularly small and medium sized firms. It was first conducted in respect of the financial year 1994-95 and the second survey (1995-96) is currently in the field.

The survey questionnaire comprises two parts - the first is a collection of growth and performance variables and the second gathers data about the factors which may affect firm growth and performance. The survey design is a two stage sample design. The first stage, in respect of 1994-95, was a stratified random sample of businesses on the ABS Business Register. The second stage, commencing in 1995-96, is a sub-sample of the original selections but further stratified by export status, innovation status and growth status. These latter characteristics were not available from the Business register for the first stage design. The 1995-96 sample is further supplemented by a stratified random sample of new businesses added to the Business Register since the initial selection.

Initial results from the first survey were published on September 3, 1996 by ABS in Small and Medium Enterprises, Business Growth and Performance Survey, Australia - Cat No 8141.0. More detailed information is available from ABS, who are providing a special data service from this survey to assist researchers with data. Results from the second survey are expected to become available in the second half of 1997. (Contact: Bill Pattinson, ABS, telephone 61 6 2525019, fax 61 6 252 7004, E-mail: bill.pattinson@abs.gov.au)

The **National Survey of Mental Health and Well-being** is a joint project between the ABS and the Commonwealth Department of Health and Family Services. It is to be conducted by the ABS during mid 1997. It will provide benchmark information on how many Australians have specific mental disorders, what disablement is associated with these disorders, what use is made of health and human services. The survey will cover adults aged 18 years and over living in private dwellings.

Institutions such as hospitals and nursing homes will be excluded.

Information will be collected on computer in a personal interview, covering mental health topics such as depression, anxiety, drug and alcohol use and post traumatic stress. Results from the survey are expected to become available in early 1998. (Contact: Marelle Rawson, ABS, telephone: 61 6 252 7995).

### CANADA (from Gordon Brackstone)

**Project to Improve Provincial Economic Statistics:** Recently, the Government of Canada and the governments of Nova Scotia, New Brunswick and Newfoundland and Labrador signed an agreement to harmonize their federal and provincial sales taxes effective April 1, 1997. The federal government will collect the sales tax revenue, set at a rate of 15%. The provinces' shares of the revenue will be calculated using Statistics Canada's provincial economic accounts. To improve the quality of the data used in the allocation formula, Statistics Canada will be undertaking a major project during the next three years, known as the Project to Improve Provincial Economic Statistics (PIPES). The project will involve a significant overhaul of the business survey program of Statistics Canada, including both the development of new surveys in industries where data gaps exist, as well as the redesign of existing surveys under a standard framework known as the Unified Enterprise Statistics Program. A key feature of the new approach is the use of the enterprise as the key statistical unit for purposes of data collection, sampling and analysis. The first year of the new program will see the introduction of a Census of Complex Enterprises, covering the 8,000 enterprises which operate in more than one industry or province, as well as new surveys in construction, services and transportation industries.

For further information on the PIPES project, please contact Don Royce, Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6. Tel (613) 951-1461. E-mail: roycdon@statcan.ca.

**The Survey of Labour Income Dynamics (SLID)** is a panel survey. Each panel, which continues for six years, consists of a sample of approximately

30,000 individuals. Respondents are interviewed twice yearly: once in January (to obtain information on labour characteristics) and a second time in May (to obtain income details). The first panel started in 1994 (i.e., collection of data began that year). SLID released its first year of data in 1996. The release of the second year of data (and hence the first release of longitudinal data) is scheduled for summer 1997. In preparation for the creation of the longitudinal file, various methodological research projects were carried out. A number of studies looked at cross-sectional and longitudinal weighting. Analyses were carried out on characteristics associated with erosion. Research was also conducted on the inclusion of an adjustment for "longitudinal" non-response in cross-sectional estimates, in order to determine its impact on measurements of level and trends. Analyses were also carried out on response errors relating to income questions. Some of these projects will continue in the coming years. In particular, efforts will be made to quantify the impact of procedures for re-interviewing non-respondents on level and trend estimates. Projects currently under way are examining imputation procedures to compensate for non-response, more specifically for the longitudinal imputation of income variables. A method of correcting aberrant changes has been included in the processing procedure. Longitudinal data also pose specific challenges for the production of microdata files, and in a study that attempts to take account of the dynamic nature of some characteristics available in a longitudinal file, the risks of disclosure of confidential information are being evaluated. A final point to be noted is that starting in 1998, the annual cross-sectional Survey of Consumer Finances (SCF) will be integrated into SLID. This is part of a process of harmonization of income statistics produced at Statistics Canada. The integration of these surveys poses particular challenges with respect to sampling and estimation. Several research activities are under way to examine various options for combining the samples for the two surveys.

For further details, contact Sylvie Michaud: Michsylv@statcan.ca, Tel. (613) 951-9482.

#### **ITALY (from Claudio Quintano)**

**The CENSUS project (ISTAT):** As is commonly known, before carrying out a census, the area to be

surveyed is generally divided into segments of different sizes and characteristics (municipalities, built-up areas, census sections, enumeration districts, blocks, etc.). Besides facilitating the organization and completeness of the survey, this subdivision offers the possibility of referencing the collected statistical information to these micro areas, thus providing information at a very disaggregated level. To this end, satellite remote sensing techniques were applied by ISTAT for the identification and mapping of urban areas within the CENSUS project which was implemented in order to define the territorial bases of the population and housing census.

The most significant innovations of the CENSUS project were: 1) identification of urban areas by means of satellite remote sensing and their digitization; 2) digitization of census section boundaries.

The urban areas were identified through the interpretation of remotely-sensed SPOT satellite panchromatic images, which indicated the actual distribution of the urban areas in Italy at the time of the census. These boundaries were digitized, plotted and sent to the municipal administration. In fact the digitized maps representing the municipal boundaries and the inhabited localities as identified by the satellite images were submitted to the municipal administration to check the accuracy of the information and to indicate the exact nature of the urban areas reported on the maps.

