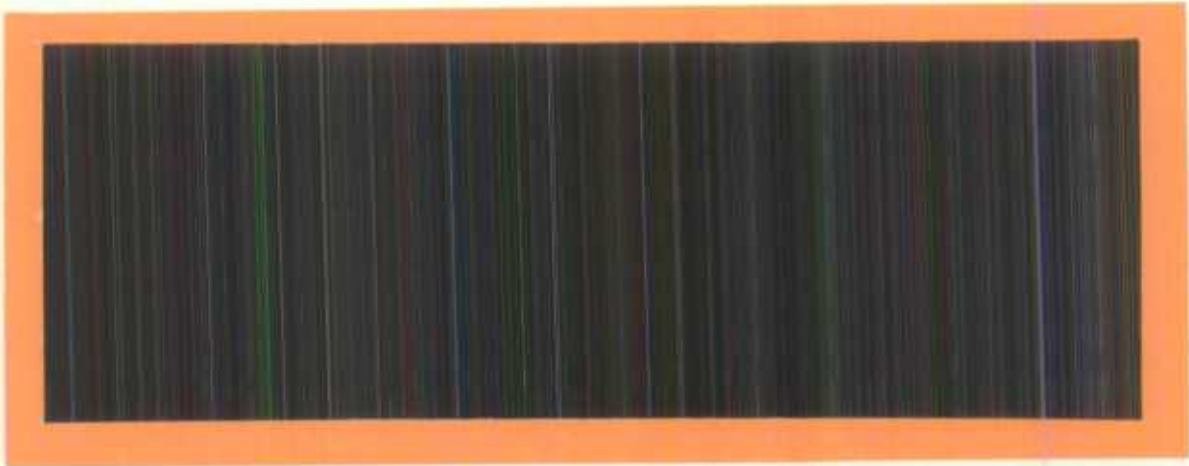# Methodology Branch

Census & Household Survey
Methods Division

Direction de la méthodologie

Division des méthodes de recensement
et d'enquêtes ménages

Canada

POST '81 CENSAL REDESIGN OF THE CANADIAN

LABOUR FORCE SURVEY

M.P. Singh, J.D. Drew, and G.H. Choudhry

Census and Household Survey Methods Division,

Statistics Canada

Post '81 Censal Redesign of the Canadian[1] Labour Force Survey

M.P. SINGH, J.D. DREW, and G.H. CHOUDHRY[1]

ABSTRACT

Following each decennial population census, the Canadian Labour Force Survey (CLFS) has undergone a sample redesign to reflect changes in population characteristics and to respond to changes in information needs. The current redesign program which culminated with introduction of a new sample at the beginning of 1985 included extensive research into improved sample design, data collection and estimation methodologies, highlights of which are described.

KEY WORDS: Continuous survey; Multi-stage sample design; Stratification; Sample reallocation; Telephone interviewing; Raking ratio estimation.

1. INTRODUCTION

The Canadian Labour Force Survey (LFS), the largest monthly household survey conducted by Statistics Canada, has been redesigned in the past following each decennial census. As a part of 1981 post censal redesign, an intensive program of research as outlined in an earlier paper (Singh and Drew 1981a) was undertaken in the areas of sampling, estimation and data collection methodologies. As the reliability of labour market data at the national and provincial levels was of sufficiently high standard, the major emphases in this redesign were on improving the reliability of subprovincial data and on making the survey more cost efficient. Towards the latter, the main thrusts were on increased automation of various steps involved in sampling, greater use of Census data in place of independently obtained information in updating the sample, and increased telephone interviewing as a regular feature of the survey. As for the improvement in the subprovincial data, alternative sampling and estimation methods were investigated leading to refinements in the earlier methods, coupled with reallocation of the sample within provinces.

