

Plans for the 1991 Redesign of the Canadian Labour Force Survey

D. Drew, J. Gambino, E. Akyeampong and B. Williams, Statistics Canada

J. Gambino, Social Survey Methods Division, 16-A, R.H. Coats Building, Statistics Canada, Ottawa, Canada, K1A 0T6

KEY WORDS: Survey design; sampling; survey operations; questionnaire content.

1. Introduction

The Canadian Labour Force Survey (LFS) undergoes a redesign following each decennial census. Redesigns have ranged in scope from making minor changes in the sample design to revising simultaneously many aspects of the survey. The last major revision, encompassing changes to the sample design, questionnaire content, processing systems and outputs, took place in the 1970s.

Work on the current redesign is proceeding on four fronts:

- (i) sample redesign
- (ii) operations, processing and outputs
- (iii) computer assisted interviewing
- (iv) questionnaire content.

This paper describes the work proposed for the current redesign. In section 2, the Labour Force Survey itself and previous LFS redesigns are described briefly. Some motivation for the current redesign is also given. The next four sections discuss items (i) to (iv) above, respectively. The major tests and milestones leading to implementation of the new design are listed in section 7.

2. The LFS and its Redesign

The Labour Force Survey began as a quarterly survey during the Second World War and became a monthly survey in 1952. The present survey obtains information from over 62,000 households each month. The data are used to produce labour force estimates at the national,

provincial and subprovincial levels for various classifications such as industry, occupation and age/sex group. Estimates are released just thirteen days after completion of data collection.

The LFS sample is selected in several stages using various methods, including probability-proportional-to-size and systematic sampling. The ultimate sampling unit, the household, remains in the sample for six consecutive months once it is selected. Thus five-sixths of the sample is common between two consecutive months. Singh *et al.* (1990) give details of the current LFS design.

The LFS system and frame are used by many other household surveys such as the annual Survey of Consumer Finances. The impact on these surveys of changes to the LFS design will also have to be considered.

The last major redesign of the LFS took place close to twenty years ago. At that time, the questionnaire was revised to provide more detailed information on the labour force. The size of the sample was increased to enable production of provincial and subprovincial estimates. The processing system was modernized with the introduction of a network of minicomputers in Statistics Canada's regional offices.

The main goal of the redesign following the 1981 census was to improve the quality of subprovincial estimates by increasing the efficiency of the survey design. In addition, the flexibility of the LFS as a vehicle for conducting other household surveys was enhanced. Singh *et al.* (1984) describe the changes that took place.

Like the redesign of the 1970s, the current redesign will see changes due to the introduction of new technologies. These will include the

introduction of computer-assisted interviewing (CAI) and the conversion of the sampling, processing and tabulation systems to a modern database environment. The aim is to improve data quality and accessibility to the data without increasing the total cost of the survey.

The content of the LFS questionnaire, which has remained virtually unchanged for close to twenty years, will be reviewed in light of new data needs, improvements identified as a result of cognitive research on the questionnaire, and the opportunity provided by the implementation of CAI. The combination of new technology and questionnaire changes will make the scope of this redesign similar to the one following the 1971 census.

Even without the technological changes being considered, a redesign is needed to bring the sample up to date. By the time the new design is implemented, the old design will be using information from a fifteen year old census. Prolonging the use of the 1981 census would lead to increased problems such as the sampling of areas that have experienced tremendous population growth (referred to as growth clusters).

3. Sample Redesign

As the time since the last redesign increases, the information on which the LFS sample is based becomes less accurate. This is especially true for areas that have experienced significant population growth, decline or shift. The redesign of the survey will update the frame, reducing the extra costs involved in the sampling of high-growth areas.

The redesign also will provide an opportunity to bring the LFS definitions of major geographical areas, such as Census Metropolitan Areas, in line with the official definitions for these areas.

Another aim of the redesign, particularly for other surveys associated with the LFS, is to make the sampling frame more flexible, making it easier to avoid unwanted overlapping of

samples, while having the samples for different surveys in close proximity so that common interviewers can be used, where desired.