Another important innovation was the digitization of the census section (CS) boundaries. During the 1991 census the municipal administration subdivided the municipal areas into census sections, either on transparent maps provided by ISTAT or on large scale maps, according to ISTAT's requirements. After ISTAT's verification of these data, the boundaries of each CS were digitized thus enabling digitized mapping showing the subdivision of all municipal areas into census sections. Since each of these micro-areas must lie fully within only one urban area, this innovation permits the integration and automatic connection between CS's and urban areas.

The table below provides a quantitative characterization of the territorial base of the census sections.

Number of census sections and its percentage variation '81-'91, by demographic size of community

| Demographic size  | Number of sections 1991 | % increase 1981-91 | Number of communities |
|-------------------|-------------------------|--------------------|-----------------------|
| <=3,000           | 48,800                  | 50.5               | 4,680                 |
| 3,001 - 10,000    | 59,931                  | 66.8               | 2,381                 |
| 10,001 - 20,000   | 28,740                  | 69.4               | 585                   |
| 20,001 - 30,000   | 18,437                  | 121.5              | 168                   |
| 30,001 - 50,000   | 39,891                  | 311.5              | 150                   |
| 50,001 - 100,000  | 43,852                  | 257.4              | 90                    |
| 100,001 - 500,000 | 51,571                  | 169.3              | 40                    |
| > 500,000         | 32,294                  | 80.7               | 6                     |
| Italy             | 323,516                 | 111.9              | 8,100                 |

Firstly it is noticeable that the number of census sections has increased more than twofold, in comparison with the 1981 census. The highest percent increase occurred in communities with population between 30,000 and 100,000. Notable increases occurred also in communities with population between 100,000 and 500,000.

Possible applications: the resulting maps are thematic maps of the Italian territory. Their integration is the first step towards the construction of a Geographic Information System by ISTAT. Among other things, this system should facilitate the development of spatial data analysis within Italian official statistics.

The digitized database of the boundaries of municipalities, inhabited localities and census sections made up the geographic frame for statistical information produced by ISTAT, and will provide the initial spatial reference of the next population census.

It is very important to observe that this system can be implemented with data bases of geocoded results referring to census sections. The integration of thematic maps with statistical geocoded data bases will take on particular importance in the near future. The system will provide thematic maps locating significant demographic, economic and social phenomena.

For more information, contact Dr. Pierpaolo Napolitano, ISTAT - Dipartimento Popolazione e Territorio, via Adolfo RavB, n. 150, Roma, 00142 Italy. Tel: +39 6 54900331; Fax: +39 6 5943257

**JAPAN** (from Chikio Hayashi)

**The History of NHK's Time-use Survey:** NHK (Japan Broadcasting Corporation) conducted its first time-use survey in 1941 (during the war time), following the example of a BBC survey. Since the practice of random sampling was not yet established, that first survey covered people only in particular social strata, and lacked numerical data for the whole of the nation. Furthermore, the classification of activities in the first survey differs from that now in use, and different classifications were made according to occupations, such as farming. In the 1941 survey, the total period of people's activities added up to 24 hours a day, as only their primary activities were considered. The survey covered activities not on particular days but on what were called "ordinary days". Questionnaires were distributed to respondents aged over 16 and then retrieved. Four rounds of inquiries were made (in spring, summer, autumn and winter). It is difficult to compare the results of that survey with those of today's surveys without detailed study of the differences between the surveys.

NHK's first post-war time-use surveys, in 1960 and 1965, did not differ substantially from the more recent type of survey conducted once every five years since 1970. (A separate survey was also conducted in 1973). But in the 1960 and 1965 surveys, the respondents aged over 10 were interviewed and there was no sub-classification of activities. The surveys since 1970 employ the method of distributing and then collecting questionnaires. Unfortunately, it is not possible to make a direct comparison of data obtained in the surveys before and after 1970 because of

differences in surveying methods and the discontinuity of figures.

Since 1960, at intervals of ten years, a large-scale survey has been conducted on a prefectural basis.

Between 1970 and 1990, we were able to analyze trends in Japanese time-use under an unchanged survey method. For example: time spent for "sleep" has continued to decline on each day of the week; time for "work" decreased on Saturdays and Sundays and increased on weekdays; time for "housework" on weekdays continued to decrease from 1980 to 1990; time for "leisure activities" generally increased for each day of the week compared between 1980 and 1990; time for listening to the "radio" decreased for each day of the week, and time for viewing "video" increased from 1985 to 1990; and time for viewing "TV" moved up and down since 1970 to 1990.

Thus NHK has been monitoring the changes in Japanese life-style over a long period of time.

For more information, please contact Tetsuo Makita, Japan Broadcasting Corporation, Public Opinion Research Institute, 2-1-1 Atago, Minato-ku, Tokyo 105 JAPAN. E-mail: MAKITA@culture.nhk.or.jp

#### LITHUANIA (from Danute Krapavickaite)

A new methodological approach was taken in the monthly **Survey on the Average Salary of Employees** in 1997. The sampling frame was obtained from the Business Register of the Lithuanian Department of Statistics and contains approximately 21,000 enterprises. The PSU is the part of the individual company engaged in a certain kind of activity. Stratified simple random sampling has been used. The problem of strata formation was examined using data from a complete survey in 1995. The main observed values - total salary and number of employees - have a high variance even in the domains of interest. The accuracy requirements in small domains dictated the stratification method. Forty-four nonintersecting domains of interest were stratified into 2, 3 or 4 strata according to the number of employees. All possible ways of stratification, satisfying the accuracy requirements, were examined and the one with the smallest sample size for a particular

domain was chosen. The coefficient of variation of the estimate in a domain was used as the measure of accuracy. The sample size was 3,990 PSUs. Approximately 25% of enterprises have changed their characteristics by the time of the inquiry, so - estimators are mainly used for estimation of totals. The coefficient of variation for 82% of the estimates does not exceed 0.03 and for 99% of estimates - 0.05.