This paper presents an overview of the findings of various theoretical and empirical investigations and field tests undertaken during the redesign program. Sections 2 and 3 provide the background information on objectives and the old design, while Sections 4, 5 and 6 highlight the findings of investigations leading to changes in sampling and data collection methodologies. Section 7 deals with estimation issues, and sample reallocations are discussed in Section 8. Implications of the changes made in the redesigned sample on other associated surveys are outlined in Section 9, and finally in Section 10 the benefits from the major improvements in the redesigned sample are briefly recounted, along with some mention of future research plans.

2. OBJECTIVE SETTING

A fundamental step in the redesign of a recurring survey is the re-establishment of survey objectives. For the LFS, this involved reassessment not only of the survey's principal role as a provider of current labour market information, but also of its use as a central vehicle within Statistics Canada for conducting household surveys (Singh and Drew 1981b).
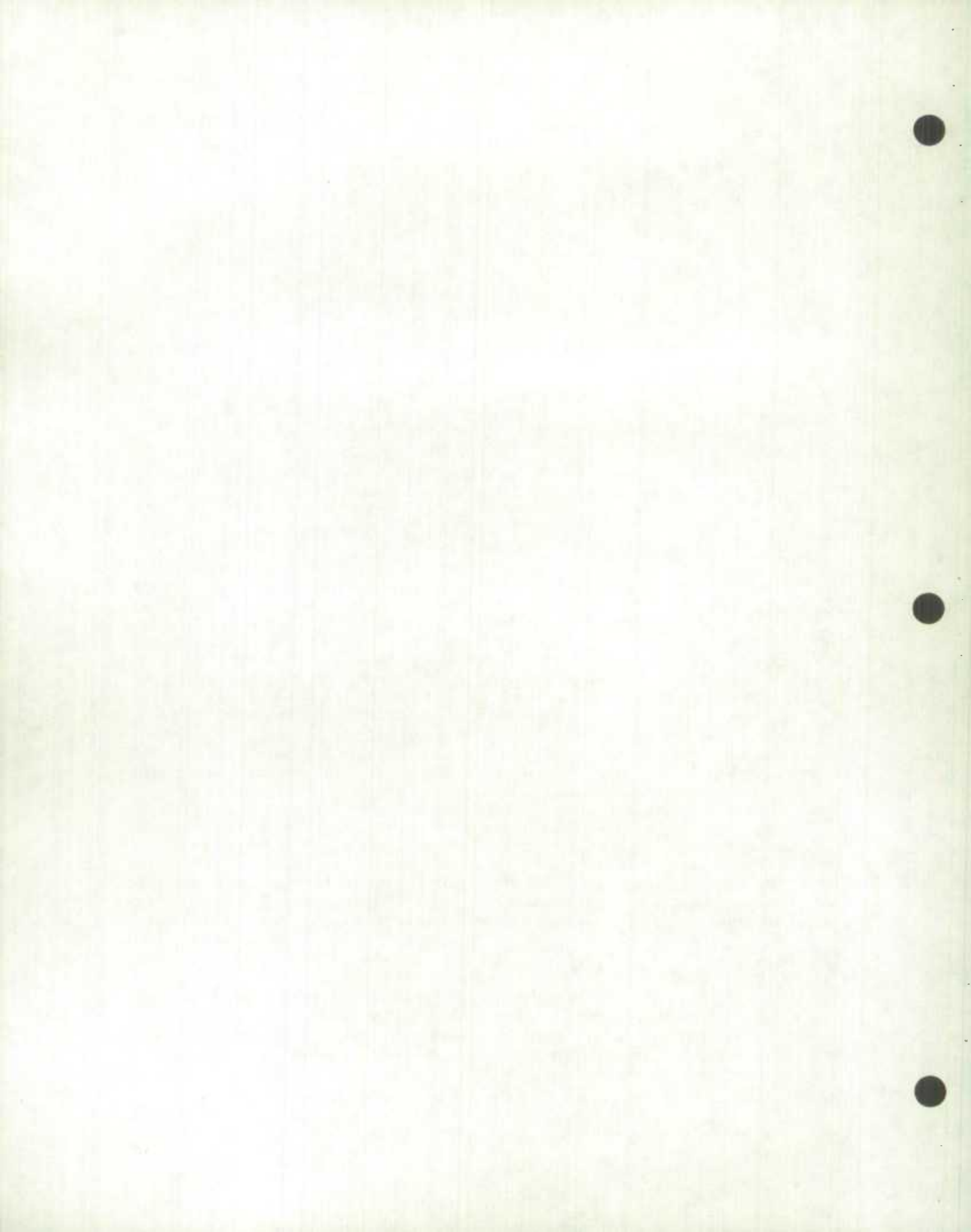
At the early stages of the redesign program it was decided that while primary orientation towards the LFS should be maintained, efforts should also be made to enhance flexibility of the vehicle for general applications. In this light, several changes are being adopted that will benefit not only the LFS but its associated surveys as well. Requirements for the labour market data are noted below, while highlights of the changes resulting in the improvements for associated surveys are given in Section 9.