It has been proposed that the new LFS frame should be a general-purpose household survey frame, with the same stratification being used by all client surveys. In this scenario, the LFS would be one of several users of this household frame. Among the advantages of this approach are the following.

- Data analysis for surveys with common designs, geography and sampled areas will be facilitated.
- Stratification will be done using a broad range of variables. Such strata are likely to be more stable over time.
- It will be easier to control or eliminate sample overlap between surveys.
- There will be cost savings because only one infrastructure will be needed.

These advantages, as well as any disadvantages, will be investigated further before a decision on the implementation of the general approach is made.

A major change being considered is the replacement of the current area frame by an address register in urban areas with populations of at least 50,000. Such an address register has been constructed for the 1991 census (Drew *et al.*, 1989). A major advantage of an address register is that it can be kept current by updating it regularly using administrative data sources such as telephone billing files.

The address register can be used as a list frame. This would allow for a much simpler sample design, but would also introduce concerns about missed dwellings. This suggests the use of a dual frame: an address register list frame and an area frame. Research on dual frame methods is being conducted at this time; a summary of

some of this work is given by Mian (1990).

As in previous redesigns, a number of basic issues and assumptions are studied in the period before the implementation of the new design. Among those under scrutiny this time are the following.

- **Stratification methods:** The choice of stratification algorithms is being studied. Since previous methods performed well, it is likely that only improvements in software will be needed. However, the choice of stratification variables is subject to change. The need for contiguity will be addressed, particularly in light of the changes being considered (e.g., the use of an address register).
- **Sample allocation:** Between late 1989 and early 1990, the size of the LFS sample was increased by over 30 per cent, with the bulk of the new sample falling in the two largest provinces (Ontario and Quebec). Thus the old allocation has been altered significantly. This was done to produce better subprovincial estimates for regions used to administer the Unemployment Insurance (UI) program. It has been determined through consultation with users that defining a common set of subprovincial regions, suitable for both the UI program and to other users of Statistics Canada data, is not feasible. Thus the sample design and allocation will be complicated by the need to simultaneously provide good estimates for two different sets of regions. This was not a constraint in previous designs.
- **Estimation and variance estimation:** The estimation method used by the LFS, based on generalized least squares, is described by Lemaître and Dufour (1987) and Singh *et al.* (1990). Variance estimation is carried out using a jackknife procedure (see Singh *et al.*, 1990). Several estimation problems will be studied, including improved methods for small areas, the variance of seasonally adjusted estimates, composite estimation,

dealing with negative weights, and the impact of imputed data on estimates.

- **Miscellaneous studies:**
 - Continuous interviewing throughout the month as an alternative to the current procedure of conducting the survey in a specific week will be investigated, looking at the impact on data quality, interviewing costs and timeliness.
 - Current quality measures will be reviewed and new measures derived in light of new data collection methods being introduced.
 - The potential benefits of using geographical innovations, such as automated mapping, will be studied.

3.1 The Address Register

The Address Register (AR) was created by merging, unduplicating and geocoding addresses from sources of administrative records containing residential information. It is being used initially for the 1991 Canadian Census of Population to check the coverage of the dwellings lists created manually by census enumerators. As a by-product, the coverage of the AR will also be improved. Following this, the AR will be updated regularly using administrative data.

With regular updating of the Address Register, its use as a sampling frame is expected to reduce some problems related to rapid population growth. Under the current area-sampling design, failure to identify growth and update the sample leads to increased sampling variance, and the cost of regular updating is prohibitive.

Two alternative designs based on the Address Register are being considered. In one design, the AR would serve as a list of addresses. In the other design, a list of postal codes or blocks would comprise a first stage of sampling units; the AR would then provide a list of addresses for each unit. Study of these two options will

be undertaken, with emphasis placed on coverage issues.

4. Operations, Processing and Outputs

Many of the processing systems for the Labour Force Survey have been in place for twenty years and use dated and even obsolete technology. As a result, it is becoming difficult to keep up with the increased demand for labour market information. Thus, the redesign period is an opportune time to modernize the processing systems. Since other changes, such as the upgrade of systems in the regional offices, will occur at the same time, it would be less efficient to build temporary interfaces now, followed by implementation of new processing systems at some future date.

