THE SURVEY STATISTICIAN

No. 34
July 1996

Table of Contents

Letter From the President

Dear IASS Member:

We have been busy with conference activity since I last wrote in *The Survey Statistician*. A very successful conference on Demography Statistics and Privacy and Surveys on Sensitive Subjects was held in Paris late last year. A summary of the conference will appear elsewhere in *The Survey Statistician*. A number of the papers are available in English, and they are shown below. Our Executive Director, Benoit Riandey, was organizer of the conference and might be able to make papers available on request. He is currently working on the proceedings.

I found the papers describing surveys on sensitive topics to be of particular interest. These include topics such as sexual behavior, violence against women, drug use, psychiatric morbidity and homeless people. There is a lot we can learn from each other's experiences.

Unfortunately, the planned Budapest conference on sample surveys in transition countries has had to be canceled because of funding difficulties.

Another big conference, where planning under Gad Nathan's leadership is well advanced, is the conference in Jerusalem on Longitudinal Surveys. It will cover both design and analysis aspects. This conference will be held in September 1997 in Jerusalem immediately after the next ISI Session. A Call for Papers appears elsewhere in *The Survey Statistician*.

The proceedings of the Surveys Measurement and Process Quality Conference held in Bristol in June 1994 have now gone to the printers. It will not be long before you will be able to purchase this very valuable volume on Survey Methods.

We may soon have a World Wide Web site. The ISI has started a site and have offered support to set up a special IASS home page. This will enable us to get news to members more quickly than at present.

One of our most important committees is the Nominations Committee which has the task of developing a slate of candidates to lead the IASS over the 1997-1999 period. Xavier Charoy is the chair of the committee, and I am sure he would welcome any suggestions for the Executive or Council members.

To conclude this note, I would like to say a few words about the future for survey statisticians.

Compared with many sciences, survey methods are reasonably mature and the techniques used in major surveys are not greatly different from those being used several decades ago. Consequently, survey statisticians have extended their research into other methodological areas arising from sample surveys and have had a significant influence. Examples include:

- non-sampling errors and the application of total survey design approaches; small area estimation;
- methodological issues arising from the introduction of technology such as CATI and CAPI;
- analysis, particularly when the data is subject to error; and
- disclosure analysis of microdata.

My one concern is that much of this work has concentrated on population surveys, rather than business surveys. However, the major issues facing official statistical agencies today focus more on economic statistics. These include the reliability of key macroeconomic statistics such as the national accounts and balance of payments, the implementation of major new international standards for both, the concern about the accuracy of price indexes particularly the CPI and the development of reliable estimates of productivity. There are many other problems in economic statistics requiring methodological support.

Should the IASS put relatively more emphasis on economic statistics? I think we should. We did sponsor a very successful conference on business surveys about four years ago but the majority of our activities have concentrated on population and social statistics. This is also true more generally of the "surveys" research community.

One of the reasons may be the reluctance of economic statisticians to involve methodologists. A

real challenge for us is to promote our skills and also to manage the relationships with economic statisticians so there is the trust and confidence required for a good relationship. There are many opportunities for us so it is worth the effort.

Enjoy the rest of the year.

Dennis Trewin
IASS President

.....

The New View on Editing
Leopold Granquist

Introduction

This article concerns the most frequent type of editing procedure in which individual data records are identified by means of edit rules for subsequent manual review. It is a shortened version of the paper by Granquist and Kovar (1996). It follows from a broad international co-operation on editing issues that is taking part under the auspices of the United Nations' Economic Commission for Europe (ECE) in the form of Annual Work Sessions Much of the research underlying this paper is documented in ECE 1994 and 1996.

A new international view on editing has developed during the last five to ten years based on the results of research on editing carried out by statistical agencies. This view, first presented by Linacre (1991), then indirectly presented at a number of international conferences, became more definitively accepted as a common international view at the International Conference on Survey Measurement and Process Quality at Bristol, April 1995.

A change in the approach to editing was necessary because the results of research showed that the heavy costs of editing were not justified by quality improvements. More specifically the reasons for this change are:

- Editing consumes 20 to 40 percent of the total direct survey budget cost. In addition there are substantial costs related to lost opportunities, increased response burden, bad will, delays in timeliness, undue confidence in data quality and respondent reporting capacity, and misuse of personnel.

- The realization that the editing process has not yet resulted in the process being less costly or more efficient.

- During the last five years, efficient methods for prioritizing edits have been developed.

- Technological developments make it possible to move the editing closer to the data source, preferably while the respondent is still available, which opens the possibility of acquiring more accurate data and knowledge about response problems.

Before the advent of computers, editing of individual forms was performed by large groups of clerks, who often had no post school education. As a result, only simple checks could be undertaken, the editing was inconsistent between clerks and over time for the same clerk, and only a small fraction of all errors could be detected. The advent of computers was seen as a means of reviewing all records consistently by applying sophisticated checks requiring computational power so as to detect errors in data that could not be found by means of manual review. Implicitly, the process was designed under the paradigm that more checks and recontacts with the respondents would result in better quality. The early systems in the sixties/seventies had high development, maintenance, and running costs, and did not solve entirely the inconsistency problem. Changes were entered in batch mode and edited again by the computer, resulting in many records passing the computer three or four times and being reviewed by different persons for each pass.

Research and development work became focused on rationalizing the EDP department's work of providing the survey managers with application programs and on developing generalized software. A great break occurred as a direct consequence of the PC revolution in the eighties, when editing could be performed on-line on personal computers, at the data entry stage, during the interview, and by the respondent in CASI modes of data collection. The process was substantially streamlined, but the gains were often used to allow the editors to process more records and to make more contacts with respondents to resolve problems. The paradigm -

the more checks and more contacts the better the quality - was still considered valid. However, evaluations said something else: 10 to 15 percent of the biggest changes made in editing of economic items contribute to around 90 percent of the total change; 5 to 10 percent of the biggest changes brought the estimate within 1 percent of the final global estimate; only 20 to 30 percent of the recontacts resulted in changed values; the quality improvements were either marginal, or negligible; many types of serious systematic errors could not be identified by editing.

Over-Editing

The main reason why editing processes were inefficient is that processes seldom were evaluated, nor were indicators or other performance measures produced. Initially, the approach seemed to be successful. Errors, even serious errors, were detected. Ambitious survey managers and designers thus maintained their confidence in editing as an outstanding tool to achieve high quality data and relied upon editing to fix any mistakes committed in earlier phases of the data collection and processing (including the survey and in particular the questionnaire design). Results from evaluations of editing processes contradicting their perceptions were not considered applicable to their surveys. They believed that investing in more checks and recontacts was beneficial to quality.

However, the opposite may be true and excessive editing may be harmful to quality. For example, editing can cause delays in publishing results and bias estimates when only certain types of errors can be detected. Errors may be introduced when editors manipulate data to pass the edits (so called creative editing) and mistakes or misunderstandings may occur in the recontact process.

Corner Stones of the New View

An absolute and fundamental requirement on editing has always been, and should always be, to identify outliers and those errors in individual data which a user with access to individual data records, but without knowledge of the particular unit, can bet on being in error. Edits aimed at identifying such erroneous data are called fatal edits. In surveys with quantitative data there might be errors which, although not fatal, significantly affect the estimates. Hence, query edits that point to suspicious data items have to be added to the editing process. It is the query type edits that are responsible for the high costs of editing. One reason why editing processes are inefficient is that the distinction between fatal and suspicious errors is not made and accordingly appropriate resources are not allocated to identify and treat each type of error.

Careful Design and Evaluation of the Set of Query Edits

The design of the entire set of query edits is of particular importance in getting an acceptable cost/benefit outcome of editing. The edits have to be coordinated for related items and adapted to the data being edited. Two measures for improving current processes are to relax bounds by replacing subjectively set limits by bounds based on statistics from the data to be edited, and to remove those edits which in general only produce unnecessary flags. It should be noted that there is a substantial dependence between edits and errors for related items. Furthermore, the edits have be targeted on the specific error types of the survey, not on possible errors.

The properties of the edits have to be controlled continuously, for example, by examining data of the outcome of edits. It is an absolute requirement that an editing system should produce statistics on error flags, changes related to edits and reasons for imputation.

Focus on Influential Item Values and Records

The edits and/or the system should have built-in prioritizing rules to focus the review on those suspicious data items or records that may have most influence on the estimates. A leading principle of many methods suggested in the literature is to begin with the worst ones and stop verifying when estimates no longer change.

Another way of prioritizing recontacts is to utilize score functions suggested, for example, by Berthelot and Latouche (1992). The idea is to run the edits as determined by the subject matter experts, assign a score to every record with at least

one flagged item according to the weight of the record, the potential impact of the suspicious item value, and the importance of the flagged items, and then review only those records with a score exceeding a threshold value. When fatal edits are included, then records containing fatal errors have to be handled automatically in cases where the score of that edit does not guarantee the threshold value is exceeded. Numerous studies indicate that applying these methods yield a reduction of the manual review by 50 percent or more without significant impact on quality.