The new **Household Budget Survey (HBS)** was started in January 1996. The objective of the HBS is to provide information on household incomes, expenditure and consumption in order to find out the living standard of the Lithuanian population and solutions for social development of the society. The survey was prepared by the staff of the Household Surveys Division with assistance of experts from the World Bank, headed by Prof. Jan Kordos. HBS is a continuous survey, carried out each month. The survey population is all private households in Lithuania. The survey does not cover collective households (nursing homes for children and elder citizens, hospitals, hotels, sanatoria, prisons, etc.). For sample selection the Population Register was used in urban areas, and available lists of households - in rural areas. The total annual sample size is equal to 10,680 households (with an expectation of obtaining approximately 8,000 responding households). In urban areas stratified one or two stage selection of persons aged 16 and over was made. PSUs were selected with probabilities proportional to the population size. Two stage sampling was introduced in rural areas: the selection of PSUs (smallest administrative units) was carried out with probabilities proportional to the number of households; simple random sampling was used for household selection. The -estimators are used in the estimation procedure.

A new estimator for our country was used in the **Labour Force Survey** carried out in September 1996. A simple random sample of 14-74 year old persons was selected from the Population Register. The household (PSU) was selected if any of its members were selected. Thus, a cluster sample with unequal selection probabilities was obtained. The effective sample contains 7,258 individuals in 2,686 PSUs. Afterwards the sample was poststratified into 13 age, 2 sex and 2 urban-rural poststrata, 52 poststrata in all. Estimates for population sizes greater than 15% have a coefficient of variation less than 0.03, estimates for

areas less than 2.5% have coefficient of variation greater than 0.06. The problems of small area estimates and computational resources for the calculation of variances in unequal probability sampling with poststratification remain to be solved.

The **22nd European Meeting of Statisticians (EMS'98) and the 7th Vilnius Conference** will be held on August 12-18, 1998 in Vilnius. The programme contains 35 sessions. They will cover the following topics: general inference; time series; statistics and probability in the life sciences; statistics and probability in natural and social science; applied probability; probability. A special session on the Problems of Official Statistics and the round table discussion "Statistics in Economy" will be held. The address of the Organizing Committee: Institute of Mathematics and Informatics, 4, Akademijos str., Vilnius, Lithuania. For further information, please contact the coordinator (chairman of the local organizing committee) - Prof. V. Statulevicius, Secretariat - Dr. M. Bloznelis, Dr. A. Plikusas. Telephone: 370-2-729207, FAX: 370-2-729209. E-mail: conf@ktl.mii.lt.

#### **MOROCCO (from Bouzaffour Seghir)**

**A new survey methodology in Morocco complements the General Census of Agriculture.** Morocco has been working for a number of years to replace its old system of data collection, a system rooted in the rural tax base, with a new system based on stable material: satellite images, topographic maps and aerial photographs of different scales. The new system, known as the "Area Survey Base", is being applied to all farming regions of the country and is very attractive on many counts: it is easy to update, it can be used to monitor farmers' returns; it allows easy retrieval from fields to gather data for estimating yield; it allows partial renewal by sample rotation; etc.

From a qualitative standpoint, the old system offered none of these features; this fully justifies the use of the area survey base in developing countries, despite the small size of the fields and the lack of natural boundaries (trails, roads, rivers, railway lines, etc.), which can sometimes limit its use.

Briefly put, the area base calls for identifying a study area and then grouping all lands used for the same purpose into the same category to form strata. These strata are then subdivided into geographic areas (called "natural zones") that are used for digitization and for first-stage selection (probability proportional to the size of each natural zone). Selected natural zones are further divided into "segments". Within each selected natural zone a sample of segments is selected randomly with equal probabilities of selection.

There are four methods of gathering data in the field: 1. the closed segment method (applies to all operators); 2. the open segment method (applies to resident operators); 3. the weighted segment method (proportionality criterion that draws on the concepts of closed and open segments); and 4. the multiple base method (combination of closed segment and list of operators).

To ensure first and foremost the representativeness of the animal husbandry and cottage farming (intensive farming, greenhouse crops, etc.) sector, Morocco has prepared an initiative broader in scope than the General Census of Agriculture. This comprehensive undertaking will provide a survey base (a list of all farmers in the country) for socio-economic surveys and surveys with a more specific purpose.

The General Census of Agriculture, which began in November 1996 and is to be completed by April 1997, will be followed by two particularly interesting surveys.

The first, a "post-census survey", will be done with maps to ensure accurate measurement of lots. The measurements will be used to establish a correction factor for adjusting farmers' returns. The sample will also be used to determine whether the process was sufficiently comprehensive.

The second, a "structural survey", will use a sample of 30,000 to 40,000 operators representing the different types of farm based on data from the General Census of Agriculture (Recensement général de l'agriculture) and will serve to establish farm structures so that the way farms function can be studied in greater detail.

## PHILIPPINES (from Gervacio G. Selda, Jr.)

The National Statistics Office (NSO) of the Philippines has developed a manual for the conduct of its **1996 Multiple Indicator Cluster Survey (MICS)**, a survey designed to generate information on the health of children under five years old, basic education of those 5 to 15 years old, and household's access to safe and convenient water and sanitary facilities. The 100-page manual covers the survey's sample design, concepts and definitions, questionnaires and their completion instructions, field operations' instructions, training schedule, and other pertinent survey information. This manual might be useful to those who are planning to conduct a similar survey. (Contact: Tomas P. Africa, Executive Director V of the National Statistics Office, Solicarel 1 Bldg., Sta. Mesa, Manila, Philippines. Fax no. 632-7130773 or E-mail: census@mnl.sequel.net).

