SOME ASPECTS OF REDESIGN OF THE CANADIAN LABOUR FORCE SURVEY

R. Platek and M.P. Singh, Statistics Canada

1. INTRODUCTION

The Labour Force Survey, established in 1945, was completely redesigned in 1963 as a multi-purpose continuing field survey with the aim of providing monthly estimates of important labour force characteristics at the provincial and the national level. At present the survey covers all civilian, non-institutional population of 14 years of age and over from the ten provinces. All the population of the Yukon Territory and the Northwest Territories plus the Indian and Eskimo population living on reserves and crown land in the ten provinces, is excluded from the coverage of the Labour Force Survey (LFS).

The LFS is comprised of two main parts: self-representing units (SRUs) and non-self-representing units (NSRUs). The SRUs are cities whose population exceeds 15,000 persons or whose unique characteristics demand their establishment as SR units. The NSRUs are the areas lying outside of the SRUs. In addition, apartment buildings in the larger SRUs and the special areas of census each form separate frames in the LFS. Special features of the present design for the different parts are briefly indicated in the corresponding sections and for a detailed description reference may be made to a report on Methodology-LFS (1965) and a paper by Fellegi, Gray and Platek (1966).

The present design of the LFS is based on the Census of 1961. The concepts, definitions, boundary demarcations, population counts, etc. of the 1961 Census were used at several stages in designing the survey. Several changes have taken place in the structure of the labour force since the Census of 1961, which affect the design to a varying degree. To a large extent the information on these changes is provided by the Census of 1971. A complete review of the LFS design is therefore necessary using the information provided by the 1971 Census and other sources.

Due to the difference in design and extent of updating used in different parts (viz. SRU, NSRU, apartment and special area) of the LFS, it was decided at an early stage of the development of the redesign program that each part would be reviewed individually and the decision on the extent of the redesign in one part would not directly influence the decision on another part.

In Section 2 the main objectives of the redesign are summarized while in Section 3 some discussion covering the use of theoretical studies in redesign is given. The effects of the changes between the 1961 and 1971 Censuses on the design of the LFS are discussed separately for the SR units and NSR units in Sections 4 and 5, respectively. Dependency of the design on censuses and the suitability of alternative schemes of selection are also discussed in these sections. Sections 6 and 7 deal with the possible improvements, on the basis of some studies undertaken, to the apartment frame and the special area frame. In Section 8, the possibility of extending the coverage of the LFS is discussed.

2. MAIN OBJECTIVES OF THE DESIGN

In a continuing large scale multi-purpose survey such as the Canadian LFS, the major objectives and priorities are decided at the inception of the survey, and the survey is designed accordingly with scope for incorporating changes depending on future requirements. Subsequent redesigns usually examine and incorporate changes with regard to (i) up-to-dateness of the sample, (ii) coverage, (iii) introduction of new methodology and (iv) level of data requirements. Changes in the design necessitated by change in the level of data requirement are not discussed in this paper. Other factors mentioned above are briefly discussed in relation to the objectives set for the present LFS redesign.

(i) Up-to-dateness: All aspects of the design which utilize 1961 Census or other obsolete data will be examined. Changes will be made using information available from the 1971 Census, other more recent sources and the results of theoretical studies. In some parts of the LFS, for example in SRUs and the apartment sample, attempts will be made to reduce the dependency of the design parameters on the census.

(ii) Coverage: As mentioned earlier, all the population of the Yukon Territory, the Northwest Territories plus the Indian and Eskimo population living on reserves and crown land in the ten provinces, is excluded from the coverage of the current LFS. Due to an increasing demand for LFS estimates for these populations the feasibility of conducting LFS in these areas will be studied.

(iii) New Methodology: Taking into consideration the results of continuing theoretical studies on LFS data and availability of vast computer resources, more refined methods of stratification, sample allocation, formation of sampling units, etc. will be examined. In addition, several new sampling methods suitable for the LFS will be compared with the existing methods for their efficiency, cost, variance estimation and operational convenience. The method (with necessary modifications) which is most suitable will be used.