Focus on Response Problems and Error Causes

The main aim of re-contacting respondents should be to collect information on respondent problems, error causes, and respondent reporting capacity. It is essential to look upstream to reduce errors in survey data, rather than attempting to clean up the data at the end, case by case. Thus, respondent data must be of high quality. In business surveys, the respondent must understand the question exactly and the underlying definitions, have the data available in the information system, understand differences in definitions between the survey and the information system, and be able to give an acceptable estimate in case of large differences. Accordingly, the survey design has to be adapted to the respondent's resources and conditions to provide accurate data. Editing has to highlight respondent problems and reporting capacity, not hide them. That can be accomplished by changing the focus of the recontact process from ascertaining whether a suspicious value is wrong to acquiring knowledge of respondent problems and causes of errors. Thus editing can be used to advantage in sharpening survey concepts and definitions and in improving the survey instrument design.

CASI modes of data collection offer an excellent tool of furnishing information about response process and the accuracy of delivered data, provided that edits and error messages are designed to help the respondent in understanding what data are requested. Possible error causes, definitions, and other information needed for answering each particular item should be prompted or provided by the system; the respondent should be allowed to easily give comments to answers; and warnings like "did you include or exclude" specific components should be included. Furthermore, CASI modes provide excellent potential for conducting experiments and/or response analysis surveys.

The New View on Editing (Summary)

Editing should be integrated with, but subordinated to collection, processing and estimation. The main task is to provide a basis for developing measures to prevent errors.

Editing should be considered a part of the total quality process, not constitute the whole quality process. Editing alone cannot detect all errors, and definitely can not fix all mistakes committed in survey design, data collection, and processing.

The paradigm - the more checks and more recontacts, the better the quality - is not valid.

The entire set of the query edits should be designed meticulously to focus on errors influencing the estimates and target existing error types which can be identified by edits. The effects of the edits should be continuously evaluated by analysis of performance measures and other diagnostics, which the process should produce.

Editing has, in priority order, the following roles:

- Identify and collect data on problem areas, error sources and causes of problems in data collection and processing, i.e., it should produce the basic information for the (future) improvement of the survey vehicle
- Provide information about the quality of the data
- Identify and handle important errors and outliers in individual data.

Besides its basic role to eliminate fatal errors in data, editing should highlight, not conceal, serious data collection and processing problems. The focus should be on the cause of an error not on the particular error per se.

For further information contact Leopold Granquist at Statistics Sweden, Research & Development Department, S-115 81 Stockholm, Sweden.

References

Economic Commission for Europe, (1994 and 1996): Statistical Data Editing: Methods and Techniques, Volume No. 1 (1994) and Volume No. 2 to appear 1996, United Nations New York and Geneva.

Granquist, L. (1995), "Improving the Traditional Editing Process," in B. G. Cox, D. A. Binder, N. Chinnappa, A. Christianson, M .J. Colledge and P. S. Kott (eds.) *Business Survey Methods*, New York: Wiley, pp. 385-401.

Granquist, L. and J. G. Kovar (1996), "Editing of Survey Data: How much is enough?" in Survey Measurement and Process Quality L. Lyberg et al. (eds.), New York: Wiley, to appear.

Latouche, M., and J.-M. Berthelot (1992), "Use of a Score Function to Prioritize and Limit Recontacts in Editing Business Surveys," *Journal of Official Statistics*, **8**, pp. 389-440.

Linacre, S. J. (1991), "Approaches to Quality Assurance in the Australian Bureau of Statistics Business Surveys," *Bulletin of the International Statistical Institute: Proceedings of the 48th Session*, Cairo, Egypt, Book 2, pp. 297-321.

.....

USDA'S Continuing Survey of Food Intakes by Individuals and Diet and Health Knowledge Survey 1994-96

Lori Borrud, Sharon Mickle, Rhonda Sebastian, and Martha Berlin

The U.S. Department of Agriculture (USDA) has been conducting nationwide food surveys since the 1930's as one means of fulfilling its responsibility to ensure the health and well-being of Americans through improved nutrition. USDA is currently conducting its 10th national survey, the Continuing Survey of Food Intakes by Individuals (CSFII) 1994-96 and its follow-up Diet and Health Knowledge Survey (DHKS). The CSFII provides detailed benchmark data on the food and nutrient intakes of the general and low-income U.S. populations. The data are used to determine the food choices Americans make and to evaluate the content and adequacy of their diets in relationship to the Dietary Guidelines for Americans and other Federal government statements of dietary policy, such as the Year 2000 Nutrition Objectives. In other food and nutrition-related program and public policy applications, the data are used to assess the nutritional impact of USDA's food assistance programs; to estimate exposure to pesticide residues, food additives, and contaminants; and to develop food fortification, enrichment, and food labeling policies. Other data uses include assessment of the demand for agricultural products and marketing facilities and food and nutrition-related research. The DHKS was the first national survey to provide information on dietary attitudes and knowledge, which can be linked to food and nutrient intakes from the same individuals in the CSFII. The specific focus of the DHKS is obtaining individuals' attitudes and knowledge about the Dietary Guidelines for Americans and on their ability to put the Guidelines into practice. For example, one section of the DHKS covers use and understanding of food labels; another covers food practices related to intake of fat.

The CSFII and DHKS 1994-96 are being conducted by Westat, Inc., a private research firm, under contract to the Agricultural Research Service of USDA. The CSFII and DHKS data collections are spread over a three-year period, beginning in January 1994, with the sample balanced over time to reflect any temporal variations in food intakes. The DHKS collects detailed information on dietary knowledge and attitudes for a subsample of the CSFII adult respondents. The objectives for the CSFII require separate estimates to be made for 40 domains defined by sex, age (10 groups), and income level. The survey includes an oversampling of the low-income population and other groups whose proportion in the population is small, but who are at greatest nutritional risk, specifically young children and the elderly. A multistage area probability sample design, including a large-scale screening of households, is used to produce the sample of individuals for the study. When the survey is completed, there will be over 16,000 CSFII respondents and over 5,000 DHKS respondents.

The objectives of the CSFII are to collect complete food intakes by sampled individuals in sufficient detail so that their intakes of 30 nutrient and other dietary components can be determined and their food choices and habits can be identified. In the past, various 24-hour recall approaches have been

used to collect the CSFII food intake data, including combinations of in-person and telephone interviews and respondent-administered food records. We describe here the methodology of the CSFII 1994-96 with an emphasis on the methods that have been developed to collect data on individuals' food intakes over a 24-hour period. The data collection instruments and procedures are administered through trained professional interviewers, an important advantage for a large-scale national survey of this type.

The CSFII 1994-96 collects food intake data for two nonconsecutive days in order to be able to estimate variability in intakes. For each day, an in-person 24-hour recall method is used, with the respondent being asked to report everything he or she ate the previous day, from midnight to midnight. The two interviews are conducted three to ten days apart, but not on the same day of the week. The range of days separating the interviews was selected to provide the minimum number required to assume statistical independence in dietary intake and the maximum number beyond which respondent cooperation for the second interview was judged to be more difficult to obtain. The interviewer's visits are scheduled to ensure that at least ten percent of the Day 1 recall interview take place on each day of the week since food intake for individuals may differ by day of the week, particularly between weekdays and weekend days. Interviewers are not permitted to make appointments for the recall interview since the respondent's eating behavior might be influenced if it were known that food intake would have to be reported.

A multiple-pass approach to recalling everything eaten and drunk during the 24-hour period was developed for use in the CSFII 1994-96 by the Center for Survey Methods Research (CSMR), US Bureau of the Census, in collaboration with ARS. The CSMR conducted cognitive research to improve the collection of the complete list of foods eaten by the respondent during the previous 24 hours. The previous approach listing the foods in chronological order was replaced with a multi-step approach. Questioning respondents multiple times can generate the recall of more items if the question is asked in different forms or probed in different ways. In the first pass of the multiple-pass approach, the respondents provide a list of all foods eaten the previous day, using any recall strategy they desire. Then interviewers get a more complete list by probing for additions to foods, such as cream in coffee, and by giving respondents the opportunity to recall food items initially forgotten. In a third pass, the interviewers review with the respondents the list of foods they reported to stimulate more reports of foods and eating occasions. Thus, at three points in the recall, the respondents perform different cognitive tasks related to the foods and beverages they consumed and are probed each time for anything they might have forgotten. This recall approach was further developed by ARS and Westat and then tested in a multi-site Pilot Study duplicating the survey design. The Pilot Study provided an opportunity to further refine the data collection methods.

Sufficient detail on food items reported by the respondent must be collected during the dietary interview in order to accurately code the items. Standardized questionnaires, a Food Instruction Booklet (FIB), and measurement aids are used to collect CSFII intake data to the required level of detail. Detailed descriptions of foods reported enable food choices and habits, preparation methods, and nutrient intakes to be tracked over time.

Interviewers use the FIB to probe respondents for a complete description of every food item reported and the amount eaten. The FIB contains the standardized and specialized probes the interviewer is required to ask to obtain the level of detail needed to link food descriptions to a food coding data base. Appropriate probes are organized under 14 food categories and 97 food subcategories. Probes vary with the type of food or beverage being recalled. For example, for cooked cereal, the respondent is asked for the kind of cereal, the brand name, how it was prepared (was it made with milk or water and if milk, was the milk whole, lowfat, or skim; were salt and fat or other ingredients used in the preparation; and was anything added to the cereal after it was cooked). Respondents estimate the amounts of foods eaten using a set of measurement aids. These include standard household measuring cups and spoons; "thickness sticks" used to measure the thickness of slices of meat, poultry, and cheese; a 12-inch ruler; a laminated card containing a set of concentric circles for measuring the diameter of food items such as pancakes and pizza; and pictures portraying a fish fillet and chicken parts. Since the survey protocol specifies that the 24-hour recall be conducted in the

home, the measurement aides can be used to measure the capacity of bowls, cups, and glasses actually used by respondents. Food labels and packaging also are utilized to obtain information on food descriptions and amounts.