The current tabulation and publication systems have been reliable but are very inflexible. Retrieval of data for users often involves a look-up on a microfiche. This introduces the possibility of transcription errors. Requests for information which is not routinely produced require writing programs for execution on the mainframe, often resulting in delays of three to four weeks. With the advent of powerful microcomputers, there are now better ways of serving users of labour force data.

A number of benefits are expected from new systems. It will be possible to respond to requests for estimates in a more timely manner and for a broader range of characteristics and levels. Similarly, it will be possible to do more data analysis before the monthly release date. Using a modern database system, it will be easier to manipulate the data, encouraging exploratory analyses. It is also expected that analyses across time will be easier to carry out. Finally, because much of the current system is "hard-coded", additions or deletions to the list of characteristics to be output monthly involves tedious programming and testing. Converting to higher-level software will make such changes easier to implement.

5. Computer Assisted Interviewing

Computer assisted interviewing (CAI) represents a major advance in survey taking. It replaces a traditional paper-and-pencil interview with one where responses are entered directly into a computer, with the computer controlling the logical flow of the interview. With CAI, some errors in the responses can be corrected during the interview by embedding consistency checks and other edits into the software. This improves the quality of the information collected.

Tests of computer assisted interviewing on the Labour Force Survey have shown data quality improvements (Catlin and Ingram, 1988). These include a decrease in the frequency of edit failures in later stages of processing, and the elimination of branching errors caused by the interviewer following the wrong path in the questionnaire. In addition, CAI resulted in decreased under-enumeration of household members. The CAI test results show no significant impact on estimates of employment and unemployment.

The 1,000 LFS interviewers across Canada will be equipped with portable computers for use in face-to-face interviewing in the homes of respondents and for use in telephone interviewing from the home of the interviewer. Survey data will be transmitted electronically between regional offices and the homes of interviewers, with data encryption to ensure security. This development will benefit other household surveys as well.

A test of a portable computer with a touch-sensitive screen instead of a keyboard is under way. The test will involve eight interviewers administering questionnaires to 3500 households over a seven-month period. The sample, which was selected from active LFS clusters, follows the usual LFS rotation pattern. This test will provide basic information which will be used in the design of more extensive tests preceding the implementation of the new LFS design.

6. Questionnaire Content

The last major changes to the LFS questionnaire were implemented in the mid-1970s. At that time, a more direct style of questioning was introduced. In addition, questions were formulated to bring the underlying concepts in line with international standards.

No changes are needed to the concepts of employment and unemployment in the current redesign. However, there has been increasing demand for new types of information which cannot be provided by the current Labour Force Survey. These include (i) data on non-standard, but increasingly common, work arrangements such as home-based work, compressed and extended work weeks and job sharing, (ii) for multiple job holders, more information on the second job, such as occupation, (iii) data on unionization, (iv) data on earnings, including compensation during periods of absence from work, (v) more information on hirings and separations and (vi) information on training by firms.

The availability of such information would benefit planning and decision-making on diverse policy issues. For example, data on earnings would provide timely tracking of wages by occupation, industry and union status. The concern that adding questions on earnings to the LFS will increase nonresponse needs to be addressed.

Improvements to the current questions will be investigated. The objective will be to reduce the potential for response error. The changes being considered include shorter and clearer wording of questions, splitting complex questions into two or more separate questions, building concepts and definitions into the questions or into a preamble to the questionnaire, and increased use of pre-coded responses. Changes in the order of several questions will also be studied. In addition, a recent study has shown that there are problems in certain parts of the current questionnaire for groups such as the self-employed. These need to be looked into further.

Following an evaluation of the questionnaire, at least one alternative questionnaire will be developed. Initially, questions will be studied using cognitive research techniques like those used recently to evaluate the current questionnaire. The alternative questionnaire(s) will be tested in the field. The end result will be the adoption of a new questionnaire for the redesigned LFS.

