SAMPLE ALLOCATION FOR THE CANADIAN LABOUR FORCE SURVEY

Ijaz U.H. Mian and Normand Laniel, Statistics Canada
Ijaz U.H. Mian, SSMD, Statistics Canada, 16-E R.H. Coats Bldg., Ottawa, Ontario K1A 0T6, Canada

KEY WORDS: Multi-stage design; Reliability of estimate; Survey redesign.

SUMMARY

The Canadian Labour Force Survey (LFS) provides estimates of the level and change in labour force characteristics of the population. The LFS is undergoing its 1991 post censual redesign. This document discuss the allocation of the redesigned LFS sample which resulted from several studies conducted for this purpose. The LFS sample is allocated to several provincial and sub-provincial areas of interest in stages. Some of the important issues regarding the sample allocation of the survey are also discussed. The steps of the selected allocation approach for the LFS and some comparisons of the results based on the current and redesigned allocations are also given.

1. INTRODUCTION

The monthly LFS is one of the oldest and major household survey conducted by Statistics Canada. It provides monthly and quarterly estimates of the level and change in labour force (LF) characteristics of the Canadian population (such as unemployment rate, total unemployment, total employment, population in LF, etc.) at the national, provincial and sub-provincial levels. It uses a stratified multi-stage sample design. The LFS is undergoing its 1991 post censual redesign. The major purpose of this redesign was to attempt to introduce new survey methodologies in order to reduce cost and to improve the reliability of estimates, especially at the sub-provincial levels, by increasing the efficiency of survey design. The objectives and plans for the 1991 LFS redesign are discussed by Drew et al. (1991) and Singh et al. (1993). The major purpose of the sample allocation studies was "to look at different alternatives for allocating the fixed national sample to provincial and sub-provincial areas of interest, and eventually to strata, from a cost-variance point of view". This is very important because the LFS sample not only provides estimates of the LF characteristics but also is the whicle for other surveys such as the Household Facilities & Equipment Survey (HFES) and others. In addition, the LFS frame is also used as a general purpose vehicle to conduct several other household surveys such as the annual Survey of Consumer Finance (SCF). The general methodology of LFS, under the current design, is described by Singh et al. (1990).

This document describes the current and redesigned sample allocations of LFS. The redesigned sample allocation resulted from two major allocation studies for the redesigned LFS sample. The first study, discussed by Mian and Laniel (1993), was carried out to allocate the fixed national sample to provinces under different allocation schemes. The second study was carried out to look at different alternatives for allocating the resulting provincial sample sizes to sub-provincial areas of interest such as Economic Regions (ERs), Unemployment Insurance Regions (UIRs) and Census Metropolitan areas (CMAs). The second study is discussed in a report by Mian and Laniel (1994). The selected allocation approach for the LFS is given. The redesigned allocation will be used by this survey starting in 1995. Some comparisons of the results are also given.

1.1 Sample Size

The total LFS sample, after the 1993 reduction, is 58850 households per month from which a sample of 16540 households is entirely funded by Human Resources Development (HRD), previously known as Employment and Immigration, Canada. The rest of the sample (42310 households) is the core LFS sample funded by Statistics Canada. The HRD funded sample was added to the LFS sample in 1989. It was added to meet the request for more and equally reliable estimates for the UIRs. In the redesigned sample allocation, we decided to keep this division of the total sample.
Thus, the two parts of the total LFS sample are:

i) Core sample (42310 households)
ii) HRD sample (16540 households)

Because of Statistics Canada's commitment with the client, HRD sample will be used only to improve the reliability of estimates for the UIRs.

1.2 Regions/Areas of Interest

Reliable estimates of LF characteristics from LFS are normally required for the following regions/areas of interest:

a) Canada
b) Provinces of Canada
c) Economic Regions (ERs) or groups of ERs
d) Unemployment Insurance Regions (UIRs)
e) Census Metropolitan Areas (CMAs)

The ERs are geographical areas, within the provinces, of similar socio-economic conditions. It should be noted that, according to the new LFS definitions, CMAs are also UIRs, except for Ottawa/Hull. The CMA of Ottawa/Hull is composed of two UIRs. It is the only region of interest which does not respect the provincial boundaries. For the redesigned sample, we decided to treat this CMA as two separate CMAs, namely Ottawa and Hull. The number of provincial and sub-provincial areas for which the estimates of LF characteristics normally required are given in Table 1.