The efficiency of processing the food intake data in the CSFII 1994-96 has been facilitated by the use of Survey Net, a computer-assisted food coding and data management system. Survey Net was developed cooperatively by the Agricultural Research Service and the University of Texas-Houston Health Science Center's School of Public Health, and was designed specifically to meet the quality control and data processing needs of the CSFII 1994-96. Survey Net is a multi-level software system, allowing food coders to access three large central data bases. These include the food coding data base (containing food descriptions, food measures, and gram weights of those measures); a recipe data base; and the Survey Nutrient Data Base. The data bases are updated throughout the survey period, allowing data to be processed using current information. The food coding data base contains over 7,200 food codes to which the reported foods can be linked, each code bearing a complete description of the food, and if relevant, the preparation method. In addition, this data base contains over 32,000 weights linked to measures associated with specific foods.

Survey Net has searching capabilities that locate suitable food codes easily, and quickly convert reported quantity information to gram-weight equivalents. Another system feature allows coders to view pre-defined recipes which list ingredients and amounts for every food code in the food coding data base, and to modify recipes to match more closely the foods eaten by respondents. Survey Net then converts the gram weight of the quantity of each food eaten to equivalent amounts of 30 nutrients and other food components. Survey Net also contains a feature that allows coder to enter descriptions of foods and quantities not present in the data base as "unknowns" for later resolution, without halting processing of the intake data.

The CSFII 1994-96 recall data are being collected by about 90 professional interviewers trained in an extensive 7-day in-person training session. Performance is monitored closely, and weekly evaluations and feedback are provided to the interviewers by supervisors. The response results from CSFII 1994 present a successful picture of the procedures and materials developed to conduct the survey. Close to 81 percent of the persons selected to participate completed the Day 1 intake interview and over 77 percent provided food intake recall data for both days.

Information about USDA surveys is available on the World Wide Web at: http://sun.ars.grin.gov/ars/Beltsville/barc/bhnrc/foodsurvey/home.htm. For further information about the surveys, contact the Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, 4700 River Road, Unit 83, Riverdale, Maryland 20737. Telephone (301) 734-8457. Fax (301) 734-5496. E-mail: fsrg@rbhnrc.usda.gov.

.....

European "Demography, Statistics and Privacy" Days: Surveys on Sensitive Subjects
Benoît Riandey

This is a report on a colloquium organized by the Institut National d'Études Démographiques (INED) and held in Paris, France on October 23 and 24, 1995. The IASS gave its support to the *European "Demography, Statistics and Privacy" Days* by sponsoring the second day, devoted to surveys on sensitive subjects.

On the first day, entitled "*Legislation and demographic data collection*," it became clear that European demographers tend to follow one of two data collection traditions. According to the first, the population is always counted by way of a census; in the second, the costly census operation is avoided by utilizing population registers and numerous administrative files. This latter methodology is prevalent in Northern Europe (e.g., Denmark, Netherlands, Sweden and, more recently, Finland).

The laws governing these two statistical traditions are rooted in different philosophies. Take for example the two laws passed in 1978 by Denmark and France. Denmark instituted a computerized national population register and required governments at all levels to use the same individual identification number. France, by contrast, passed a very strict law on *Informatique et Libertés* (electronic data processing and freedoms), which prevented government from using a national

identification number which had been in general use for more than thirty years. The Danish legislature was thus expressing its confidence both in the State and in the professional ethics of government statisticians. The French legislature was expressing doubts about the former, if not the latter!

For its part, the Parliament of the European Union in 1995 adopted a "directive" on *data dissemination and privacy protection*. After difficult negotiations, the member states agreed on this summary document, which makes exceptions for statistics and research. This should make it easier to construct European statistics. This was the wish expressed by the Director General of Eurostat (the European statistics bureau), which sponsored this colloquium. However, in the round table discussion on which this day ended, participants were convinced that national statistical traditions would to some extent be maintained.

This issue also concerns survey statisticians, since the interest that demographers and sociologists show in marginal groups and sensitive behaviors raises real questions regarding the protection of privacy. Depending on the country, statisticians frequently are called to make decisions with respect to collection of sensitive data or else they must obtain quite specific authorizations from the legislature or a watchdog authority. However, the same European regulations apply to all countries, although not to the exclusion of national laws or practices. Consider the following example relating to surveys: a watchdog authority is mandated to verify that the data collected are "*relevant, adequate and not excessive in relation to the purposes for which they were recorded*" [trans.]. This clause seeks to provide considerable protection against possible abuses by government, but it can restrict the information available for administrative statistics. Applied rigidly, it may also lead to unforeseeable decisions regarding surveys conducted for research purposes.

The European regulations identify a list of sensitive variables on which respondents may be questioned only after giving their consent in writing: ethnic or racial origin, religious or political opinions, sexual behavior, etc. The European directive does offer greater flexibility for the collection of statistical information, but it leaves it up to member countries to define the "appropriate guarantees." The question remarks whether the flexibility offered will be beneficial to statistical work in more regulated countries, such as Germany or France.

The German legislation has been reluctant to establish a fixed list of sensitive variables, since the sensitivity of a question varies depending on the context. Sensitivity also varies broadly from one country to another. For example, with reference to medical privilege, physical disabilities are considered in France as a very sensitive matter. This position surprised many Swedish, American, English and Canadian participants. In Canada, such information is even included in the Aboriginal Peoples Survey and the Census of Population. In a similar vein, Howard Meltzer limited his presentation to the issue of studying the prevalence of mental diseases, since physical disabilities did not seem to him to come within the scope of the topic of the day!

Some matters are almost universally considered sensitive, but with very diverse consequences. This is the case with surveys on sexual behavior, with the exception of Finland, with its long-standing tradition of sexology research. In France, the Ministry of Health made this subject its statistical priority in 1990, and the French watchdog authority judged that a survey was necessary, despite its strong reservations as to the intrusive nature of the survey questions. By contrast, the British and American governments withheld public funding for such a survey. On the other hand, Statistics Canada took it upon itself to conduct a groundbreaking telephone survey on violence against women. Many considered it remarkable that a statistical institute would undertake to examine such a sensitive subject. Such an undertaking is risky because of concern that it is ethically intolerable for only partial or whimsical statistics to be disseminated on a sensitive subject, even for a humanitarian purpose.

In France, there was a serious lack of statistical information on the field of immigration prior to the survey conducted by INED in co-operation with INSEE. For years the collection of data regarding nationality was subject to enormous constraints, that are now recognized as excessive. Statistics Canada conducted a survey of aboriginal peoples (Indians, Inuit and Métis), based on a question on ethnic origin included in the Census to address its statistical information needs.

Other survey topics discussed at the colloquium underline the extent to which ethical and methodological issues are linked: poverty measures in Poland, surveys of the homeless, the prevalence of mental disabilities, and surveys on addiction. These survey situations have been described as difficult. Sending out an enumerator who is also a psychiatrist is no trivial procedure. Similarly, infiltrating a network of drug abusers draws on very tentative ethnographic methods. On the other hand, the study of poverty by means of the European panel study of households mainly gave rise to problems of European harmonization or of circulation of banalized data between the member teams of the panel. Restrictive legislation makes such circulation limited and unbalanced.

Beyond the interest that each of these survey topics represents in and of itself, the colloquium brought out common problems that these topics pose owing to their sensitivity: censorship or self-censorship in carrying them out; the difficult trade-off between methodological rigour, the observance of very general regulations on the protection of individuals and the ethical requirement that the respondent be completely informed; but also the use of modern technology to better guarantee confidentiality during the interview; the psychological comfort of both the respondent and the enumerator; elaborate tests of the relevance and the neutrality of the questionnaire; and rigorous research on data reliability.

The colloquium ended on a noteworthy exchange on ethics, in which the openness of society, scientific rigour, loyalty to the respondent, and accountability researcher were debated. Following on the discussions at Amsterdam in 1985 when the International Statistical Institute's declaration of ethics was adopted, these exchanges served to carry the debate forward. This is useful, since the issues involved cannot and should not ever be settled once and for all.

The proceedings of the symposium will be published by INED once the translations into English and French have been completed. For further information, contact Benoît Riandey, INED, 27 rue du Commandeur, 75675 Paris cedex 14, France.

.....

Statistics in Transition
An International Statistical Journal in English

Jan Kordos[1]

Statistics in Transition is the Journal of the Polish Statistical Association and is published by the Central Statistical Office of Poland. The Journal is expected to provide a forum for an exchange of ideas and experiences in various fields of statistics. It publishes articles on statistical methods and their application in investigating social and economic processes, especially those relevant to the economies in transition. The journal publishes also papers presented at conferences and seminars if they are important for the development of statistics. The journal publishes book reviews and reports, usually from international meetings connected with statistics in transition countries. It intends to inform about the activities of statistical societies in different countries.