Objectives relating to provision of labour market data were established in consultation with the statistical focal points within each of Canada's ten provinces, and with key federal user departments, through annual Federal/Provincial Conferences on Labour Statistics and bilateral follow-ups. In general, these consultations revealed satisfaction with data reliability for provincial and national data, but a strong desire for improved subprovincial data. Specific data reliability objectives adopted for the redesigned sample are as follows:

(i) for Canada and each of the ten provinces, no reduction in current reliability for monthly estimates of level and estimates of month-to-month change in total employment and unemployment.

(ii) for the 24 Census Metropolitan Areas as defined by the 1981 Census, monthly estimates of unemployed with coefficients of variations (CV's) of 20% or less.

(iii) for 66 subprovincial Economic Regions agreed to in consultation with the provinces, monthly estimates of unemployed with CV's of 25% or less.

(iv) for cities with population of 60,000 or more in Quebec and Ontario, and 25,000 or more in the remaining provinces, quarterly estimates of unemployed with cv's of 25% or less.

Attainment of these objectives necessitated a reallocation of sample within the provinces that entailed a shift of sample from larger cities and Economic Regions to smaller ones. This, coupled with the desire to reduce the cost of the LFS as discussed in Section 8, created high expectations from the research projects for identifying more cost efficient strategies for data collection and production of statistics from the LFS. In the following sections, these issues are addressed for the two main parts of the survey namely, Self-Representing (SR) and Non Self-Representing (NSR) Areas.

3. OLD LFS DESIGN

A complete description of the old LFS design is given by Platek and Singh (1976). Salient features are noted in this section to provide a context for discussions in later sections.

The Self-Representing Units (SRUs) in the old design corresponded to those cities which, when designing the survey, were sufficiently large to yield an expected sample of 20 dwellings, the minimum thought acceptable for an interviewer. Minimum SRU sizes ranged from a population of 10,000 in the Atlantic Region to 25,000 in Quebec and Ontario.

Within large SRUs, deep geographic stratification was carried out by grouping together contiguous Census Tracts — geostatistical areas with populations in the 3,000 - 5,000 range, whose stability from one Census to the next makes them a convenient operational unit — without any regard to optimality of characteristics. Primary sampling units, called clusters, corresponding approximately to city blocks, were delineated on the basis of field counts obtained in 1973. A two stage sample of clusters and dwellings was selected following the Rao, Hartley and Cochran (1962) pps random group method. In addition to the area frame, an open-ended list frame of apartments was maintained in the larger cities.