**Development of a New Master Sample Design for Household Surveys:** The National Statistics Office (NSO) has developed a new master sample design for its Integrated Survey of Households last July 1996. The new design is intended to: (a) meet the demand for an expanded scope and coverage of household surveys; (b) address the need for more reliable estimates at more detailed levels of disaggregation; and (c) maximize the use of limited resources available for the conduct of surveys. The new master sample includes a core sample and an expanded sample that can provide reliable estimates at regional and provincial level, respectively, using a frame based on the results of the 1995 Population Census. The development was done in two phases. The first one involved the development of the sampling frame, selection of sample barangays/enumeration areas (EAs) and selection of sample households from EAs. The second phase involved the formation of primary sampling units (PSUs) with a prescribed minimum measure of size, segmentation of PSUs, selection of sample segments, and selection of sample households from sample segments. The rotation scheme to be adopted for segments of households is also part of the second phase. The results of the implementation of the design showed that: (a) the expanded sample as used in the July 1996 Labour Force Survey (LFS) round improved the overall precision of survey estimates at the provincial and city/municipality domains, and (b) the core sample as used in the October 1996 LFS round gave

reliable estimates at the regional level but less precise provincial estimates. There was an increase in the efficiency of the sampling design from the implicit stratification for the LFS estimates at the national level and for most of the regions. (Contact: Tomas P. Africa, Executive Director V of the National Statistics Office, Solicarel 1 Bldg., Sta. Mesa, Manila, Philippines, Fax no. 632-7130773, E-mail: census@mnl.sequel.net).

The National Statistical Coordination Board (NSCB) of the Philippines conducted the **Survey on the Statistical Practices of Asia-Pacific Countries** in July 1996. The objective was to find out information on the availability, frequency, mode of dissemination, time lag, and disaggregation of some key indicators, such as labour force statistics, foreign trade statistics, consumer price index, poverty statistics, food balance sheet, labour productivity statistics, wage statistics, national accounts, social accounting matrix, input-output tables, balance of payments, and environment statistics. Preliminary results based on a 40% response rate were distributed at the 10th Session of the Committee on Statistics held in Bangkok, Thailand last November 25-29, 1996. (Contact: Secretary-General Romeo A. Virola of the National Statistical Coordination Board, Midland Bldg., 403 Gil Puyat Ave, Makati City, Philippines. Fax no. 632-8909397/8952395 or E-mail: nscb@mozcom.com).

## SPAIN (from Montserrat Herrador Cansado)

The National Statistical Institute (INE) of Spain has created **Centralized Collecting Units (URCEs)** in the bigger provincial statistical offices as well as in the Central Services in Madrid to collect the data of the annual economic surveys. The usual approach adopted by INE for carrying out these surveys involves mailing out questionnaires to respondents and then following up through telephone calls, personal interviews and reminder letters from the 52 provincial statistical offices. Subsequent recording and processing are conducted by the Central Services in Madrid.

Several reasons led to the creation of URCEs:

- To improve timeliness by conducting simultaneously activities that were previously performed serially;

- To improve the quality of the data by making interviewers responsible not only for the actual collection of the questionnaire but also for the quality of the data contained therein;
- To increase efficiency by enabling the peaks of activities faced by provincial statistical offices to be evened out;
- To launch "small" surveys for which decentralization would be more expensive than centralization. Decentralization requires several selection and training sessions for interviewers and rather complicated follow-up and control systems.

Before creating the actual URCEs, in 1995 an experiment was carried out with the 1994 Survey on Annexed Transport Activities, with 4000 sampling units. Its aim was to test the feasibility of the project as to the need for personal interviews, working methods, collection quality and cost of the new method. Given the results, four URCEs have been created with room for 160 survey agents. To this end, each workstation has been provided with a computer and a telephone. The above mentioned experiment gave rise to the decision to help interviewers and inspectors in the following activities :

- Checking of the respondent units' status. Each contact with the enterprises causes the shifting of units from one state to another according to the available information. The states are defined beforehand and refer to the most usual situations such as SE (the enterprise advises that it will send the questionnaire), EX (the enterprise advises that the questionnaire was misplaced), EN (the enterprise advises that the questionnaire has already been sent), RE (the questionnaire has been received), NC (it has not been possible to contact the unit), etc. Each state requires the performance of different activities;
- Scheduling of the work programme. It establishes for each interviewer the activities of the day, according to each unit's state and to whether an appointment has been granted or not;

- Recording and processing of the questionnaires. The processing is conducted with the aid of several screens which inform the interviewer of the controls to which the units are to be corrected. Two types of error have been distinguished: type I errors do not allow interviewers to mark the questionnaire as clean and then the questionnaire must be revised; type II errors force interviewers to get into contact with the enterprise to confirm the data or to obtain additional information. After this, a questionnaire may be considered clean or not;

- Reports for follow-ups and controls;
- Maintenance of the data base of the respondent units;
- The inspector is in position to make new reminder calls, confirm or return clean questionnaires as well as the shortcomings marked by interviewers, etc.

Up to now the data for the following surveys have been collected by URCEs (as a whole or partially):

- 1994 Survey of Annexed Transport Activities;
- 1995 Survey of Salary Structure;
- 1995 Survey of Financing and Spending for Private Education;
- 1995 Survey of Intermediary Consumptions;
- 1995 and 1996 Annual Industrial Survey of Enterprises;
- 1995 and 1996 Annual Industrial Survey of Products;
- 1996 Survey of Labour Costs.

The information obtained from these experiences (relative to the number of necessary contacts per information unit, the form of the collecting cycle, etc.) is proving very useful in improving the method and in planning new surveys.

**Computer assisted data collection in the Labour Force Survey:** The main objectives of computer

assisted data collection replacing the traditional system with paper questionnaires, are the following:

- To improve the quality of the information thanks to directed interviews and internal control, thus avoiding flow errors and some inconsistencies.
- To integrate the complete documentation used in the Survey, into a single integrated questionnaire, thus avoiding the need for the Interviewer to use different forms and reference books.
- To allow codification and editing of the data in the course of their collection.
- To collect and register the information simultaneously and to do this as close to the source and as early as possible.
- To reduce the delay in making data available.