The Journal is published twice a year: in June and December, but also "Special Issues" are published (they are distributed free of charge) if we have sponsors and appropriate materials connected with statistics in transition countries.

Why We Need a New International Statistical Journal

During the last five years, a number of international conferences, seminars, and workshops were devoted to different statistical issues in Central and Eastern European and the former Soviet Union countries. An important goal of those meetings was to bring together representatives from different countries, producers, and users of statistical data, both from developed countries and from the countries in transition. All of us can learn useful lessons from one another. However, only a few statisticians from the countries in transition could attend those international meetings, and proceedings from those meetings were usually published very late, or were not published at all.

I attended some of those international meetings and usually found them interesting. I also found that it

---

[1] Prof. Jan Kordos is the Editor of the International Statistical Journal in English *STATISTICS IN TRANSITION*. His address is as follows: Al Niepodleglosci 208, 00-925 Warsaw, Poland.

was not easy for statisticians from developed countries to understand statistical issues in transition countries. There are different reasons for that that do not need to be discussed here. As a result, I understood that statisticians from transition countries need a special international statistical journal devoted mainly to statistics in transition countries. However, such a journal should be edited not only by statisticians from the countries in transition, but also from other countries willing to co-operate with the countries in transition. I have discussed those problems with statisticians from different countries and we decided to establish a new international statistical journal in English. We came to the conclusion that it should not be only "another international statistical journal" but it should also be mainly devoted to statistical issues in transition countries. Statisticians from other countries also have a chance to publish their articles in this journal.

Editorial Board and Associate Editors of the Journal

The Editorial Board of the Journal is an international one. It consists of statisticians from the countries in transition to market economies (Bulgaria, Czech Republic, Estonia, Hungary, Latvia Lithuania, Poland, Romania, Russian Federation, Slovak Republic, Slovenia, Ukraine) and from other countries (Canada, India, Italy, Japan, Mexico, the United Kingdom, United States).

First Two Volumes of the Journal

The following assumptions have been accepted while editing the journal: each regular issue published in June or December will be devoted mainly to selected topic but also other articles sent to the Editorial Board will be published. Proceedings of international conferences, seminars and workshops devoted to statistics in transition countries will be published as "Special Issues".

The first issue of the journal was published in June 1993 and was devoted mainly to Polish statistics which celebrated the 200th anniversary of its existence and the 75th anniversary of the establishment of the Polish Central Statistical Office (GUS) in 1993. The second issue, published in December 1993, was mainly devoted to papers presented at a meeting during the 49th ISI Session on the "Special Issues in Transition Countries".

The first "Special Issue" was published in March 1994, and was mainly devoted to the results of the meeting on the "*Round Table on Fundamental Principles of Official Statistics and Related Legal and Deontological Aspects of Statistics in the Countries in Transition*" held near Warsaw, Poland, in September 1993. The June 1994 issue was mainly devoted to *the Labour Force Surveys (LFS) methodology*. The methodology of LFS in five countries in transition (the Czech Republic, Hungary, Poland, Romania and the Slovak Republic) was published. Experts with a long tradition in the LFS methodology commented on the papers. Papers on the LFS methodology in Canada and the USA were also published.

The next issue of the journal, published in September 1994, was a "Special Issue" mainly devoted to some *aspects of poverty measurement in the transition countries* and Poland was considered as a case study. The papers included in this issue were presented at the international conference held in Warsaw in July 1994.

The December 1994 issue, which closed the first volume of the journal, was mainly devoted to *small area estimation*. Three other articles were also published as well as book reviews and reports, among others, a report on the International Conference on Statistics to Commemorate the 100th Anniversary of Professor Jerzy Neyman's Birthday, held near Warsaw in November 25-26, 1994.

The second volume opened with a "Special Issue" published in March 1995, and was mainly devoted to the *ILO International Conference: Restructuring of Labour Statistics in transition Countries: First results and Emerging Challenges* held in Minsk, August 31-September 2, 1994. The June 1995 issue was devoted mainly classifications and analysis methods. Next two "Special Issues" were devoted to a Workshop on National Accounts (on Implementation of SNA/ESA in transition countries), and to time use surveys (selected papers presented at the International Scientific Conference on *Methodological Issues of Time Use Surveys: Design and Analysis* which was held near Warsaw, Poland, from 31st May to 2nd June 1995). The

December 1995 issue was devoted to nonresponse issues in statistical surveys.

The June 1996 issue is mainly devoted to business surveys methodology and some problems of statistics in transition countries. New issues are in preparation. They will be devoted to the methodology of household budget surveys, panel surveys, some methodological aspects of sampling surveys, agricultural surveys, and some aspects of statistical analysis. We expect also to publish some "Special Issues" connected with statistics development in transition countries.

.....

Question/Answer

As in the previous issues, we continue to use selected questions from the IASS workshops in Cairo and Padua. While good questions remain which may suffice for a couple of more issues of *The Survey Statistician*, it is time to renew the request to readers for further interesting and practical questions. Please send questions and observations on this column to Prof. Vijay VERMA, ESRC Research Centre, University of Essex, Colchester CO4 3SQ, U.K. Fax: 44 1206 873 151. E-mail: vjverma@essex.ac.uk.

Q34.1 Serious coverage errors occur in the use of area frames when the boundaries of the enumeration areas (EAs) cannot be clearly identified during field work. What would you suggest to address this problem in a survey?

This is quite a common problem in surveys based on outdated area frames, or when the available maps and descriptions are of inadequate quality. My first advice is to try and isolate the problem. It often happens that the problem with area boundaries is serious only in some domains. For instance, census EAs in rural areas may correspond to individual localities and therefore have relatively easily identifiable boundaries, and in the main the problem may exist only in the more densely populated (and more rapidly changing) urban areas.

Secondly, rely on all the information available, not just the maps. Often census maps are accompanied by helpful descriptions. More importantly, census household lists, even if outdated for actually selecting a sample of households, can be extremely useful in defining and identifying the original boundaries of census areas.

Next, we can look for different, larger, area units as PSUs in place of individual EAs, units which are likely to have less serious boundary problems. From a given frame of basic area units such as EAs, larger units can be created in two ways: by selecting EAs in contiguous groups rather than individually from the EA list; or by introducing a higher stage of selection such as taking whole towns or villages, each composed of a number of EAs, as the PSUs. The larger areas can then be subsampled as appropriate. For instance, the selected units may be segmented and one or more segments selected per area in an additional sampling stage, or the whole area may be listed for the direct selection of households.

In exceptional circumstances, when no practical solution could be found and the problematic domain was only a small part of the total population, I have resorted to approximate solutions. For instance, in a country no suitable census EA or other area frame was available in the domain of small/medium towns, which accounted for around 8 percent of the total national population. However, reasonably reliable information was available on the current population of each town. A sample of towns was selected with probability proportional to the current population, and within each selected town a fixed number of locations (points) were selected at random using the best available maps. An area of appropriate size was demarcated around each selected point, using more or less arbitrary but clearly identifiable boundaries. This was followed by listing and selection of a fixed number of households from each such area.

Q34.2 What would be the best way of integrating a general household survey with an agricultural survey (of households engaged in agriculture), or with an establishment survey (of households operating small-scale establishments)?

This is a very broad and complex question, and the answer depends on specific circumstances. Perhaps, it can generate a continuing discussion in this Column. For the moment, let me take on some particular aspects.

Consider a survey of small-scale, largely household-based establishments covering diverse non-agricultural sectors. In terms of the sample design requirements and mode of data collection etc., such a survey has many features in common with a typical household survey. Indeed, often the most appropriate way to survey small-scale establishments is through a sample of households. Nevertheless, the incorporation of such an establishment survey within the framework of a general household survey program must take into consideration some important differences between the two.

As survey units, small-scale establishments tend to be less stable than typical households. While there is likely to be a one-to-one correspondence between household based economic units and households where the former exist, many households will have no units eligible for the establishment survey. Small-scale economic activity, though widely dispersed, is often far from being uniformly distributed across the population, and the sample design must take into consideration the patterns of concentration, which can differ from one type (sector) of activity to another. These can also differ by size and other characteristics of the units. The units involved are heterogeneous by virtue of differences in the type or sector of activity, size, location, type of premises, criteria of eligibility for inclusion in the survey etc. Each type of units may have its own sample size requirements. The content and form of the information required may also differ to some extent for different types of units.

The integration of an establishment survey with a household survey of the general population requires incorporation of special requirements of the former into a compromise design. Let me describe this in relation to one aspect of stratification of area units. Apart from usual reasons for stratification, stratification has a very specific additional objective in an establishment survey: namely, to identify and separate out different patterns and degrees of concentration of different types of units. Sampling procedures and rates can then be varied across different types of strata to improve efficiency of the design and meet specific precision requirements of particular sectors or types of establishments. Consider a typical multi-stage design, involving the selection of area units (in one or more stages) followed by the listing and selection of ultimate units (households, establishments) within each sample area. Suppose that, apart from population-based information on the numbers of persons and households, the frame for the selection of area units also includes information (from a past economic census, for instance) on the numbers of establishments of different types. On the basis of this information, we can identify the type of establishments (j) which predominates within each area unit. This may be done, for instance, in terms of a relative measure such as the following:

R_j = ratio of the number of type j establishment in the area, to the average number of such establishments per area in the population.