Table 1: The number of ERs, UIRs and CMAs within each province.

Province	# of ERs	# of UIRs	# of CMAs
Newfoundland	4	3	1
Prince E. Island	1	1	0
Nova Scotia	5	5	1
New Brunswick	5	4	1
Québec	16	13	6*
Ontario	11	18	10*
Manitoba	8	3	1
Saskatchewan	6	4	2
Alberta	8	4	2
British Columbia	8	6	2
Canada	**72**	**61**	**26**

*The Ottawa/Hull CMA is treated as two CMAs.

Two neighbouring ERs in each of the provinces of Québec, Manitoba, Saskatchewan and British Columbia are mostly rural or remote areas with very small populations. The allocation procedure would allocate very small sample to these ERs which would not be sufficient to publish the estimates for these regions. On the other hand, large sample sizes for these ERs are not feasible due to response burden considerations and high data collection costs in these regions because of their rural/remote nature. As it is done under the current LFS design, we decided to combine the two neighbouring ERs into one region.

1.3 Primary Stratification

The first level of stratification for LFS consists of the ten provinces of Canada, excluding Yukon and Northwest Territories. Each province is then geographically divided into one or more ERs of similar economic nature. Furthermore, the 25 big cities and surrounding areas are defined as CMAs. It should be noted that most of the CMAs do not respect the geographical boundaries of the ERs.

A totally different geographical division of each province is into UIRs. All the UIRs respect the geographical boundaries of the provinces but most of them do cut across ERs. Because of overlap between the boundaries of ERs and UIRs, we have several [ERxUR] intersections within each province. For illustration purposes, Table 2 shows 8 non-overlapping [ERxUIR] intersections with non-zero populations in the province of Saskatchewan. As it will be seen later, these intersections had played an important role in the redesigned LFS allocation.

Table 2: The [ERxUIR] intersections in the Province of Saskatchewan.

ER	UIR			
	48	**49**	**50**	**51**
710	INT-1		INT-2	
720			INT-3	
730		INT-4	INT-5	
740			INT-6	INT-7
750&760				INT-8

Note: Only the intersections which has non-zero populations are numbered.

As mentioned earlier in the prvious section, two of the ERs in each of the provinces of Québec, Manitoba, Saskatchewan and British Columbia were collapsed (combined). After collapsing, there is a total of 128 [ERxUIR] intersections with non-zero populations across Canada.

1.4 Reliability Objectives

The data reliability objectives for the purpose of redesigned LFS sample allocation were:

For Canada and Provinces: No reduction in the CVs of the estimates at the provincial and national levels. The currently achieved CV of the estimate of total unemployment at the national level is about 2%. The current CVs at the provincial levels varies considerably from one province to another. However, the achieved CVs for the monthly estimates of total unemployment are much smaller than 8% (usually 4% to 7%), except for the province of Prince Edward Island. For this province, the CV was allowed to be slightly higher than 8%. This is mainly to keep the sample size small in order to reduce the response burden on the population of this small province.

For ER (or Group of ERs): The quarterly estimates of the total unemployment with CVs of 25% or less. A minimum required core sample size to achieve this CV was fixed at 200 (300 in Alberta) for each ER.

For UIR and CMA: The quarterly estimates of the total unemployment with CVs of 15% or less. A minimum required total sample size to achieve this CV was fixed at 600 for each UIR. It should be noted that the HRD funded sample will be used only to achieve this objective. Since all CMAs are UIRs, the objectives for CMAs are same as the objectives for UIRs.

2. CURRENT ALLOCATION

The current provincial sample sizes of LFS are based on several historical changes in them over several years. The core sample size for this survey was 47760 households per month during mid 1986 to September 1989. In October 1989, the sample was supplemented by HRD funded sample of 16540 households. The core sample was later reduced to 46460 households in April 1990. The most recent sample decrease in the core sample was in April 1993 which resulted in a core sample of 42310 households. The core and total provincial sample sizes under the current LFS allocation are given in Table 3.