In short, the quality of the Survey is improved and the results are obtained sooner, thereby facilitating the interviewer's task and the follow-up.

The application incorporated into the portable computer consists in an integrated, directed and controlled electronic questionnaire. The questionnaire is integrated because it contains all the forms and books needed to carry out the LFS, such as: the document indicating the addresses of the dwellings to be visited in each section; the LFS questionnaire proper; the refusals questionnaires, used when there is refusal to co-operate for the first time; the Rent Card questionnaire, an auxiliary tool for collecting data on rented dwellings; the Dwelling Card, recording the interviews carried out throughout the 6 quarters of the survey, as long as the group living in the dwelling remains the same; the Interviewer's Field Work Manual; classifications or coding systems for a series of variables; and consulting lists.

It is a **directed** questionnaire because the application proper directs the interviewer's work telling him which questionnaire to use in each case and when and how to replace them. It is a **controlled** questionnaire because it contains several controls during and after the field work.

This new system of data capture started in the first quarter of 1996 in five provinces. In the first quarter of 1997 this procedure has been extended to another 22 provinces. During this year we hope the system will be implemented in the remaining provinces.

#### **UNITED STATES (from Dan Kasprzyk)**

The Agency for Health Care Policy and Research (AHCPR), in conjunction with the National Center for Health Statistics (NCHS), sponsors a project which provides the health policy and research community with detailed information on the medical utilization and expenditures of the United States population on an ongoing basis. Data collected include detailed information on: health services that Americans use, how frequently they are used, the cost of these services, and method of payment for services rendered, as well as on the cost, scope and breadth of private health insurance held by and available to the United States population. It is designed to help understand how the growth in managed care, changes in private health insurance, and other dynamics of today's market driven health care delivery system have affected the types, amounts, and costs of health care that Americans use. The Medical Expenditure Panel Survey (MEPS) obtains this information through in-person interviews of households, nursing homes, and medical providers as well as telephone interviews and/or mail surveys of establishments. For more information, contact Steve Cohen at 301-594-1400 or look for more information on the Web at [www.meps.ahcpr.gov](http://www.meps.ahcpr.gov).

~~~~~

IASS Invited Papers Program in Istanbul

The IASS has organized five invited papers sessions for the Istanbul Meetings. Each paper treats a specific operation of the survey process. These papers are described below.

Topic 42: Imputation and Missing Values
Organizer: J.N.K. Rao (Carleton University)

Steven G. Heeringa and Roderick J. Little (University of Michigan). Bayesian Estimation and Inference for Multivariate Coarsened Data on U.S. Household Income and Wealth.

Jun Shao (University of Wisconsin). Variance Estimation for Survey Data with Imputed Missing Values.

Eric Schulte Nordholt (Statistics Netherlands). Imputation: Methods, Simulation Experiments, and Practical Examples.

Discussants: David Binder (Statistics Canada) and Seppo Laaksonen (Statistics Finland).

Heeringa and Little apply a Bayesian approach to estimation and inference for multivariate coarsened data using the general location model. They apply these methods to both the Health and Retirement Survey and the Study of Aging and Health Dynamics. These surveys ask respondents to report dollar values for variables such as income and assets. Income and asset questions commonly generate high rates of item nonresponse. These nonresponse rates can be reduced by methods like coarsened data. The term "coarsened" refers to data that are a mixture of actual value responses, interval-censored responses (to questions that are designed to place a financial amount in a known interval whenever a respondent is unwilling or unable to provide an exact response) and completely missing data.

Shao addresses variance estimation for Horvitz-Thompson type estimators of totals from survey data with both imputed responses and non-negligible sampling fractions. Shao proposes a method based on variance decomposition. The method can be applied to complicated situations where a composite of deterministic or random imputation methods is used, including using

imputed data as donors for missing data. The U.S. Census Bureau's Annual Motor Freight Transportation and Warehousing Survey is used to demonstrate the method.

Nordholt describes two simulation experiments of the hot deck method. In the first study, data are randomly generated and various percentages of missing values are then non-randomly "added" to the data. In the second study, the quality of the imputation method is examined by running a simulation experiment on a subsample of data from the Dutch Housing Demand Survey.

Topic 43: Generalized Estimation in Survey Sampling and Associated Software
Organizer: David A. Binder (Statistics Canada)

Nico Nieuwenbroek, Robbert Renssen, Han Hölsgen and Maaïke Miedema. A General Weighting Package Inducing Estimates for Population Totals and Corresponding Variances.

D. R. Bellhouse (University of Western Ontario), M. A. Hidioglou (Statistics Canada) and J.E. Stafford (University of Western Ontario). Generalized Estimation Systems with Future Enhancements Using Symbolic Logic.

D. Morganstein and M. Brick (Westat). Computing Sampling Errors from Clustered, Unequally Weighted Data Using Replication: WesVarPC.

Discussant: Susan Linacre (Australian Bureau of Statistics).

This topic covers three different generalized estimation systems. Nieuwenbroek et al. describe a software package that combines weighted sampling data using auxiliary information and variance estimation based on a resampling method. Its usefulness for practical situations is discussed.

Bellhouse and Hidioglou discuss the Generalized Estimation System (GES) using auxiliary information to produce estimates for one-stage designs. The package uses regression estimators known as GREG (Generalized Regression Estimator) and includes the wider family of calibration estimators. This allows the extension of GES to the analysis of complex data.