A 'stratum of concentration' can be formed for each type, j, of establishments by grouping together areas in which the R corresponding to that type has the largest value. Sampling procedures and rates can then be varied as required from one such stratum to another, depending on the type of establishments concentrated in the stratum concerned.

Of course, being defined in terms of sample areas rather than individual establishments, any stratum will contain establishments of different types, and only a proportion (p) of establishments of a given type will be 'captured' in its stratum of concentration. (Changes in the population since the preparation of the area frame also have the same effect.) The effectiveness of the stratification depends on the proportion of units of each type j 'captured' in the corresponding stratum of concentration. The higher this proportion, the more effectively can the sampling of units of a given type be controlled by choosing an appropriate design for the corresponding stratum of concentration.

In practice, various refinements can be introduced in the above scheme to improve its effectiveness. For instance, often the proportion captured can be improved by further stratification taking into account not only the first but also the second highest R_j value within each area. Incidentally, in a large-scale application, I also found the method to work better when the mean in the denominator of R_j is computed after excluding 'empty areas', i.e. areas containing no establishment of the type concerned.

.....

Country Reports

AUSTRALIA (from Susan Linacre)

The Australian Bureau of Statistics is conducting a national **Women's Safety Survey** in 1996. It will collect information from 6,000 women about fears for and threats against their personal safety; experience of physical and sexual assault; actions taken after experiencing violence; and the effects on women of violence. Information is collected about violence against women by both men and women. A survey reference group, including representatives from relevant government departments, crime research agencies, women's services and academics, has advised the ABS on the information collected and the methodology used. Given the sensitive nature of the subject matter, ABS interviewers working on the survey have received special training. Procedures have also been put in place to ensure the safety of women participating in the survey. Results are planned for release in December 1996. For more information contact Marion McEwin, ABS, P.O. Box 10, Belconnen ACT Australia 2616. Tel. 61-6-252-7068, Fax 61-6-252-5172.

Relationship between survey design and analysis of time series. For the past couple of years, the ABS has been exploring the relationship between the design of sample surveys and the time series characteristics of survey based estimates. So far, the investigations have traversed: the effect of sample rotation on serial correlations and hence on the variances of seasonally adjusted and trend series derived from survey estimates; how to inform the X11 seasonal adjustment package about design-induced time series features; and how to "tune" the filters used to derive current trend rates of growth by taking account of sample-induced and other variability. For more information, contact Andrew Sutcliffe, tel. 61-6-252 7646; e-mail andrew.sutcliffe@abs.telememo.au.

The Australian Bureau of Statistics is assisting on a consultancy basis in the application of the **Delphi surveying technique** in the Australian maritime industry. The technique involves two approaches to respondents, the first to obtain expert opinions on the likelihood, benefits and constraints of a development occurring in the future; the second again seeking opinions from experts but this time with knowledge of the aggregated opinions from round one. The objective of the technique is to achieve a consensus' opinion. Similar Delphi surveys to determine the future directions of industries have been undertaken in Japan, Germany and the United Kingdom. For more details, please contact Gemma Van Halderen, P.O. Box 10, Belconnen ACT Australia 2616. Tel. 61-6-252 6302, Fax 61-6-253 1033.

The Queensland Government Statistician's Office has developed a SAS system for automated **survey management** of CATI, mail and mixed mode surveys. The system also has automatic dialing and remote listening and monitoring capabilities. A codeless questionnaire generator is currently being built. The system was successfully used to conduct the Queensland Migration Survey based on a random digit dial sample of 110,000 numbers. For further information, contact Ranjit de Mel, GSO, P.O. Box 37, Albert Street, Brisbane 4002, Australia. Fax 61-7 3221 9516.

BASQUE COUNTRY - SPAIN (from A. Iztueta Azkue)

In 1995 EUSTAT-Basque Statistics Institute began a general reflection process to define clearly the objectives and critical success factors for the Institute for the period 1995-1998, and to identify the necessary actions to reach its targets in this period.

This process has counted on a wide participation of the personnel of EUSTAT, as well as on the collaboration of some outstanding economic and social analysts of the Autonomous Community of Euskadi, from the academic, business and government sectors as well as the mass media.

The process was carried out in a typical TQM (Total Quality Management) framework, which has as a high-priority target attention to the different types of customers of EUSTAT: University, Public Administration, Enterprises, Unions and society as a whole.

As a result of this process, there are at present four large action programs underway, each one of them including a variety of activities, as well as 12 immediate actions aimed at improving a series of very concrete aspects of the work of the Institute, both in production and in the area of internal human

relationships. The four large action programs are the following:

1. Definition of the general framework of EUSTAT: institutional position, internal organization, self-financing, etc.

2. Statistics dissemination: analysis of supply and demand, redesign of products and of the forms of dissemination, information service, statistics data bank, etc.

3. Foreign action: coordination, cooperation and exchange with other international and state statistics agencies, marketing, etc.

4. Production and quality: automation of the statistical production processes, quality control of the productions, etc.

On the occasion of the XIII INTERNATIONAL STATISTICAL SEMINAR in 1995, EUSTAT organized three courses. The relevant publications are available to those interested.

I. ESSPROS (European System of Integrated Social Protection Statistics) by Mr. Alfonso Barrada, Mr. Cor N. Gorter and Mrs. Mercedes Alcalde.

II. Capture-Recapture Models by Dr. Kenneth H. Pollock.

III. Residents registration systems and Statistics by P. Myrskyld and W. Mohr.

For more details about this report, please contact: EUSTAT-Basque Statistics Institute (Anjeles Iztueta Azkue)Dato, 14-16, 01005 Vitoria-Gasteiz (Spain).

CANADA (from Gordon Brackstone)

The **1996 Census of Agriculture** (CEAG) features several innovations in data collection, editing, and quality assurance taking advantage of several generalized systems developed at Statistics Canada. The Census of Agriculture is conducted in conjunction with the Census of Population which is testing a new centralized edit collection methodology in Eastern Ontario with a view of using this approach for the 2001 Census. The data collection and capture operation was redeveloped using Statistics Canada's generalized data collection and capture system (DC2), taking advantage of on-screen edits and of sample key verification. The editing rules were both simplified and reduced. The Farm Register data files will be updated with the 1996 CEAG farm data using probabilistic linkage routines in the Generalized Record Linkage System (GRLS). This will translate into significant improvements over past methodologies in both quality and efficiency. Geographic editing has been reworked to take advantage of Farm Register data automatically when possible. Imaging technology will also be used for the first time to replace the library function and will simplify and accelerate most manual operations, in particular the work of validators. Applications of other generalized systems such as the Generalized Edit and Imputation System (GEIS) and Automated Coding Text Recognition (ACTR), introduced in the 1991 Census, were reviewed and optimized. For more information on this project, please contact Pierre Daoust or Serge Legault, Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6. Tel: (613) 951-4755 or 951-1460.

The **Generalized Estimation System** (GES) developed at Statistics Canada can now meet the estimation requirements for a large number of surveys. The system can be used for estimation of means, ratios and totals for arbitrary domains, making use of auxiliary data through the theory of Generalized Regression Estimation or Calibration Estimation (including weight constraints). Variance estimates are available using either linearization or jackknife methods. The GES can be applied to surveys with stratified random samples, PPS sampling and cluster sampling.

The system may be used interactively. It has been developed to have compatibility between DOS-based and UNIX-based computers. This system is part of the suite of software products supported by Statistics Canada. For more information, please contact Francine Hardy, Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6. Tel: (613) 951-1709.

CENTRAL AMERICA AND PANAMA (from David Fitch)

A small sampling experiment was undertaken jointly by the Ministry of Health, Department of El Quiché, Guatemala and Instituto de Nutricion de Centro America y Panama (INCAP). Mothers were interviewed in some 30 clusters about their practices and beliefs with regard to having their children vaccinated. Mothers had records showing only 39% of their children had been fully vaccinated. The two methods used were, (1) the non-probability EPI method (Expanded Programme on Immunization) promoted by the WHO in which a random house in a random direction is selected as the first house, with subsequent houses being the one closest to the last, and (2) a systematic procedure facilitated by the use of a hand-held computer that selects one house at random out of each interval, the interval having been computed to give the desired probability within the cluster.

Efforts and methods to improve vaccination coverage are of great importance. A large number of children die in rural indigenous areas like those of the Department of El Quiché of diseases which are preventable by vaccination. Measles killed thousands there in 1990.

INCAP's hand-held computer method, when controlled for sample size, showed somewhat less variances due likely to spreading the sample over each village and permitting the larger villages and the 6 zones of the Department's capital town to be used as separate strata.

The study itself is perhaps not that sophisticated, but could have important consequences. The Department's health Director, already with doubts about the EPI method through study of an earlier INCAP report, plans to report their research in a meeting to which he is inviting officials from UNICEF, WHO, USAID, and their contractors.