Under the current LFS design, each province is divided into self representing (SR) and non-self representing (NSR) areas. An SR area is composed of one or more SR units. The SR units are usually large urban centres having a population of 15000 or more in most provinces. The NSR area is the remaining area of the province which itself is composed of special areas, small cities (which do not qualify as SR units) and rural areas. At the time of 1981 LFS redesign, the optimum sample sizes for SR and NSR areas within the provinces were determined. These optimum sampling rates (based on SR and NSR sample) were maintained, where possible, throughout the province and served as a basis for the within province allocation. Some adjustments in the sample sizes for some sub-provincial areas were made to meet data reliability objectives for those areas. This was done by fixing the minimum required sample size to achieve the acceptable CV or by pooling some of the neighbouring areas.

As mentioned earlier, the core sample was supplemented by HRD funded sample in October 1989. A model was developed to figure out the additional sample required to attain a minimum 15% or better CVs for the quarterly estimates of total unemployment for all UIRs. The sample was increased only in those UIRs which required the additional sample to meet minimum 15% CV objectives. This resulted in big sample increases for the larger provinces such as Ontario, Québec and British Columbia.

In order to incorporate the sample decrease of April 1993, the core sample sizes in most of the ERs were reduced. This was done with have a small impact on the reliability of estimates for UIRs.

It should be noted that the sample was not directly allocated to UIRs at the time of last LFS redesign. Thus the stratification under the current design does not respect the boundaries of UIRs. The estimates for these regions are produced using domain estimation techniques.

3. SAMPLE ALLOCATION STUDIES

The redesigned LFS sample allocation, discussed in Section 4, is a result of two major sample allocation studies. Several sample allocation options were investigated for between and within province allocations. The studies are summarized in this section.

3.1 Provincial Allocation Study

As mentioned in the previous section, the LFS sample has gone through many changes (increases and decreases) in the past several years. It was felt that these changes in sample sizes and the population sizes may have taken the provincial allocation away from what is optimum for the provinces. For this reason, their might be a considerable loss in reliability for the provincial and national estimates. A study was carried out to investigate this and to find a compromise allocation which can provide more reliable estimates of LF characteristics at the provincial and national levels. Several allocation schemes for allocating the fixed national sample between provinces were considered. The sample allocation schemes considered were

a) Revised current allocation,

b) Neyman allocation,
c) Proportional allocation,
d) Kish allocation (Kish, 1976),
e) Power allocation (Bankier, 1988),
f) Square-root allocation

and few others. The core national sample size was allocated to provinces under the above mentioned schemes and the results were compared. It was found that the proportional and Neyman allocations give very similar results. Furthermore, as expected, these schemes allocate very small sample sizes to smaller provinces and large sample sizes to bigger provinces. The square-root, after some adjustment for the province of Prince Edward Island, give sample sizes which are somewhat closer to the current allocation. Furthermore, the compromise (Kish, Power, etc.) allocations give promising results for certain values of the compromising parameters. The details of this investigation are given in a report by Mian and Laniel (1993).

3.2 Within Province Allocation Study

The basic approach was to first allocate the core provincial sample sizes to ERs and UIRs. Then the HRD sample was used to supplement the core sample in those areas which needed extra sample to produce reliable estimates for UIRs. The core provincial sample sizes were allocated to sub-provincial areas using proportional and square-root allocation schemes. The allocation of the sample directly to three sub-provincial areas (ERs, UIRs and [ERxUIR] intersections) was considered. The proportional allocation was found to be better for provincial and national estimates. However, the square-root allocation was slightly better for sub-provincial estimates as it give more emphasis to the reliability of sub-provincial estimates.

The minimum required sample size for each UIRs was fixed to achieve a prespecified CV. The core sample was then supplemented with HRD funded sample to get a minimum required sample for each UIR. The prespecified CV was revised (increased or decreased) until all the HRD sample was allocated. This resulted in big sample size increases for Ontario, Québec and British Columbia. On the other hand some provinces did not get any additional sample. The details of within province sample allocation study are given in another report by Mian and Laniel (1994).

4. REDESIGNED ALLOCATION

This section describes the steps for the redesigned sample allocation of LFS. The redesigned sample allocation will be used by LFS starting in 1995. The general strategy is to first allocate the core sample to provincial and sub-provincial areas with the objective of reliable estimates of LF characteristics at the provincial and national levels. Second, use the HRD funded sample to supplement the core sample for those UIRs which can not produce the estimates with desired reliability otherwise.