Morganstein and Brick address the problem of inference from clustered, unequally weighted data (e.g., survey data adjusted for nonresponse or clinical data used in trials). These problems are familiar for data analysts. The authors propose WesVarPC, which is Windows-based, offers a user-friendly graphics interface, computes functions of weighted totals, and uses replication techniques to fit models to clustered, unequally weighted data. Their discussion covers the program's capabilities and limitations as well as some of the statistical theory underlying the use of replication methods.

Topic 44: Improvement of the Production of Gender Statistics

Organizer: Hozta Ayhan, (Middle East Technical University)

Birgitta Hedman (Statistics Sweden) and Francesca Perucci (Italy). **New Challenges in the Improvement of Gender Statistics.**

Discussant: Hermann Habermann (UNSTAT).

Alvaro Gonzalez-Villalobos (FAO), Manoel A. Soares da Cunha (Brazil), and Sara Novaro de Cosarinsky (Argentina). **Improvement of Gender Disaggregated Statistics Obtained from Agricultural Surveys: The Latin American Experience.**

Discussant: Cathryn S. Dippo (U.S. Bureau of Labor Statistics).

Yoichi Ito (Japan). **Gender Disaggregated Statistics, Its Present Stage of Development, and Tasks: The Japanese Experience.**

Discussant: Vijay K. Verma (India).

The availability of disaggregated gender statistics is becoming an increasingly important issue, as reflected in the Beijing Declaration and Platform for Action of the Fourth World Congress on Women which explicitly states that improved gender statistics is a critical concern. Ito and Villalobos present progress reports and Hedman and Perucci discuss future challenges in gender statistics.

Topic 45: Questionnaire Design, Computerized Interviews, and Data Quality

Organizer: Antonio Giusti (Department of Statistics, Firenze, Italy)

Thomas Piazza (University of California, Berkeley). **New Methodological Possibilities Offered by Computer Assisted Interviewing.**

Jean Ritzen (Statistics Netherlands). **Assessment of Quality Criteria in Redesign of Computer Assisted Survey Data Collection.**

Luigi Fabbris and Francesca Bassi (University of Padova, Italy). **Data Quality Issues in Computer Assisted Data Collection.**

William L. Nicholls II (U.S. Bureau of the Census). **Meanings of Data Quality in Assessments of New Data Collection Technologies.**

Discussant: Colm O'Muircheartaigh (London School of Economics and Political Science).

Piazza discusses aspects of experiments and other methodologies in survey interviews that affect electronic questionnaire design. Ritzen covers aspects like on-line response control (editing in computer assisted surveys is often integrated with the data collection operation). The other two papers cover various aspects of data quality in data collection. Nicholls examines the different interpretations of data quality in studies designed to assess the effects of new data collection technologies on survey data quality. He looks at three different definitions of survey quality in this context.

Topic 67: Metadata and the Internet

Organizer: Cathryn S. Dippo (U. S. Bureau of Labor Statistics).

Cathryn Dippo (U.S. Bureau of Labor Statistics) and Easley Hoy (U. S. Bureau of the Census). **Providing Metadata to Survey Staff via the Internet.**

Karsten Boye Rasmussen (Danish Data Archives). **Storage and Access to Metadata and Beyond.**

As statistical agencies rush to make statistics and microdata available via the Internet, the concern arises that there does not seem to be commensurate interest in making the requisite metadata available. Nor is much effort devoted to

investigating the general public's (the social science student in a classroom, someone in a public library, etc.) ability to search and retrieve statistical information, their uses of this information and what statistical agencies can do to promote and facilitate the public's understanding and use of the metadata.

Dippo and Hoy discuss providing metadata on production processes for the staffs of the Bureaus of Labor Statistics and the Census who work on the Current Population Survey (designers, managers, data collectors, regional office staff). Numerous issues must be addressed when designing and developing an information system that supports ongoing production and research activities. When the public can also access to the system, additional issues arise. This paper discusses these issues primarily from the user's perspective. If survey statistics are to continue to be of interest and importance in the next century, we must resolve these and many other issues related to metadata.

Rasmussen discusses providing metadata for public use microdata sets. He describes initiatives like those of the Council of European Social Science Data Archives, the International Association of Social Science Information Service and Technology, and the Network of Social Science Tools and Resources.

Organizing Committee Chair: Anders Christianson, Statistics Sweden, Box 24300, S-104 51 Stockholm, SWEDEN, Tel. +46 8 783 44 83, Fax. +46 8 783 45 99, E-mail: a.christianson@scb.se.



Symposium '97: New Directions in Surveys and Censuses; Statistics Canada, Ottawa, Ontario, Canada; November 5-7, 1997

Statistics Canada's XIV annual international methodology symposium will explore new trends and directions in designing and conducting surveys and censuses, in processing and analyzing the collected data, and in disseminating the results. Emphasis will be given to new methods and procedures that attempt to reduce costs while maintaining high standards of quality. This year's symposium will be expanded to feature both contributed and invited sessions, running in parallel,

and will be held at the Palais des Congrès conference facilities in Hull, just minutes from downtown Ottawa. For more information contact: Jack Gambino, Statistics Canada, 16th floor R.H. Coats Building, Ottawa, Ontario, Canada, K1A-0T6; telephone (613) 951-0334; fax (613) 951-3100; internet E-mail: sympos97@statcan.ca.



**IASS/IAOS Conference
Statistics for Economic and Social Development
Aguascalientes, Mexico, September 1-4, 1998**

First Call for Topics and Papers

This is a call for papers for a joint IASS/IAOS Conference to be held in Aguascalientes, Mexico, from 1 to 4 September 1998. The theme, "Statistics for economic and social development", was chosen to be of interest to both developing and developed countries, and especially Latin American countries.