FINLAND (from Kari Djerf)

The General Assembly of the International Statistical Institute decided at the Beijing Session that the **52nd Session of the ISI** will be organized in Helsinki in 1999. The preparations for the Session are successfully proceeding under the direction of the ISI Headquarters. The recently elected local bodies represent both the scientific community, official statistics, and governmental and other relevant sources and users of statistics. Mr. Timo Relander, Director General of Statistics Finland, was elected as the Chairman of the Council of the National Organizing Committee, and Professor Hannu Niemi, University of Helsinki, as the President of the Local Programme Committee. The Session provides a great challenge for the Finnish statistical community and we expect to meet all the readers of Survey Statistician at the famous Finlandia Hall in Helsinki in August 1999. For further information, please contact Mr. Kari Djerf, e-mail: kari.djerf@stat.fi.

The **16th Nordic Conference of Mathematical Statistics** will be held on August 18-21, 1996 in Lahti. The Finnish Statistical Society and the Department of Statistics of the University of Helsinki are the principal organizers. The conference covers a range of topics relevant for survey statisticians such as Markov chain Monte Carlo methods (Julian Besag, University of Washington) and multilevel modeling (Harvey Goldstein, University of London). Contributed Papers Sessions also contain interesting topics, e.g. sampling, survey methodology and quality control. Participation is open to statisticians from all countries. For further information, please contact the Department of Statistics, P.O. Box 54, FIN-00014 University of Helsinki, E-mail:ncms@noppa.helsinki.fi.

Statistics Finland has recently been a host to many international meetings. Two of those arranged in autumn 1995 deserve closer attention: the Blaise Users' Group Conference and the Sixth Workshop on Household Survey Non-response. The proceedings of the Blaise conference were published already. The proceedings of the non-response workshop presents some very recent work and will be available in June. The next conference, **International Conference on Comparative Analysis of Firm Data,** will start on June 17 and last for three days. Those interested in analyzing enterprise panels, innovations, wage formations from data on firms, etc., please contact the organizers at The Midnight Sun Conference, P.O. Box 3 A, FIN-00022, Statistics Finland to obtain additional information or directly from the home page: http://www.stat.fi/sf/home.html.

A **new MSc program in statistical systems** was launched last autumn at the University of Jyväskylä. Its primary goal is to provide the students with sound theoretical knowledge and practical skills on data collection, analysis and dissemination. Students will get a good overview of both official statistics and decision making. The program consists of the following modules: a) theoretical statistics, b) theory of statistical systems, c) software systems and databases, d) organizations and their management, and e) practical work and MSc thesis. Each module has courses of which about half are obligatory, including courses on sampling theory and survey design. The elective courses can be chosen in such a way that they support working practice and the subsequent writing of a thesis. For further information, see home page: http://www.jyu.fi/list.html and choose **New MSc Programs in Information Technologies.**

FRANCE (from Benoit Riandey)

The findings of two interesting retrospective surveys were published in 1995. The methodology of these surveys distinguishes them from the strong Anglo-Saxon tradition of panel studies; in retrospective surveys, data collection, while economical, is subject to errors of memory and also to a selection effect due to mortality and emigration.

The survey entitled **"Jeunes"** (young people) describes the phenomenon of young persons taking longer to become independent: in an era when unemployment at the start of working life is high, young people are tending to postpone certain key events: leaving home, forming relationships and marrying. The survey shows that these stages in life are being delayed and that the events themselves are becoming blurred. Formerly easy to pinpoint in time, these events have become slow and reversible processes: for three years, the first jobs are precarious, with numerous periods of unemployment. Returns to the parental home are numerous, and not only for students. Going back to school has become a more frequent practice.

This survey, conducted in a single wave by the Institut Statistique (INSEE), focused on 10,000 young persons aged 18 to 29 selected from the sample for the employment survey (Enquête sur la force de travail - labour force survey). The respondents describe their experience since age 16: education, place of residence, employment, relationships. As a result of remarkable co-operation among various research institutes, co-ordinated by Monique Méron (INSEE), the survey brought together researchers who had worked with other questionnaire methodologies (those of CEREQ, CNRS, INED). The INSEE periodical *Economie et Statistique* published nine articles on this survey in its July 1995 issue (No 283-284) and will soon bring out a second issue on this most troubling subject.

For further information, contact Monique Méron, INSEE, 18 Boulevard A. Pinard, 75675 Paris cedex 14, France.

The survey **"Mobilité géographique et insertion sociale"** (geographic mobility and social integration) deals with immigrants' life histories after their arrival in France, as well as those of their French-born children. This survey examines a topic tainted by a number of false notions and controversy, owing to a former lack of statistical data which the survey serves to correct. In tackling this topic, it has clarified the underlying concepts: in a country in which acquisition of citizenship was easy, statistics relating solely to the foreign population tend to give a skewed picture of personal histories, since immigrants who have become French citizens exhibit specific behavior patterns (ones that are "more French" than those of immigrants who have maintained their foreign status). In addition, the immigrant population is highly heterogeneous, owing to the large number of migratory flows from different countries in recent decades. This attention to underlying concepts has profoundly shaped the methods used in conducting the survey and subsequent analyses. For example, the analyses differentiate immigrants by country of birth. For this purpose, the sample, drawn from the census of population, distinguishes seven subsamples of immigrants from different countries, three subsamples of their French-born children, and a control sample of persons from non-immigrant backgrounds. There is no subsample of recent immigrants, since observing the latter would not yet be of great biographical interest. When youth from immigrant backgrounds are compared to young persons in the preceding survey, this will answer important questions for social cohesion in the coming years.

This survey of the Institut National d'Etudes Démographiques (INED) was carried out with the co-operation of the Institut de Statistique (INSEE), and with the assistance of a network of interpreters for some of the interviews.

The survey findings are reported in *Faire France* by Michéle Tribalat, Editions La Découverte, 9 rue Abel Hovelacque, 75013 Paris, France.

For further information, contact Michéle Tribalat, INED, 27 rue du Commandeur, 75675 Paris cedex 14, France.FRANCE (par Benoit Riandey)

ITALY (from Claudio Quintano)

The Italian National Institute of Statistics (ISTAT) is undertaking a new experimental Family Budget Survey based on a sample of about 12,000 families. The principal aims of this experimentation are to control the field-work and reduce non-sampling errors.

The project has four main aspects:

1. About 3,000 of the 12,000 families are going to be reinterviewed by the CATI method in order to check interviewers' work.

2. To reduce recording errors the BLAISE package interactive procedures will be used.

3. A new sample of 2,000 families will be interviewed by professional interviewers from a private company in order to test a modified version of the current questionnaire in which new individual income information is requested, and to evaluate the interviewers' performance.

4. About 300 of the 2,000 families will be interviewed by a questionnaire developed within BLAISE and using the CAPI method.

For further information, contact Giuliana Coccia, FAM Service, ISTAT, via Ravà 150 - 00142 Rome, Italy. Tel. 39-0-6-5943009, Fax 39-0-6-5430660.

JAPAN (from Chikio Hayashi)

The "Survey on Japanese Life" has been conducted every year since 1958 by the Prime Minister's Office. Its purpose is to assess the actual condition and future trends of basic Japanese consciousness in the quality of life. The survey method and questionnaire have remained almost the same from the beginning.

The survey method is face-to-face interviews with 10,000 males and females over 20 years of age throughout Japan. The response rate is between 70% and 80%. The sampling method is a two-stage stratified probability sampling. At the first stage, census enumeration districts are stratified by geographic region and urban city level and a sample of districts is selected. At the second stage, potential respondents are selected at random from each sampling unit using an equal interval method.

The questionnaire covers life satisfaction, life style, prospect of the future life, and demographic and socio-economic status. The special focus of the latest survey was "A Comfortable Life in Japan".

The results of the survey indicate changes in the mainstream of life style. The mainstream had shifted to a culture-oriented life style from a materialistic life style in the early 1980's. A majority of Japanese have come to attach more importance to leisure and leisure activities. Since 1983, leisure and leisure activities have been the top priority for the future life in the survey. Trends in Japanese perception and concerns provide basic information for use in government policy. Thus, the "Survey on Japanese Life" is a most important source of sociological data in Japan.

For further information, contact Akio Konsei, Research Department, Central Research Services, Inc., 3-14-13 Higashi-Gotanda, Shinagawa-ku, Tokyo, 141, Japan. Tel.: 81-3-3449-4511; Fax: 81-3-3449-4516.

LATVIA (from Janis Lapins)

Interest in the probability sampling approach has grown rapidly in Latvia in recent years. In 1995, probability sampling was used for the first time in five surveys carried out by the Central Statistical Bureau (CSB) of Latvia: a survey of small enterprises, a survey of retail shops, an agricultural survey, a household budget survey, and a survey of the labour force.

The main objective of the new **Household Budget Survey** (HBS) of Latvia is to estimate the total value as well as the structure of households' income, expenditure and consumption. The survey started in September 1995 and is a continuous survey. The results will be published quarterly. The survey is prepared by the staff of the Department of Social Statistics of the CSB in close co-operation with a team of experts from the World Bank headed by Prof. Jan Kordos. Two survey instruments are used to collect data: the Household Questionnaire (HQ) and the Diary. The HQ consists of two parts -- the Introductory Interview and the Final Interview. The Introductory Interview is used to collect data on the socio-demographic characteristics of the household. The Final Interview collects data on household income, possession of durable goods, and access to public facilities. The Diary consists of a two-week daily recording form, administered twice a month, covering expenditures, consumption, and farming activities. Households are paid 3 Lats (about 6 USD) at the end of the final interview for their participation in the survey. The total annual sample size is equal to 7992 households (666 per month with a complete monthly renewal till December 1996). Half of each month's sample for the year 1996 will be used as a panel in the sample of the corresponding month for the next 3 year period.