A major advantage of the selection method for the area frame lies in its flexibility for changes in the sample size (Singh and Drew 1977), and for sample updating (Platek and Singh 1977, Drew, Choudhry, and Gray 1978). Sample updating is necessary in SRUs because over time the design counts used in the pps selection become out of date, leading to higher sampling variances for survey estimates. Since sampling is done independently in each random group, a Keyfitz (1951) sample update can be carried out, under which revised probabilities of selection based
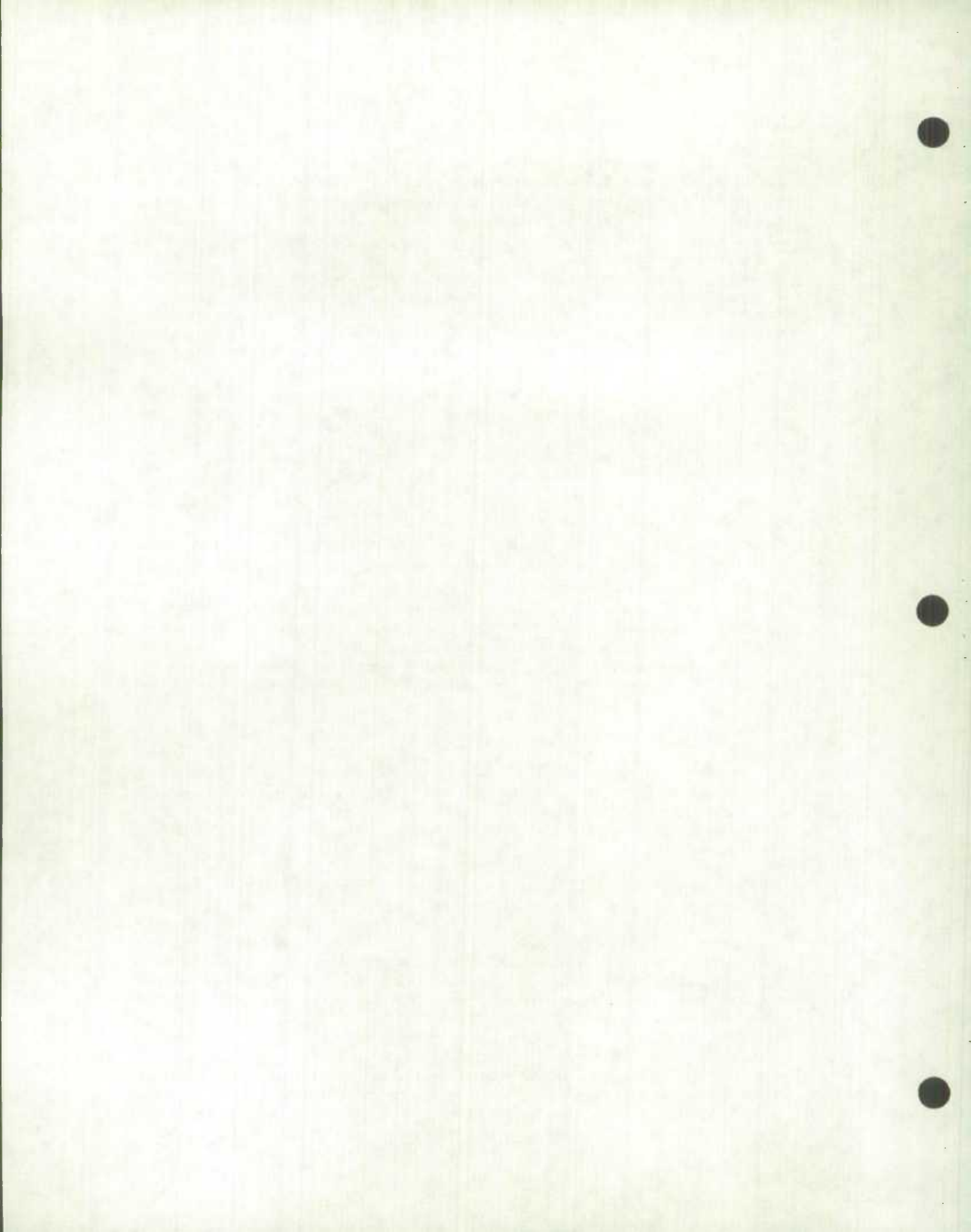
on recent dwelling counts can be incorporated, while maximizing retention of already selected units and avoiding overlap of selected dwellings between the pre- and post-update-samples. Regular updating of high growth SR areas occured from 1978 until the beginning of the redesign period in 1982, during which time almost half of the frame was updated. While the intensity of updating was not sufficient to reduce survey design effects to levels experienced during the first 4 years of the survey, it was sufficient to arrest further deterioration which had been averaging 7 – 8% per year for unemployed.