The Conference will open on Tuesday, 1 September with two keynote speakers. One will be Carlos Jarque, President of INEGI; the other, Ivan Fellegi, Chief Statistician of Canada.

Geoff Hole (Statistics Canada) chairs the Program Committee whose members are: Paul Cheung (Department of Statistics, Singapore), Cynthia Z.F. Clark (U.S. Bureau of the Census), David Fitch (INCAP, Guatemala), Enrique Ordaz (INEGI, Mexico), Paulo Garonna (ISTAT, Italy), Chris Scott (Consultant, UK), Pedro Silva (IBGE, Brazil). Mario Palma (INEGI, Mexico) chairs the Local Organizing Committee.

More information about the conference is soon to be available on the three following web sites:

- English: http://www.statcan.ca/english/iass_iaos/mexico98/
- Français: http://www.statcan.ca/francais/aise_aiso/mexique98/
- Español: <http://www.inegi.gob.mx>

This IASS/IAOS Conference will feature both invited and contributed sessions. Planning of invited paper sessions has begun. The Program Committee would welcome ideas on topics of interest, names of possible organizers and of papers to be included in invited sessions.

Abstracts of proposed papers should include full information on authors and their affiliations, a contact address (including E-mail and fax), key words and text of 200-300 words. Abstracts should be submitted, preferably via e-mail (in ASCII or TEX), by fax or by mail to: Geoff Hole, Social Survey Methods Division, Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6. Fax: 1-613-951-0653, E-mail: holecj@statcan.ca.



Statistical Terminology

IASS members may be interested in a book that provides translation of statistical terms between the two official languages of our Association, English and French. This book, entitled *Statistics and Surveys Vocabulary* (Terminology Bulletin 208) was written and produced by Canada's Translation Bureau. It contains approximately 4,300 entries with 730 definitions exclusively in the field of statistics and surveys. The terms come from unilingual, bilingual and multilingual dictionaries and Statistics Canada publications. Some concepts, especially those relating to survey vocabulary, are particular to Statistics Canada, and the terms that

designate these concepts were selected because they are commonly used.

Paperbound - 14 cm x 21.5 cm - 555 p - 1992.

Cat. No. S52-2/208-1992

ISBN 0-660-57072-6

PRICE: Canada \$29.95 (Other countries) US \$29.95.

To order:

By mail: Groupe Communication Canada -
Publishing
Ottawa, Ontario, Canada
K1S 0S9

By telephone: (819) 956-4800

By fax: (819) 994-1498
1-800-565-7757 for Visa or
MasterCard orders

By Internet: Web Site: <http://www.gc.ca/termium>

By E-mail: edition@ccg-gcc.ca



In Other Journals

Journal of Official Statistics

An International Review Published by Statistics Sweden

JOS is a scholarly quarterly that specializes in statistical methodology and applications. Survey methodology and other issues pertinent to the production of statistics at national offices and other statistical organizations are emphasized. All manuscripts are rigorously reviewed by independent referees and members of the Editorial Board.

Contents Volume 12, Number 3, 1996

Limitations of Balanced Half Sampling <i>Richard Valliant</i>	225
Cross-Classified Sampling <i>Esbjörn Ohlsson</i>	241
Response Effects in Consumption Surveys: An Application of the Beta-Binomial Model to Self-Reported Drinking Frequencies <i>Timo Alanko and Paul H. Lemmens</i>	253
Estimation of Measurement Bias in Self-Reports of Drug Use with Applications to the National Household Survey on Drug Abuse <i>Paul Biemer and Michael Witt</i>	275
Changes in Interview Setting Under CAPI <i>Mick P. Couper</i>	301
Book Reviews	317
In Other Journals	327

All inquires about submissions and subscriptions should be directed to the Chief Editor: Lars Lyberg, R&D Department, Statistics Sweden, S-115 81 Stockholm, Sweden.

Survey Methodology

A Journal Published by Statistics Canada

Contents Volume 22, Number 2, 1996

In This Issue	105
Weighting and Estimation	
A.C. SINGH and C.A. MOHL Understanding Calibration Estimators in Survey Sampling	107
D.M. STUKEL, M.A. HIDIROGLOU and C.-E. SÄRNDAL Variance Estimation for Calibration Estimators: A Comparison of Jackknifing Versus Taylor Linearization	117
B.R. JAYASURIYA and R. VALLIANT An Application of Restricted Regression Estimation in a Household Survey	127
G. CHEN and J. CHEN A Transformation Method for Finite Population Sampling Calibrated With Empirical Likelihood	139
K.J. THOMPSON and R. FISHER The Application of McNemar Tests to the Current Population Survey's Split Panel Study	147
J.L. ELTINGE and D.S. JANG Stability Measures for Variance Component Estimators Under a Stratified Multistage Design	157
Y.G. BERGER Asymptotic Variance for Sequential Sampling Without Replacement With Unequal Probabilities	167
A. COWLING, R. CHAMBERS, R. LINDSAY and B. PARAMESWARAN Applications of Spatial Smoothing to Survey Data	175
J.M. BRICK, J. WAKSBERG and S. KEETER Using Data on Interruptions in Telephone Service as Coverage Adjustments	185
G.S. PANDHER Optimal Sample Redesign Under GREG in Skewed Populations With Application	199
Acknowledgments	205

IASS/IAOS CONFERENCE

1-4 September 1998, Aguascalientes, Mexico

STATISTICS FOR ECONOMIC AND SOCIAL DEVELOPMENT

A joint IASS/IAOS Conference is to be held on Statistics for economic and social development in Aguascalientes, Mexico, from 1 to 4 September 1998. Carlos Jarque, President of INEGI (the National Statistical Office) has kindly offered to host the Conference. It will be the first Conference held in Latin America by both Associations. Mario Palma of INEGI will be looking after the local arrangements. INEGI is the official agency responsible for integrating Mexico's systems of statistical and geographic information and, in addition, for promoting and orienting the development of informatics in Mexico.

The Conference will open on Tuesday, 1 September with two keynote speakers. One will be Carlos Jarque, President of INEGI, who will speak on statistical developments in Mexico. The other will be Ivan Fellegi, Chief Statistician of Canada, who will speak on statistical developments in Canada.

Geoff Hole (Statistics Canada) chairs the Program Committee whose members are: Paul Cheung (Department of Statistics, Singapore), Cynthia Z.F. Clark (U.S. Bureau of the Census), David Fitch (INCAP, Guatemala), Enrique Ordaz (INEGI, Mexico), Paulo Garonna (ISTAT, Italy), Chris Scott (Consultant, UK), Pedro Silva (IBGE, Brazil).