Historically, total unemployment is used to allocate the LFS sample. The main reason for using this allocation variable is that it is the key variable for which the LFS produces estimates. Once again, we decided to use this variable for allocation purposes.

In contrast to the current LFS allocations, we decided to use [ERxUIR] intersections as strata. This will enable us to produce estimates for all sub-provincial areas of interest without using domain estimation techniques. The major steps to achieve the redesigned sample allocation are given below.

Step 1: Allocation of Core Sample to Provinces

This is basically the current provincial allocation of the core sample with revised sample sizes for the provinces of Manitoba, Saskatchewan, Alberta and British Columbia. The sample sizes for these provinces were revised such that the pairs Manitoba & Saskatchewan and Alberta & British Columbia have equal CVs for monthly estimates of total unemployment. The equal CV objective for these pairs of provinces was achieved through an iterative process by shifting the sample between provinces until they have equal CVs.

The equal CV objective for these pairs of provinces was motivated by the fact that a detailed review of current allocation showed big differences between the CVs of the estimates of total unemployment. The CVs for Manitoba, on average, are approximately 20% higher than Saskatchewan. Similar conclusions hold for Alberta and British Columbia. This was a direct consequence of imbalance in the current provincial sample sizes.

Step 2: Allocation of Core Sample to ERs

First the fixed provincial sample sizes were allocated proportionally (proportional to the number of households) to ERs. Then the ER sample sizes were adjusted to satisfy the minimum sample size requirement. The minimum required sample size was set at 200 (300 for Alberta) households in each ER.

Step 3: Allocation of Core Sample to [ERxUIR] Intersections

The resulting core ER sample sizes from Step 2 were proportionally (proportional to the number of households) allocated to [ERxUIR] intersections.

Step 4: Allocation of HRD Sample to UIRs

This step consisted of calculating the minimum required sample sizes for all the UIRs to obtain a specified CV (i.e., desired reliability) of the quarterly estimates of total unemployment. This minimum required CV was set about 10% in order to be conservative and to allocate all the HRD sample. A minimum sample size requirement of 600 households in each UIR was also set.

Step 5: Allocation of HRD sample to [ERxUIR] Intersections

The supplemented HRD sample sizes for UIRs from Step 4 were proportionally (proportional to the core sample sizes) allocated to corresponding [ERxUIR] intersections. This process gave us the total sample sizes for each [ERxUIR] intersection. The total sample sizes for higher levels were obtained by summing the sample of appropriate intersections.

5. CV CALCULATIONS

The reliability of estimates, measured by CV, is one of the major considerations for any survey design. We decided to calculate the CVs of total unemployment for national, provincial and sub-provincial areas. Values of certain population parameters are required to calculate these CVs. The parameter values required are: i) sizes of LF population (P), ii) number of households (N), iii) unemployment to population ratio (R) [R = unemployment rate (UR) x participation rate (PR)], iv) design effect (DEF) and v) variance reduction factors (VRF) for the quarterly estimates. Values of these parameters are needed for all [ERxUIR] intersections. In this sample allocation, the sizes of LF population (P_{INT}) and total number of households (N_{INT}) are taken from the 1991 Census of Population. The other data for intersections are not available. We decided to use R, DEF and VRF of the corresponding ERs to approximate them for intersections. The average URs and PRs of the ERs from years 1984 to 1992 are used to approximate them. Similarly, the averages of smoothed DEFs and VRFs of the estimates from years 1989 to 1992 are used. The estimates of total unemployment, U_{INT}, and its variance, Var_{INT}, in all the intersections were calculated by the formulas:

$$U_{INT} = P_{INT} \times R_{ER},$$

$$Var_{INT} = (ISR_{INT} - 1) \times DEF_{ER} \times U_{INT} \times \left[1 - (U_{INT}/P_{INT})\right].$$

Here ISR_{INT} is the inverse sampling ratio in the intersection. The estimates of total unemployment and its variance (hence CV) at the higher levels were obtained by summing the estimates and their variances over appropriate intersections. For example, the monthly estimate of total unemployment, its variance and CV for the province are given by

$$U_{PR} = \sum_{\text{All INT} \in \text{PR}} U_{INT},$$

$$Var_{PR} = \sum_{\text{All INT} \in \text{PR}} VAR_{INT},$$

$$CV_{PR} = \sqrt{VAR_{PR}} / U_{PR}.$$

In order to calculate the CVs of the quarterly estimates, the variances Var_{INT} were multiplied by VRF_{ER} before applying the summations.