The sampling procedure differs in urban and rural areas. In urban areas stratified one or two stage random selection of persons aged 16 and more from the population register is made. Towns are used as the PSU. PSUs are selected with probabilities proportional to the population size. Simple random sampling of persons aged 16 and more is carried out within PSU at the second stage. The households where selected persons are living formed the sample for the HBS in urban areas. In rural Latvia stratified two stage random sampling of households is used. The rural area is divided into 466 PSUs (the smallest administrative units or their union). The selection of PSUs is made with probabilities proportional to the number of households. Simple random sampling of households is carried out within PSUs at the second stage. For field work and data processing the techniques of Living Standards Measurement Study surveys developed by the World Bank are adopted to increase quality and overall efficiency of the HBS. There are 79 interviewers each surveying, on average, either 7 households in rural PSUs or 9 households in urban PSUs. The field work is organized and controlled by 12 full-time and 8 half-time supervisors. Data entry is organized in Riga and 14 regional statistical offices. Horvitz-Thompson type estimators are used.

For more information, contact Dr. Janis Lapins, Central Statistical Bureau of Latvia, Lacplesa 1, LV1301, Riga, Latvia; Fax: 371-78 30137.

NEW ZEALAND (from Len Cook)

1996 is the year of several key initiatives for Statistics New Zealand. Many are associated, either directly or indirectly, with the Census of Population and Dwellings, the 30th population census in New Zealand.

Data Capture by Imaging: For the first time, the data capture phase of census processing is being carried out using imaging technology. About 1.2 million dwelling and 3.5 million personal double-sided A3 forms are to be scanned, and an electronic image taken of each document using ICR (Intelligent Character Recognition) for ticks and numeric answers (approximately 80% of Census data). At the subsequent code and edit phase alphabetic answers will be captured by conventional key entry. This, and all data editing done at this stage, will use the questionnaire images. Significant cost and time savings are anticipated, with provisional census data expected to be released up to three months earlier than in the 1991 Census.
[Roy Middleton, Statistics New Zealand, tel.: 64-3-374 8764]

Questionnaire development using cognitive research: In previous years census forms and individual questions were evaluated by a series of large scale pretests and pilot tests. This produced collections of completed forms which were then analyzed for problems like item non-response and visible errors. For the 1996 Census, questionnaire development relied mostly on cognitive testing with small groups of people, yielding qualitative rather than quantitative information. Cognitive testing aims to understand how people carry out tasks such as answering questions and completing forms, and involves direct observation of people in these tasks. These observations gave insights into respondent interpretation and reaction to individual

questions and the reasons for failure to follow 'skip' instructions (common routing errors). This method of evaluation allowed for more flexibility in the development timetable, with continuous refining of forms and purposive testing on selected population groups for particular questions.
[Kim Saffron, Statistics New Zealand, tel.: 64-4-495 4712]

Bilingual (Maori-English forms): Although census forms in Maori (the language of the indigenous people of New Zealand) were used earlier this century, these had always been for separate censuses of the Maori population. In 1996, for the first time the English language census forms have parallel Maori language forms, asking the same questions and using the same layout. The bilingual forms, not issued generally but available on request, comprise the English form inserted inside its mirror-image Maori form in such a way that the translation of each question is easily identifiable on the facing page. As only the English forms are designed to be electronically scanned, responses on Maori forms will be transcribed to the English form, which is then detached for scanning.
[Janette Briggs, Statistics New Zealand, tel.: 64-4-495 4743]

The Post Enumeration Survey: New Zealand's first Post Enumeration Survey is being held two weeks after the census to estimate the accuracy and completeness of the census count. Sample selection and data collection methods are independent of the Census. Approximately 12,250 households (permanent private dwellings) are being interviewed about their completion of Census forms. Dwelling address, and name, age, sex and ethnic group of household members will be used for matching and searching purposes, and to produce estimates of the level of coverage of the 1996 Census. Post Enumeration Survey forms will not be imaged, but will be matched with the imaged census forms. [Judith Marbeck, Statistics New Zealand, tel.: 64-4-495 4650]

The Disability Survey: Two to three months after the Census, New Zealand's first comprehensive survey of people with disabilities is being carried out. Approximately 25,000 adults and children in private households throughout the country will be interviewed. The sampling frame is the census database, using responses to the two disability filter questions to select respondents who identify as having a disability. A number of those who respond that they do not have a disability will also be selected for interview. The survey will provide estimates of the number of people with disabilities and their demographic characteristics, the nature, duration, severity and origin of the disability, barriers faced in daily life, costs incurred because of disabilities and their need for technical aids and support services. Questions cover a range of disabilities, both physical and psychological or intellectual.
[Jenny Sanders, Statistics New Zealand, tel.: 64-4-495 4629]

REPUBLIC OF KOREA (from Seung-Kon Lim)

As the statistical environment in Korea progresses rapidly, the existing Statistics Act, last changed in 1975, is again being revised for the following purposes: to simplify the approval process for statistical compilations; to strengthen the protection of personal information provided in the course of statistical activities; and to reinforce the efficient utilization of statistical methods, resulting in systematic development of the national statistical regime and the improvement of statistical operations. The new Statistics Act will be effective from April 1996.

Main aspects of the new Statistics Act. Under the existing Act, even statistics compiled under other Laws require approval of the Commissioner of the National Statistical Office. However, under the new Act, only items such as survey methods, as referred to in the Presidential Decree, along with items not stipulated by the Laws, require consultation with the Commissioner of the National Statistical Office (Article 9).

As a result of Article 11 of the new Act, the Commissioner of the National Statistical Office may conduct education on statistical matters, or recommend the head of a statistical agency to conduct it, in order to strengthen the ability of persons engaged in the statistical work of a statistical agency.

In order to protect confidential information of individuals, legal persons or organizations, the penalty for disclosing confidential information has been increased to three years from the previous one year imprisonment (Articles 14 and 23).

For statistics with high reliability, the Commissioner of the National Statistical Office may waive the requirement for consultation with him prior to publication, resulting in simplification of the process of statistical compilation (Article 15(2)).

In order to improve the efficiency of the implementation of the new Statistics Act, the penalty for certain minor violations such as failure to present material, submission of false material, interference with the submission of material, and refusal of an investigation conducted by a person engaged in statistical work was reduced to the fine for negligence rather than imprisonment.

For further information, contact Sang-Sik Kim, National Statistical Office, Republic of Korea, Hanta Bldg., 647-15, Yoksam-Dong, Kangnam-Gu, Seoul 135 723, Korea. Tel.: 82-2-222- 1840 and Fax 82-2-538-5343 or 538-3874.

SPAIN (from Alberto Gonzalez)

The General Population Survey provides an updated sampling design for the execution of population surveys between censuses. The data from the **1991 Population and Housing Census** have been used to revise this sampling design, in order to adapt it to the current population structure. This revision has given rise to the following modifications:

1. Selection of a fresh sample of primary units (census sections)

   In order to measure adequately the evolution of population characteristics, the General Population Survey makes use of a standing sample of census sections. Only for the Census or Register data, and with the purpose of adapting the sampling structure to that of the population, are a probability study and a fresh sampling allocation carried out every five years, involving renewal of about 10% of the sections in the sample.

   However, using data from the 1991 Population and Housing Census, the sample has been renewed totally, rather than partially as before. This meant the selection of 3,200 new census sections.

2. New stratifications

   While maintaining the existing approach, that is grouping the sections by strata according to the importance of the population size of the municipality they belong to, new strata have been introduced in a number of provinces. The number of strata has increased from 260 to 271 in the new design.

3. New sub-stratification approach

   Within each stratum, the sections are grouped into sub-strata taking into account the socio-economic category of the section's working population. The sub-stratification of the previous design has been replaced by a new approach which includes a new sub-stratum made up of all the sections whose non-working population is three times as large as the working population. This new approach makes use of a new grouping of the 18 socio-economic condition categories giving rise to 16 sub-strata in each stratum instead of the 7 former ones.

   This new design is being used in the Labour Force survey since the first quarter of 1995, and in the Family Budget Survey since the first quarter of 1996.

The European Household Panel is a fresh statistical operation, harmonized at the European Union level and instigated by EUROSTAT. It provides a statistical tool to allow the study and follow-up of living and working conditions, as well as of social cohesion, taking into account the information requirements of the European Union's active policies in these fields and their effects on the population.

The size of sample has been set at 8,000 households thus making it possible to provide general information for major regions, and a limited number of characteristics for Autonomous Communities. Initially, three waves were planned, two of which were carried out in October 1994 and 1995 respectively. Preparatory work for the third wave has started. For the sampling design, the methodology of the General Population Survey has been followed.