The Non-Self Representing (NSR) units are the areas outside the SRUs comprised of rural areas and smaller urban centers. In NSR areas, stratification based on industry classifications was carried out within Economic Regions, subject to the restriction that strata should be contiguous land areas. Within strata, Primary Sampling Units (PSUs) were delineated such that each PSU represented its stratum to the extent possible with respect to the ratio of rural to urban population and important LF characteristics. While the rural portions of a PSU were comprised of collections of contiguous rural EAs, urban portions could not always be made contiguous to the rural due to the restrictions placed on maintaining the rural to urban population ratio. Frequently larger urban centres had to be shared amongst several PSUs within the stratum.

At the time of the 1973 Redesign, two PSUs were selected per stratum using the randomized pps systematic method (Hartley and Rao 1962). In 1977, the LFS sample size was increased from 33,000 to 55,000 households per month, with the additional sample being allocated so as to improve data reliability at the province level. Thus the smaller provinces received larger proportionate sample size increases. The increase was achieved in NSR areas by selecting 1 – 4 additional PSUs per stratum (Gray 1975).

Within selected PSUs, urban and rural portions were sampled independently. In the urban portion of the selected PSU a two stage sample of clusters and dwellings was selected, whereas in the rural portion of the PSU a three stage design was used with secondaries (which are Census EAs or combinations), clusters (identifiable land areas having up to 20 dwellings) and dwellings as the sampling units. Except for the last stage, randomized pps systematic sampling was used in the selection.

4. REDESIGN OF THE SELF REPRESENTING AREAS

The criterion for cities to qualify as SRUs in the new design was raised to a minimum sample of 50 dwellings, since analysis of cost data indicated significantly higher unit costs for smaller SR assignments. The composition of the SR universe remained largely unaffected, however, due to the off-setting influences of the increase in the LFS sample size from 33,000 to 55,000 households during the late 1970's, and due to the reallocation of the redesigned sample.

For reasons noted earlier, the basic design in the SR areas remained the same and the main thrust for these areas was to update the size measures without resorting to a costly independent field count as used in the last redesign. In order to exploit the data collected during the 1981 census for this purpose different approaches were used in the block-faced cities (larger cities where data were available at block face level) and non-block-faced cities. The choices of updating method and sampling units by the two approaches are given below, whereas the two level stratification adopted for SR areas is given in Section 5.

Block-Faced Cities

The availability of Census data at the block face level in the built-up portions of the larger cities was the key factor in deciding to completely redesign the sample in these cities, which account for ⅔ of the SR frame. The redesign of these cities also provided the opportunity to introduce improved stratification as described in the next section.

For block-face portions of the cities, Census blocks were adopted as clusters (i.e., PSUs). Variance components under a two stage RHC random group design were studied for different choices of clusters — Census Blocks or EAs — by simulating the LFS design using 1976 Census data for the SRUs of Halifax and Saskatoon (Choudhry, Drew, Lee 1984). Study results, for the case of up-to-date size measures, showed little difference between the EA and block in terms of sampling variances. Hence, the decision in favour of blocks was made on the basis of operational considerations. The blocks provided a ready made frame (with splitting or combining needed in only 5 - 10% of cases), permitting a highly automated design with very low redesign costs. Importantly as well, data for future Censuses will be retrievable for the geostatistically stable blocks (but not for EAs which as operational units change at each Census), permitting low cost quinquennial updating of the sample. The built-up portions comprise 86% of these cities.

The EA was adopted in the non-block faced portions of the cities. Whereas the study results considered only the up-to-date case, it was felt that the EA being a larger unit than the block would be more robust to the highly clustered growth which can occur in the non-built-up portions of cities. Also adoption of the EA in these areas permitted very low redesign costs, since roughly 80% of these areas fell in Quebec and Ontario where due to the low sampling rates very little splitting was needed.