The provincial sample sizes with respective CVs under the current and redesigned sample allocations are given in Table 3. Furthermore, Table 4 gives the results of current and redesigned allocations for the ERs within the Province of Ontario.

REFERENCES

Bankier, M.D. (1988). Power allocation: Determining sample sizes for subnational areas. *The American Statistician*, 42, 174-177.

Drew, D., Gambino, J., Akyeampong, E. and Williams, B. (1991). Plans for the 1991 redesign of the Canadian Labour Force Survey. *Proceedings of the Survey Research Methods Section, American Statistical Association*.

Kish, L. (1976). Optima and Proxima in linear sample design. *The Journal of Royal Statistical Society*, Series A, 139, 80-85.

Mian, I.U.H. and Laniel, N. (1993). 1991 LFS-redesign: Provincial sample allocation study. Methodology Research Report, *Statistics Canada*.

Mian, I.U.H. and Laniel, N. (1994). The sample allocation studies for the 1991 post censual LFS redesign. Methodology Research Report, *Statistics Canada*. (To be completed)

Singh, M.P., Drew, J.D., Gambino, J.G. and Mayda, F. (1990). *Methodology for the Canadian Labour Force Survey, 1984-1990*. Ottawa: Statistics Canada, Catalogue 71-526.

Singh, M.P., Gambino, J., Laniel, N. (1993). Research studies for the Labour Force Survey Redesign. *Proceedings of the Survey Research Methods Section, American Statistical Association.*

Table 3: Core and total (core+HRD) provincial sample sizes (n) with respective percentage CVs of the monthly estimates of total unemployment, based on current and redesigned LFS sample allocations.

Provinces	Current Allocation				Redesigned Allocation			
	Core		Total		Core		Total	
	n	CV	n	CV	n	CV	n	CV
Newfoundland	2240	5.4	2582	5.1	2240	5.2	2240	5.2
Prince E. Island	1421	7.5	1421	7.5	1421	7.5	1421	7.5
Nova Scotia	3101	5.4	4002	5.0	3101	5.1	4050	4.7
New Brunswick	3095	5.2	3441	5.0	3095	5.2	3480	5.0
Quebec	6474	4.1	11356	3.5	6474	3.7	11630	3.2
Ontario	8517	4.1	17388	3.3	8517	3.7	17246	3.0
Manitoba	3276	6.5	3897	6.3	3870	5.2	4428	5.0
Saskatchewan	4527	5.1	4563	5.0	3933	5.2	3987	5.1
Alberta	5205	4.5	5225	4.5	4745	4.3	4745	4.3
British Columbia	4454	5.1	4975	4.6	4914	4.3	5623	4.1
Canada	**42310**	**2.0**	**58850**	**1.7**	**42310**	**1.8**	**58850**	**1.5**

Table 4: Core and total (core+HRD) sample sizes (n) with respective percentage CVs of the quarterly estimates of total unemployment for ERs within Ontario, based on current and redesigned LFS sample allocations.

ER	Current Allocation				Redesigned Allocation			
	Core		Total		Core		Total	
	n	CV	n	CV	n	CV	n	CV
510	837	10.0	1774	6.8	884	9.6	1544	7.5
515	426	13.7	926	9.3	347	15.2	988	10.1
520	491	10.2	744	8.2	281	13.4	456	10.5
530	1891	6.6	2302	6.0	3379	4.9	4184	4.8
540	831	9.6	2541	5.5	717	10.4	2385	6.1
550	996	8.5	2657	5.2	1022	8.4	2181	6.0
560	475	12.6	1235	7.8	474	12.6	1165	8.1
570	538	9.2	1203	6.2	483	9.7	967	7.2
580	585	13.1	1150	9.3	236	20.7	516	15.7
590	942	8.9	1917	6.2	493	12.3	1889	7.0
595	505	11.7	939	8.6	200	18.7	972	9.7