3. THEORETICAL STUDIES

A regular feature of the LFS is to carry out theoretical studies using LFS data in order to assess the suitability of the ongoing design and implement changes when necessary. The studies undertaken on a continuous basis are (i) binomial factors, (ii) components of variance and (iii) cost components. Binomial factors measure the overall effects of stratification, allocation and the methodology. Past studies on binomial factors do not provide direct evidence of deterioration of the design. In the components of variance study (Gray, 1971) the total variance is split into different com-

ponents and therefore the relative magnitude of these components would help to decide upon the sample allocation at different stages of sampling. The magnitude of these components, together with the corresponding cost components would determine the optimum size of the sampling units at different stages of sampling.

The LFS redesign procedure will not be based only on the results of detailed mathematical analysis of cost and variance. However, the results of these studies, together with some operational studies, will indicate the best of several alternatives that might be acceptable. In the finalization of the program an adjustment will be made to reconcile the theoretically desirable and operationally feasible.

4. SELF-REPRESENTING UNITS

4.1 Special Features: The self-representing units (SRUs) are cities whose population exceeds 15,000 persons or whose unique characteristics demand their establishment as SRUs. Larger SRUs are further sub-divided into compact subunits which in turn are sub-divided into segments (mostly city blocks with well defined boundaries). From each subunit six or a multiple of six segments are selected systematically with probability proportionate to size (PPS). The size measure is related to the number of households as of 1961 Census. The households within sampled segments are selected systematically. The selected households within each segment remain in the sample for six consecutive months when they are replaced by another group of households from the same segment. In addition, there is a similar scheme for the rotation of segments. On the basis of partial field counts (discussed in Section 4.3) the sample is regularly updated to take account of extraordinary growth in the subunits.

Due to regular updating of the present SRU sample and simplicity in the selection procedure, at an early stage of planning for redesign it was decided that the SRU sample should not be redesigned simply because a particular feature of the design requires revision. A decision was made to examine different features of the present design individually and then decide collectively the extent to which the SRU sample should be redesigned. Brief discussions on the choice of a selection scheme and problems of updating are given in the following sub-sections.

4.2 Selection Scheme: The present method of selecting segments, namely PPS systematic (referred to as Method 1) is quite simple to operate but suffers from certain disadvantages on theoretical considerations. From several 'unequal probability without replacement' selection methods available, two well-known methods were chosen, keeping in mind that six (or multiple of six) segments are to be selected. These methods are

Method 2: Random order--PPS systematic (Hartley and Rao, 1962).

Method 3: Random group--PPS (Rao, Hartley and Cochran, 1962).

Method 2 is essentially Method 1 applied to a random ordered list of segments within the subunit. In Method 3, the segments within a subunit are to be grouped into six (or multiple of six) random groups and then one segment selected with PPS from each group. These methods were examined for their suitability for the LFS and the following points noted with regard to the theoretical and operational considerations for each of three methods.

Theoretical Consideration:

(a) Comparison of Methods 2 and 3 under a superpopulation model indicates that in many situations met with in practice, Method 2 is likely to have smaller variance than Method 3. However, Method 3 possesses the exact theory for finite populations and hence exact variance and variance estimate can be obtained without any difficulty. Method 2 requires asymptotic approach and the variance and variance estimates are valid only for a large number of segments within each subunit.

(b) The efficiency of Method 1 currently in use is very much dependent upon the magnitude of serial correlation coefficient, which does not necessarily increase or decrease with sample size, as it is a function of variation within and between segments in addition to the size of the sample. Due to this the behaviour of the variance of the estimate is likely to be somewhat irregular for this method.

(c) Unbiased variance estimation is not possible in Method 1 whereas both Methods 2 and 3 provide unbiased variance estimators.

(d) Both Methods 2 and 3 are more efficient than PPS sampling with replacement.

(e) Method 3 provides simple and stable variance estimator as compared to many other unequal probability schemes.

(f) The joint inclusion probabilities are zero for most pairs of segments in Method 1 and in Method 2 their computation is quite involved when the number of segments is large.