The variance study results, combined with cost results from the Time and Cost Study (Lemaître 1983), showed variances per unit cost to be quite flat in the range of 2 - 8 selected dwellings per cluster. Hence it was decided to retain the density of 4 - 5 dwellings per cluster used in the old design in strata where clusters were blocks, but to increase the density to 6 - 8 dwellings in the case of the EAs, due to their larger sizes.

Non-Block-Faced Cities

Since over 70% of the non-block-faced SRUs were either new or had changed boundaries, and since most of those remaining had not been updated since the 1973 redesign, it was decided to completely redesign these cities. Clusters were taken as individual or combined Census blocks in the built-up portions of the cities, with the dwelling counts obtained directly from visitation records and maps completed by Census enumerators. In the non-built-up portions, EAs or split EAs were taken as clusters, with field counts sometimes being required to do the splitting.

The use of Census visitation records, while more costly than procedures followed in the block-faced areas, nevertheless resulted in significant cost savings over the procedure followed in the old design of obtaining independent field counts.

5. STRATIFICATION

5.1 Algorithm and Choice of Stratification Variables

A modified version of a non-hierarchical algorithm due to Friedman and Rubin (1967) was adopted for stratification purposes in both SR and NSR areas, on the strength of findings reported by Judkins and Singh (1981), and Kostanich, Judkins, Singh and Schautz (1981) who evaluated several stratification algorithms for use in the U.S. Bureau of the Census' Current Population Survey. New features incorporated into the algorithm were a capacity to form geographically contiguous and/or compact strata, and the option to form either homogeneous clusters (i.e., strata) or heterogeneous clusters (i.e., primary sampling units within NSR strata). A detailed description of the method, and results of empirical evaluation studies are given by Foy, Bélanger, Drew and Joncas (1984). An overview is presented below.

The algorithm starts with a random partitioning of units into a specified number of strata. An iteration consists of examining in turn each stratification unit and moving it to any stratum which will reduce a weighted multivariate within stratum sum of squares, while continuing to satisfy constraints on strata sizes. The algorithm converges at a local optimum when moving any one of the units would increase the within strata sum of squares. Based on the findings of Judkins and Singh (1981), local optima were improved upon by the use of a moderately large number of random starts (i.e., 30).

For the contiguity option, a matrix is inputted specifying for each unit, all others contiguous to it. Contiguous initial strata respecting the size constraints are then built up from units chosen as random starting points. During the optimization step, an extra condition is imposed on transfer of units that contiguity be maintained. To achieve compactness, population centroids (longitude and latitude) are added as variables in the weighted sum of squares to be minimized.

For both NSR and SR areas, a multivariate stratification has been carried out using 1981 Census data for up to 17 stratification variables. Population variables include: total employed; employment income; persons with secondary education; population $15 +$, $15 - 24$ and $55 +$; and labour force in agriculture, forestry-fishing, mining, manufacturing, construction, transport, and services. Dwelling related variables include: total dwellings, dwellings rented, one person households, and two persons households.

Any industry variables accounting for less than 2% of the labour force of the area being stratified were dropped and other variables were scaled to be equally important in the optimization process. Unemployed was not included as a stratification variable due to its instability. Study findings showed that strata formed without unemployed when evaluated at the next Census were more efficient not only for other characteristics, but for unemployed as well. The inclusion of the neighborhood type variables, on the other hand, did result in improved efficiency for unemployed.

5.2 Two Level Stratification in SR Areas

In the larger SRUs with sample sizes of 300 or more households, two levels of stratification were adopted. Primary strata, with sample yields of $150 - 170$ households from the area and apartment samples combined, are comprised of collections of geographically contiguous Census Tracts. As such, primary strata are designed to correspond to two interviewer assignments. Three to four non-geographic secondary areal strata each yielding six or a multiple of six sampled clusters are formed within primary strata, with Census Tracts as stratification units, and with optimization based on the 1981 Census characteristics of non-apartment dwellers.