UNITED STATES (from Dan Kasprzyk)

In 1991, staff at the Research Triangle Institute (RTI) began to develop Audio Computer-Assisted Self-Interviewing (ACASI) to overcome the challenges of illiteracy and item sensitivity. Using headphones like those with portable tape-recorders, responders listen to survey questions and instructions that have been digitally recorded and stored in the memory of a personal computer. As respondents proceed through the interview, they enter their answers using the personal computer. The software that runs the interview evaluates each response based on programmed routing logic and determines the next appropriate question to present to a respondent. Similar technology has been used to conduct telephone interviews (Telephone ACASI). Instead of listening to text through headphones and entering answers on a computer keyboard, respondents listen to digitally recorded questions through their telephones and respond by pushing keys on their touch tone telephones. In recent applications of this technology, respondents preferred using ACASI when responding to sensitive questions, and have little (if any) difficulty using the computer software and hardware.

Charles F. Turner, H. Miller, and T. Smith from Research Triangle Institute (RTI), and Joseph Catania from the University of California San Francisco are evaluating the use of these technologies (ACASI and Telephone ACASI) from two perspectives: 1) respondents' reactions to the methodology and 2) the technical capabilities and limitations associated with the systems. In both respects, testing to date has indicated high levels of success.

On one study, ACASI was used to gather data on sensitive behaviors from a general population sample from the Baltimore, MD area. In a second study a sample was drawn from the Chicago metropolitan area and telephone interviews were completed. Levels of reported behavior to the sensitive items and respondent impressions of the technology were compared during the analysis of the data. Respondents' overall preferred method for reporting sensitive behaviors was the computer-assisted technique. Reported incidents were substantially higher using the computer-assisted technique.

For more information about this technique and these projects contact Paul Biemer of the Research Triangle Institute, P.O. Box 12194, Research Triangle Park, North Carolina 27709-2194, USA.

.....

Call for Papers

IASS/IAOS Satellite Meeting
on Longitudinal Studies
Jerusalem, August 27-31, 1997

The satellite meeting on Longitudinal Studies will follow the ISI session in Istanbul. It is intended to cover aspects of both the design and analysis of sample surveys (prospective or retrospective), in which the same units are investigated with respect to several points of time, or over periods of time. This includes fixed or rotating panel surveys and longitudinal studies based on administrative data or on censuses. Although the term *Longitudinal Studies* is to be regarded in a general sense, papers proposed should be limited to studies with a predominantly statistical component. The major focus will be on surveys with a longitudinal component, such as panel surveys, and the special problems posed by the longitudinal nature of the survey.

The first two days (*August 28-29*) will be devoted to papers on **new** methodological research results that relate specifically to longitudinal surveys. These could include theoretical models, design issues, practical questions of collection and processing, longitudinal analysis and frameworks for inference, measurement and other non-sampling errors. There is special interest in papers devoted to the problems of weighting for panel surveys, to take into account population and household composition changes, attrition, non-response, etc. Research efforts that focus on imputation strategies to compensate for panel/wave nonresponse will be given special consideration. Proposals for papers devoted to methodological issues specific to certain subject matter areas, such as household economics, birth cohort studies or establishment panel surveys, are also welcome.

The last day (*August 31*) will be devoted to a workshop aimed at the presentation and discussion of *case-studies*. These should reflect methodological innovations and novel

developments of current practice in longitudinal surveys. Operational issues in the design and conduct of longitudinal studies will also be given attention (e.g. cost-benefit analyses; frequency of data collection/length of recall period and impact on data quality; efficient design strategies to minimize attrition due to change in location). Proposals for papers reporting on surveys that highlight these topics and those covered in the first two days are welcome. Methodological aspects, applied problems and the relationships between design and analysis should be emphasized. Of particular interest would be papers dealing with plans for new longitudinal studies and with longitudinal surveys in the settings of national and international statistical systems.

The meeting is sponsored by the *International Association of Survey Statisticians* (IASS), the *International Association of Official Statistics* (IAOS), the *Israel Central Bureau of Statistics*, the *Israel Statistical Association* and *the Department of Statistics at the Hebrew University*. Members of the International Program Committee are: *Lidia Barreiros* (EUROSTAT), *David A. Binder* (Statistics Canada), *Steven B. Cohen* (AHCPR, USA), *Jean-Claude Deville* (INSEE, France), *Susan Linacre* (ABS, Australia), *Graham Kalton* (WESTAT, USA), *Gad Nathan* (CBS, Israel), *Danny Pfeffermann* (Hebrew U., Israel), *Fritz J. Scheuren* (George Washington U., USA) and *Chris Skinner* (U. of Southampton, UK).

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. The deadline for submission is 31 December 1996. Earlier submissions are encouraged and notifications of acceptance will be sent as soon as possible. Acceptance is conditional on the attendance at the meeting of at least one of the authors. Abstracts should be submitted, preferably via E-mail (in ASCII or TeX), by Fax or by mail to:

Gad Nathan, Central Bureau of Statistics, 91905 Jerusalem, Israel; Fax: +972-2-6522-319; E-mail: gad@olive.mscc.huji.ac.il, or to any other member of the Program Committee.

Additional information is available on the Internet Web site: *http://pluto.mscc.huji.ac.il/~gad/smls.html*.

.....

Conference on Demography, Statistics and Privacy and Surveys on Sensitive Subjects
October 23-24, 1995

The following papers are available in English:

Statistics, ethics and policy: case of French situation
by **M. Jacod**, *INSEE, France*

Situation in European countries
by **G. Als**, *Luxembourg*

The new European directive: aims, descriptions and perspectives
by **C. Engelage**, *Eurostat*

Census by matching registers: The Danish experience
by **L. Thygesen**, *Dunmark Statistik*

Population registers or census
by **A. Laihonen**, *Statistics, Finland*

Survey on violence against women
by **B. Petrie**, *Statistics Canada*

Surveys of drug use
by **R. Hartnoll**, *Great Britain*

Surveys on physical or mental handicap
by **H. Meltzer**, *Great Britain*

Surveys about homeless people
by **M. Burt**, *USA*

Surveys about autochtonous people
by **B. Petrie**, *Statistics Canada*

.....

Francophone Colloquium On Surveys
Rennes, June 19-20, 1997

With the support of: IASS, IRISA

Not since 1991 has a general colloquium on the subject of surveys been held in France, and this is why the ASU and the SSF felt that it was time to take the initiative and organize one at Rennes in June 1997.

Scientific Program:
The Francophone Colloquium on Surveys has the following objectives:

- to promote exchanges between researchers and practitioners;
- to take stock of the latest theoretical and practical advances;
- to draw up a list of problems to be studied in the years to come.

The scientific program covers all aspects of these issues, from the experiences of users to software tools to the development of new statistical findings.

The colloquium will be of interest to any person concerned with surveys, whether he or she is from a university, a national statistical institute or the world of business and industry.

The scientific committee will examine papers on the following aspects of surveys:

1. Theoretical problems and statistical tools
2. Data collection
3. Analysis and processing of survey data
4. Algorithms and software
5. Applications and practical problems

Location: Université Rennes 2
Proposed papers and suggestions should be sent to Gildas Brossier (Université Rennes 2, 6 avenue Gaston Berger, 35043 RENNES CEDEX), E-mail: gildas.brossier@uhb.fr

Scientific Committee:
Chair: A.M. Dussaix (Essec, Paris)
L. Lebart (CNRS, Paris)
J.C. Deville (INSEE, Paris)
G. Nathan (Univ of Jerusalem)
J.M. Gauthier (HEC, Paris)
C.J. Skinner (Univ of Southampton)
P.J. Gomez (INE, Porto)
Y. Tillé (Free Univ of Brussels)
M. Hidiroglou (Statistics Canada, Ottawa)

Organizing Committee:
Chair: G. Brossier (Univ. Rennes 2)
M. Bennani Dosse (Univ. Rennes 2)
C. Danysz (ENSAI - Rennes)
S. Joly (Univ. Rennes 2) - treasurer
A. Morin (Irisa - Rennes)

Colloquium Secretariat:
Michèle Cassin and Nicole Piton
Université Rennes 2 - 6 avenue Gaston Berger
35043 RENNES CEDEX - Tel: 99 14 17 86, Fax: 99 14 17 85

.....

Quatrièmes Journées de Méthodologie Statistique
INSEE (Paris, France)
December 11-12, 1996

The 5 sessions, to be held in French, will cover the following topics:

- Time series
- Inequality Measures
- Questionnaire design
- Surveys on Sensitive Subjects
- Small Area Estimation

For further information, contact: Jean-Claude DEVILLE, INSEE, Unité de Méthodologie Statistique, Box F401, 18 Boulevard Adolphe Pinard 75675 Paris, Cedex 14, FRANCE. Fax: 33-1-41-17-62-87

.....

Planned Programme of Short Courses, Istanbul, ISI 1997 Meetings

The two short courses offered at the 1995 Beijing meetings are planned to be presented again at the 1997 meetings in Istanbul. They are: 1) Workshop on Survey Sampling in Developing Countries; and 2) Variance Estimation in Complex Surveys.

Other ideas for short courses are encouraged. Please send ideas (and suggestions for instructors) to Fritz Scheuren, Visiting Professor of Statistics, The George Washington University, 1502 Ruffner Road, Alexandria, VA 22302, E-mail: Scheuren@aol.com, Fax: 703-548-1120.

.....

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 11, Number 3, 1995

Contents Volume 11, Number 4, 1995

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 21, Number 2, December 1995

Contents Volume 22, Number 1, June 1996