Operational Consideration:

(a) In the context of LFS these three methods are equally simple to operate. Methods 2 and 3 would be expected to take approximately the same time as Method 1. Randomization needed in Methods 2 and 3 may not require extra time as some randomization is done also for Method 1 in the process of formation of rotation groups. In Method 3 the random groups themselves serve the purpose of rotation groups.

(b) Both Methods 2 and 3 are also applicable and simple for the rotation program used in the LFS.

(c) Method 3 possesses three special features, namely,

(i) Only one segment is selected with PPS from each random group.

(ii) After the formation of groups, sampling within the groups is done independently.

(iii) Each random group by itself is a subsample of segments from the subunit.

Use of these features can be well explored in updating the LFS sample and also in conducting special studies or surveys related to the labour force characteristics. Because of feature (iii) the LFS sample may serve as a suitable vehicle for special studies or surveys as they may be conducted in any one (or more) groups.

As mentioned previously, the LFS sample is continuously updated but due to the complexity of the updating method, among other reasons, it

is limited only to areas of extraordinary growth (100% or more). Because of features (i) and (ii) above it would be possible to apply directly the Keyfitz method of revising selection probabilities to the individual random groups which have developed noticeable changes in the selection probabilities, without affecting the selections in the remaining groups of the subunit. This method does not affect the selections in the remaining groups of the subunit, also, because it is simple to execute even minor changes in selection probabilities could be revised. This aspect is discussed further in Section 4.3.

Remark 1 On the basis of the above discussions, both Methods 2 and 3 appear to be more suitable than Method 1 which is currently being used in SRUs. Comparing Method 2 and 3, from a theoretical standpoint one is equally preferable to the other. However, from operational viewpoints Method 3 seems to have very definite advantages over Method 2.

4.3 Uneven Growth and Updating Procedure: Population growth usually occurs at a faster rate in self-representing units than in non-self-representing units. Along with natural growth there is a marked tendency for persons living in rural or urban centers of NSRUs and also in smaller SRUs to migrate to larger SRUs. In addition, there is frequent movement of population within a particular SRU because of developments in certain parts of the SRU, location of area of work and other facilities. The result of all these is an uneven rate of growth in different parts of the SRU.

Uneven growth, in addition to creating administrative problems concerning enumerators' assignments, increases the sampling variance of the labour force estimates. This is because the size measures used in the present design were accurate when the LFS was redesigned in 1963 but they tend to become consistently less accurate as a result of uneven growth. Continued use of these original size measures does not affect the unbiasedness of the LFS estimates, but as most labour force characteristics are correlated with the size measure, the more out of date the measures become, the lower the correlation between them and hence the higher the resultant variance of the LFS estimates.

Partial Field Count: In order to check the deterioration of the design as a result of uneven growth, a scheme of partial field count is in operation in SRUs. In this scheme every subunit is checked annually, according to a fixed rotation program, and revised dwelling counts are then obtained only for those areas where growth is 100% (or more) of the original count. The sample in such subunits (or part thereof) is then revised using a prescribed method which does not affect the unbiased aspect of the estimates.

One of the major drawbacks of the present partial field count scheme is that revised counts are not available for segments for which growth is less than 100%. The updating procedure therefore allows for an increase in sampling variance of the LFS estimates to a magnitude which in some cases may be serious. This limitation of the updating procedure based on partial field counts necessitates the revision of the sample after every census in order to check the deterioration of the design.

Complete Field Count: This scheme is envisaged as a regular feature of the LFS replacing the present partial field counts. In this plan it is intended to obtain the dwelling counts, segment maps and other necessary information for all segments of the subunits irrespective of the rate of growth. According to a fixed rotation program each subunit will be completely checked in the field annually. The data collected from the first rotation will be used in the present redesign and the data from subsequent rotations will be used in the regular updating of the sample.

For the purpose of the present redesign, on an average the field counts so obtained would be only six months out of date, whereas the data obtained from the 1971 Census, if used, would be about three years out of date by the time the last province is redesigned. Field counts would thus have advantages over the census counts with regard to their up-to-dateness, and since the sample could be selected in stages as the data are received throughout the year, it would have a better stabilizing effect and uniform workload.