The theme, Statistics for Economic and Social Development, was chosen to be of interest to both developing and developed countries, and especially Latin American countries. Of course the Program Committee wants the conference to be of interest both to survey statisticians and to those working on official statistics. We also would like to include topics of particular interest to statisticians in Mexico, the host country, and to statisticians in Central and South America.

Call for Papers and Topics

Included in this issue is a 1st Announcement and Call for Papers. Planning of invited paper sessions is still at an early stage. Hence the Program Committee would welcome your suggestions on topics of interest and if possible, your ideas on who might organize a particular suggested topic. Please feel free to get in touch with Geoff Hole or any other member of the Program Committee with your ideas on topics. Abstracts of papers for invited or contributed sessions may also be sent to any one of us.

Local Arrangements

INEGI is located in a modern building with excellent conference facilities in Aguascalientes City with a population of 582,628 people. Close to the conference venue are hotels offering a high standard of accommodation. The average temperature is between 25 to 30 C during the day and the evenings can be cool.

As well as the scientific program there will be an accompanying persons program, to include excursions in Aguascalientes and its environs and to nearby colonial cities, and a series of social events.

Information on the Conference

Bulletin 1 will be available in time for the ISI conference in Istanbul. Those planning to attend should note that the language of the conference is English. The possibility of having some simultaneous translation is being investigated, but due to the cost, it may be limited to translation from English to Spanish for some of the sessions.

The Program Committee and Local Organizing Committee have heeded the plea of Dennis Trewin, President of the IASS, in a recent issue of this Newsletter for use to be made of the IASS Web Site (<http://www.cbs.nl/isi/iass.htm>). Information about the conference is posted there. This site will also point to the three following web sites where information on the conference will be updated as it becomes available in the languages indicated.

English: http://www.statcan.ca/english/iass_iaos/mexico98/
Français: http://www.statcan.ca/francais/aise_aiso/mexique98/
Español: <http://www.inegi.gob.mx>

This information will cover such topics as

- Preliminary Program
- Registration
- Accommodation in Aguascalientes
- Abstracts of Papers Being Presented
- Tourist Information

It is hoped the site will eventually hold all Papers Presented and so constitute electronic proceedings. The latter may be available on CD-ROM. The possibility of having printed proceedings is also being explored.

The Program and Local Organizing Committees hope you will send us your ideas and abstracts and hope to see many of you and your colleagues at the IASS/IAOS Conference in Mexico in 98.

Mario Palma & Geoff Hole

IASS AND IAOS CONFERENCE

STATISTICS FOR ECONOMIC AND SOCIAL DEVELOPMENT

Aguascalientes, Mexico, 1-4 September 1998

1st Announcement and Call for Papers

Plans are well underway for a Joint IASS and IAOS international conference to be hosted by INEGI from 1-4 September 1998, in Aguascalientes, Mexico. The theme, Statistics for Economic and Social Development, was chosen to be of interest to Statisticians in developing and developed countries, and especially Latin American countries.

The IASS/IAOS 98 Conference will feature both invited and contributed papers. The Program Committee welcomes offers of papers for the conference. Bulletin 1 of the IASS/IAOS 98 Conference will be issued in the Fall of 1997. Planning of invited paper sessions is still at an early stage. Hence suggestions on topics of interest and wherever possible the name of a likely organiser would be welcomed by the Program Committee. Some preliminary suggestions for topics follows to provide an idea of the scope of this IASS1/IAOS Conference.

ECONOMY

- Impact of globalization in Latin American countries
- Supplementing GDP as a measure of national progress
- Growth of the information society and its implications for official statistics
- Developing a whole economy price index/problems arising around index numbers
- Sources of data and estimation methods for SNA 93
- Measuring the effects of structural adjustment
- Updating cost-of-living weights in developing countries
- Usefulness and relevance of statistics for business planning
- Leading indicators for developing economics

SOCIAL

- Monitoring poverty and researching other disadvantaged groups in society
- Measuring unpaid labour
- Household consumption/expenditure surveys - approaches in developing and developed countries
- Income - Expenditure Surveys
- Demographic and Health Surveys
- Child Health
- Education
- UN Statistical Commissions Minimum National Social Data Set to monitor targets and support policies and decision making

ENVIRONMENT

- Environmental indicators

META DATA AND DATA ANALYSIS AND PRESENTATION

- Data analysis, data synthesis and data dissemination
- Quality of official data - how should it be measured and how should it be communicated to users?
- Geographical representation of statistical data

SURVEY DESIGN, METHODS & STANDARDS

- Planned innovations for the 2000 Census round
- Methodological problems and experience with the development, use and updating of major classifications used in official statistics
- Meeting the demand for longitudinal data: use of survey and administrative data
- Meeting the demand for data at the level of intermediate geographical subdivisions, in developing countries
- International statistical standards: relevance and challenges
- Recent innovations in business surveys
- Application of advanced information technology in statistical operations: generalised systems
- Sampling in villages
- Possibilities for superior developing world sampling
- Total quality management approach to taking surveys

As well as the scientific program there will be an accompanying persons program, a series of social events.

For further information, please contact:

Mario Palma, Chair of the Local Organising Committee, INEGI, Aguascalientes, Mexico, Fax: 011 39 6 4673 2636, E-mail: mpalma@diapc.inegi.gob.mx or Geoff Hole, Chair of the Program Committee, Social Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6, Canada. Fax: 1-613-951-0653, E-mail: holegjc@statcan.ca