Apartments are sampled separately, in the form of a pps systematic sample from an open-ended list, which generally comprises a single stratum for the entire SRU. Sorting of the apartments existing at the time of design by primary strata yielded an implicit geographic stratification to the apartment sample.

In the smaller block-faced SRUs which warranted neither separate apartment sample nor geographic primary strata, optimal non-geographic areal strata were formed directly. In the non-block-faced cities, with considerably less scope for stratification, simple geographic strata were opted for.

The two levels of stratification in the larger SRUs had appeal on both operational and technical grounds. The relaxing of geographic constraints over those existing in the old design permitted greater optimality to be achieved, while the retention of contiguity at a higher level will provide a suitable unit for sample updating purposes later in the decade, and will facilitate the planning of interviewer assignments. Also, in the old design, SR strata were likely to be covered entirely by a single interviewer, and hence the variance estimates did not reflect the correlated response variance component of total variance. To the extent that within strata, interviewer assignments are geographic and secondary strata are non-geographic, an interpenetration of strata and interviewer assignments will be achieved in the new design without incurring any additional data collection costs, resulting in this component being better reflected in the variance estimates.

Table 1 presents study results for two SRUs — Ottawa and Quebec City — comparing efficiencies of the geographic strata used in the old design with those of optimal two-level strata, formed using 1971 Census data. Percent reductions in the first stage variance due to stratification, calculated at the time of the 1981 Census, indicate largest improvements under the optimal stratification for income and rented dwellings. The only marginal gains for other characteristics including employed and unemployed point to the strength and robustness of the simple, but deep, geographic stratification in the old design.
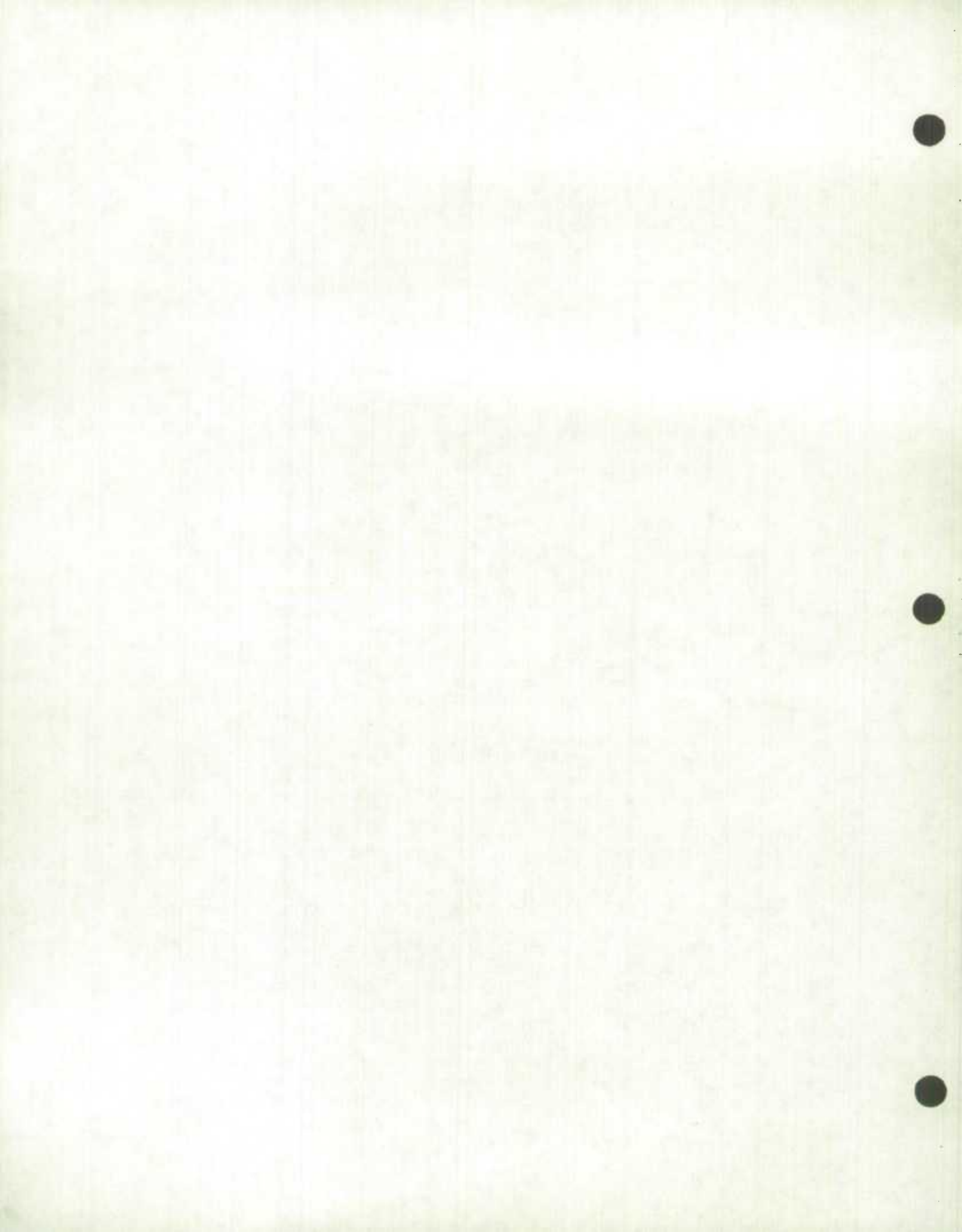
Table 1