7. Tests and Milestones

A number of tests will precede implementation of the new design. A small-scale test of a portable computer was mentioned in section 5. In this section, brief descriptions of subsequent studies are given.

Focus testing of questionnaire alternatives (1992): Following an evaluation of the current questionnaire in early 1991, and considering the content issues discussed in section 6, alternative versions of questions will be prepared. These will be studied using focus groups. Based on the findings, one or more alternative versions of the questionnaire will be developed during 1991 and tested early in 1992.

Test of sampling from the Address Register (1992): This test will involve sample selection and dwelling listing, but no interviews will take place. An important aim of the test is to evaluate methods of improving the coverage of the Address Register (i.e., to minimize the number of missed dwellings).

Data Quality Test of Computer Assisted Interviewing (1992-1993): This test will measure the effects on data quality of computer assisted interviewing. It will be a follow-up to the operational test of CAI being conducted during 1991. The primary purpose of the test is to determine if there are differences in the estimates of labour force characteristics based on data obtained via paper-and-pencil and CAI interviews. The test will be based on a sample of 3600 households per month for the CAI treatment, with a similar number of households

serving as a control group. The test-based results can also be compared to the ongoing LFS.

If the test shows that CAI has no impact on the survey estimates, then the implementation of CAI will be moved up to 1993 or 1994. This is due to the cost savings CAI will bring, particularly because of the elimination of data capture as a separate activity.

Should the test indicate that CAI does have an impact on survey estimates, then its implementation will occur simultaneously with the introduction of the new design and the revised questionnaire.

Combined CAI/questionnaire test (1993): This will be a large-scale test of both computer assisted interviewing and alternative questionnaires. The effects on data quality and survey estimates will be studied.

Parallel run (October 1994 to December 1995): It is currently planned to phase in an overlap sample in the last quarter of 1994, followed by a parallel run during 1995. The objective is to conduct both the old and redesigned surveys in parallel to study the effect of the redesign on survey estimates. The size of the parallel run, which should be sufficient to measure differences due to the redesign at the national and provincial levels for the major characteristics, has yet to be determined.

Phase-in of the new sample (October to December 1995): The new sample will be phased in off-line (i.e., not used for official estimates) during the last three months of 1995. The redesigned survey will be fully implemented in January 1996 (i.e., as of that date, published estimates will be based on the redesigned sample).

8. Conclusion

The redesign of a survey as large and complex as the Labour Force Survey is an expensive and

time-consuming, but necessary, endeavour. The aim will be to produce a new survey capable of meeting the needs for current labour market data over the next decade in a cost-effective manner, and capable of providing the infrastructure to support an expanded program of household surveys in the 1990s and beyond. By introducing new technology, we expect both reduced operational costs and improved data quality, as well as better service to data users. By bringing the frame up to date, and by making it flexible enough to be used by most household surveys, we expect to reduce costs, have a more efficient design, and reduce response burden. Finally, by modifying the questionnaire, we expect to improve the quality of the items currently collected, and to add new information to satisfy an increasing demand.

References

- Catlin, G. and Ingram, S. (1988). The Effects of CATI on Cost and Data Quality. In *Telephone Survey Methodology*, edited by R.M. Groves *et al.*, New York: John Wiley and Sons.
- Drew, J.D., Royce, D.R. and Van Baaren, A. (1989). Address Register Research at Statistics Canada. *Proceedings of the Survey Research Methods Section, American Statistical Association*.
- Lemaître, G. and Dufour, J. (1987). An Integrated Method for Weighting Persons and Families. *Survey Methodology*, 13, 199-207.
- Mian, I.U.H. (1990). *Dual Frame Estimation of Proportions in Sample Surveys*. Statistics Canada, SSMD, Technical report.
- Singh, M.P., Drew, J.D. and Choudhry, G.H. (1984). Post '81 Censal Redesign of the Canadian Labour Force Survey. *Survey Methodology*, 10, 127-140.
- Singh, M.P., Drew, J.D., Gambino, J.G. and Mayda, F. (1990). *Methodology of the Canadian Labour Force Survey, 1984-1990*. Ottawa: Statistics Canada, Catalogue no. 71-526.