During subsequent years (rotations) the data collected would enable revision of the sample even when the growth is (for example) five to ten percent and thus the LFS estimates would retain their original efficiency.

Remark 2 Complete count scheme, together with random group method of selection and Keyfitz method of revising the sample (Section 4.2, item 3) should ensure that the sample in the self-representing units is up-to-date for a long period of time. As the design in SRUs will not depend on the Census of 1971, it would not require redesigning even after the next census, with the exception of the areas annexed to SRUs.

5. NON-SELF-REPRESENTING UNITS

5.1 Special Features: Non-self-representing units (NSRUs) are areas lying outside the SRUs. These are a combination of rural areas and small urban centers. Due to the relatively low density of population the sample in NSRUs is selected in four stages from each stratum of an economic region. The required number of primary sampling units (PSUs) are delineated and 2 PSUs are selected without replacement with unequal probabilities following Fellegi's method. Segments (rural and urban separately), clusters and households form the subsequent stages of selection.

The extent of redesign required in the NSRU sample, unlike the SRUs, would depend upon a single factor, namely the extent of restratification needed as a result of changes in boundaries of economic regions and census enumeration areas and also due to annexations. These factors were examined in detail and a brief resume is presented in Sections 5.2 to 5.5.PSU delineation and selection are discussed in subsequent sections.

5.2 Changes in the Economic Region Boundaries: Each province in Canada is divided into a number of economic regions (ERs). These ERs usually consist of a group of contiguous counties (or divisions) with a similar economic structure.

In the present design of LFS the DDP ERs of 1961 were used as primary strata. These ERs were further subdivided into strata following a particular principle ensuring minimum variation in the labour force characteristics within a stratum.

A review of DDP economic regions was carried out applying 'rigorous statistical techniques' leading to the delineation of regions. In November 1971, the Regional Statistical Policy and Coordinating Committee recommended a new set of regions. Main recommendation of the committee is that these 1971 ERs be 'utilized uniformly by the bureau divisions as an additional level of spatial reference for the purpose of data tabulation, development of new data and the survey design wherever its application is appropriate'.

For the purpose of LFS, use of these subprovincial regions has added advantages from both theoretical and operational considerations, such as:

(a) because the new set of ERs is formed on 'statistical principles' on the basis of more recent information, they are suitable as primary strata
(b) an ER as an area to be further stratified is more manageable than the province as a whole
(c) as the sample would be representative of each region of the province, estimates with a known degree of precision could be easily obtained at the subprovincial level. In addition it would allow more flexibility in the design since changes could be introduced, if needed, in any one (or more) regions without affecting the design of the other regions. Also it would be useful in conducting special studies and comparisons at the ER level.

In many cases the composition of the ERs in the new set is quite different from that of the old set of DDP regions. Where there are changes in the ER boundaries, the present strata will cut across the new ER boundaries. To avoid this and retain the efficiency of stratification it is essential that changes in the strata boundaries be incorporated to accommodate the 1971 ER boundaries. The extent to which restratification would be necessary can be obtained by comparing the composition of the two sets of ERs. The number of ERs with boundary changes is given in the following Section 5.3.

5.3 Annexation to SRU: The boundaries of many SRUs have changed due to growth or some other administrative and political considerations. As a result, parts of the NSR strata or in some cases the entire strata have been included in a neighbouring SRU. Consequently, the stratification in such ERs is adversely affected which could result in poor efficiency of the LFS estimates. The number of SRUs affected due to annexation is given in the following Table 1. Figures in column (5) refer to those ERs which are affected by annexation of 10% or more. Annexation in the province of Quebec has not been obtained as all ERs are affected by boundary changes and will require restratification.

Table 1: 1971 ERs Affected

Province	Total No.of ERs		No. of 1971 ERs affected due to		
	1961	1971	ER Boundary Change	Annexation to SRU	Joint consideration
(1)	(2)	(3)	(4)	(5)	(6)
Newfoundland .	6	6	-	1	1
P.E.I.	1	1	-	1	1
Nova Scotia ..	4	5	2	2	4
New Brunswick.	4	5	3	1	4
Quebec	10	10	9	n.a.	9
Ontario	10	10	-	4	4
Manitoba	7	7	-	-	-
Saskatchewan .	6	6	5	-	5
Alberta	7	8	5	1	5
B.C.	9	9	8	1	8
Total	64	67	32	11	41

5.4 Extent of Restratification: All the ERs which are affected by either boundary changes or annexation will require restratification. Table 1 shows that in 1971, 41 out of 67 ERs fall into this category. Of the remaining 26 ERs, 6 ERs consist of a single stratum and 6 ERs do not have any NSR area (these are distributed in different provinces). Thus, in 1971 there are only 14 out of a total of 67 ERs which do not require restratification on account of boundary changes or annexation.

Other operational and theoretical considerations must be taken into account when deciding upon restratification of an ER. Some of these considerations are boundary changes in enumeration areas, deterioration in the Index of Stratification, formation of new SRUs or extention of the coverage of existing SRUs.

Of the 14 ERs mentioned above, 8 were examined for the extent of changes in the boundaries of 1961 Census EAs: Ontario (4) Alberta (1) New Brunswick (1) and Manitoba (2). As the information on such changes between the Censuses of 1961 and 1971 is not yet compiled, the frequency of such changes was obtained between the Censuses of 1961 and 1966. If the relative frequency of changes between 1966 and 1971 is assumed to be the same as between 1961 and 1966, then the examination revealed that a percentage of the 1961 EAs which have changed their boundaries in the Census of 1971 varies from a low of 20% in some ERs to a high of 52% in others.

The above considerations show that for one reason or another almost all 1971 ERs would require restratification. As a result, it is contemplated that in NSR areas restratification would be carried out in all provinces with new strata formed in each of the 1971 economic regions. A project is undertaken to examine and extend the use of computer in stratification and PSU formation.

5.5 PSU Delineation and Selection: The primary sampling units (PSUs) in NSR areas consist of a group of contiguous rural enumeration areas and reasonably nearby urban areas within the stratum. While forming the PSUs in the present design attempts were made to make each PSU a replica of the stratum to which it belongs with respect to the 'discriminating characteristics' and the rur-

al-urban population ratios. Restratification of NSR areas would imply redelineation of the PSUs. It is felt that the contiguity of the enumeration areas within PSU could be relaxed to some extent without affecting the cost of enumeration.

With complete restratification and redelineation of PSUs, the partial replacement of the sample (with maximum retention of the sampled PSUs) was found to be undesireable from both operational and theoretical considerations. A decision was made to reselect PSUs and units at subsequent stages.

As mentioned earlier, in the present design, two PSUs are selected from each stratum following Fellegi's method. This method has worked quite satisfactorily in NSRUs and is not difficult to operate for the selection of two PSUs. Unless, sufficient evidence is found in favour of some other method the PSUs in NSRUs will be selected following the present method. Consideration was given to random group method (Method 3, Section 4.2) for selection of PSUs. However, studies in different context have shown that Fellegi's method is likely to be more efficient than random group method for situations where (i) the coefficient of variation of the size variable is small and (ii) the sampling ratios (for the PSUs) is large. Both these conditions are likely to be satisfied for most of the strata because of the very principle of the PSU formation,and also the sampling ratios in terms of PSUs.

The segments of the sample PSUs will be placed in two groups (urban and rural) as in the present design. For reasons mentioned in Section 4.2, the required number of segments would be selected from each group, following Method 2 rather than Method 1, as at present. Use of Method 3 is not suitable here because of the small number of segments within each group. The clusters within sampled segments may be selected with stricter classification of possible configurations for two dimensional array of cluster selection and random start. The systematic sampling of households, the ultimate stage of sampling, will remain until a superior method is found.

6. APARTMENT SAMPLE

6.1 Definition and Coverage: In addition to the regular sample in SR areas a separate frame of apartment buildings, which have at least 30 units and 5 floors, is set up in 12 of the larger SRUs. The purpose of setting up this frame is to ensure the representation of apartment dwellers as they may be different in many characteristics from persons living in single dwellings or row housing.

6.2 Stratification and Selection Procedure: A study was conducted to examine the suitability of stratifying the apartment frame of each SRU on the basis of size of the apartment buildings, size being the number of suites in them. Frequency distribution for the SRU was obtained for 15 size classes and efficiency of stratification was calculated following two well known rules of stratification, namely (i) $\sqrt{f}$ rule and (ii) stratum total rule, for constructing two and three strata.

Both rules were found to be equally efficient for 6 out of 12 SRUs and the remaining 6 SRUs were equally distributed between the two rules. On forming two strata substantial gains over the unstratified sample were noticed in all cases. Division of the frame into three strata also showed some improvements over two strata. On practical considerations it was decided to form two strata in the seven medium sized SRUs and three strata in the two largest SRUs of Toronto and Montreal. In three smaller SRUs stratification was not considered necessary. A preliminary study conducted to examine the spatial representation of the sample revealed that before selection the two larger SRUs of Toronto and Montreal should be geographically stratified.

In the present design a two-stage probability sample is selected from the apartment frame. Apartment buildings (segments) are selected systematically with probability proportionate to size (number of suites) and households are selected systematically. The apartment frame is continuously updated according to a prescribed procedure. Consideration was given to the use of Method 3 (or 2) discussed in Section 4.2, for the selection of apartment segments from each stratum of a SRU. Due to open-endedness of the apartment frame within each strata, it is felt that PPS systematic sampling has an advantage over other methods. At the second stage, to ensure a representation of the different types of suites in the sample, a serpentine listing of the suites in the sampled apartment buildings is proposed.

7. SPECIAL AREA SAMPLE

The frame for this sample consists of special enumeration areas as defined in the Census of 1961 and the remote areas. Census special enumeration areas are grouped into three types (strata), namely, military establishments, hospitals and other institutions. The remote areas form the fourth stratum. The frame of special areas is a closed ended one. In most of the provinces, each stratum is divided into two substrata and samples are selected in two-stages from each substratum following the method used for the apartment sample. Although the labour force population in special areas is less than 2% of the total labour force, it was decided to treat them separately in order to have proper representation of the institutional population.

8. EXTENSION OF LFS TO NORTHWEST TERRITORIES AND YUKON TERRITORY

These areas differ significantly from other provinces in respect to population density, living conditions, transport and communication, seasonal variations, mobility of population, etc. These factors will have a significant bearing on the overall methodology of the LFS in these areas. As little information is known about such factors and their impact on the estimates, a pilot survey will be conducted for these areas with a view to (i) examine the operational feasibilities and to decide upon the enumeration methods, frequency of survey, rotation plan, etc., (ii) to determine the cost and other design parameters and (iii) to assess the general applica-

bility of the LFS concepts and definitions.

Acknowledgements: We are grateful to Dr. I.P. Fellegi for his comments and encouragements during the completion of this paper. To our colleagues we are also grateful for enlighting discussions and active participation in the various projects undertaken.

REFERENCES

[1] Fellegi, I.P.: "Sampling with Varying Probabilities Without Replacement--Rotating and Non-Rotating Samples"; Journal of the American Statistical Association, Vol. 58, 183-201.

[2] Fellegi, I.P., Gray, G.B. and Platek, R.: "The New Design of the Canadian Labour Force Survey"; Journal of the American Statistical Association, Vol. 62, 421-453.

[3] Gray, G.B.: "Variance Components and Variance Functions"; talk presented at the 1st Canadian Conference on Applied Statistics, Sir George Williams University, 1971.

[4] Hartley, H.O. and Rao, J.N.K.: "Sampling with Unequal Probabilities Without Replacement"; Ann. Math. Stat., Vol. 33, 350-374.

[5] Keyfitz, N.: "Sampling with Probability Proportional to Size--Adjustment for Changes in Size"; Journal of the American Statistical Association, Vol. 46, 105-109.

[6] Rao, J.N.K., Hartley, H.O. and Cochran, W.G.: "On a Simple Procedure of Unequal Probability Sampling Without Replacement"; Journal of the Royal Statistical Society, Vol. 24 (2), 482-491.

Sampling and Survey Research Staff; Dominion Bureau of Statistics: Methodology - Canadian Labour Force Survey (1965).