% Reduction in First Stage Variance Due to Stratification
Old vs New Methods

Variable	Stratification Method		Variable	Stratification Method	
	old	new		old	new
total employed	9.1	12.6	agriculture[1]	5.9	3.9
employment income	18.1	30.4	forestry/fishing[1]	3.1	2.4
secondary education	39.4	42.1	mining[1]	4.8	3.0
population 15 +	9.2	12.6	manufacturing	23.5	23.1
population 15 – 24	12.9	17.6	construction	11.9	11.2
population 55 +	25.3	29.7	transport	4.2	6.4
total dwellings	28.5	33.1	services	14.5	19.8
dwellings rented	20.9	28.8	unemployed[1]	7.1	9.7
1 person households	33.7	38.4			
2 person households	27.5	29.6			

[1] characteristics not used in optimization for new method

In the NSR areas, the same clustering algorithm was used within each Economic Region to form either rural, or mixed urban and rural strata, depending on the design adopted, as discussed in Section 6.3. Also the adaption of the clustering algorithm for use in PSU formation is described in Section 6.5.

## 6.   DESIGN CONSIDERATIONS IN NSR AREAS

6.1 Extension of Telephone Interviewing

Telephone interviewing for months 2-6 in the sample was introduced during the early 1970's in Self Representing Areas, primarily to reduce cost. However in NSR areas, all interviewing continued to be done in person due to concern over the high instance of party lines vis-à-vis the confidentiality of the data being collected. Nevertheless, in recognition that not only immediate cost benefits from telephoning were at issue, but so also were the longer term potential benefits from the use of new technologies such as Random Digit Dialing and Computer Assisted Telephone Interviewing (CATI), it was decided to test the feasibility of extending telephone interviewing to NSR areas.

A first field test was restricted to urban areas having over 80% private lines. The test was conducted on a portion of the actual LFS sample, with the principal objective of assessing the data quality implications of telephone interviewing. To facilitate this analysis, interviewer assignments were split between the telephone and personal procedure.

This test ran from January 1982 to June 1983 with a gradual phase-in to ensure no adverse impact on the ongoing survey. Principal findings were: lower non-response rates for the telephone sample (3.4% versus 4.3% for the control sample); a high instance of households with telephones (96% for all provinces but one); a low instance (1%) of households not agreeing to telephone interviewing; and no detectable differences in estimates for labour force characteristics.

A second test carried out in the rural areas had comparable findings. Based on the positive findings from both tests, the decision was taken to introduce telephone interviewing across the board in NSR areas during the remainder of 1983 and early 1984.
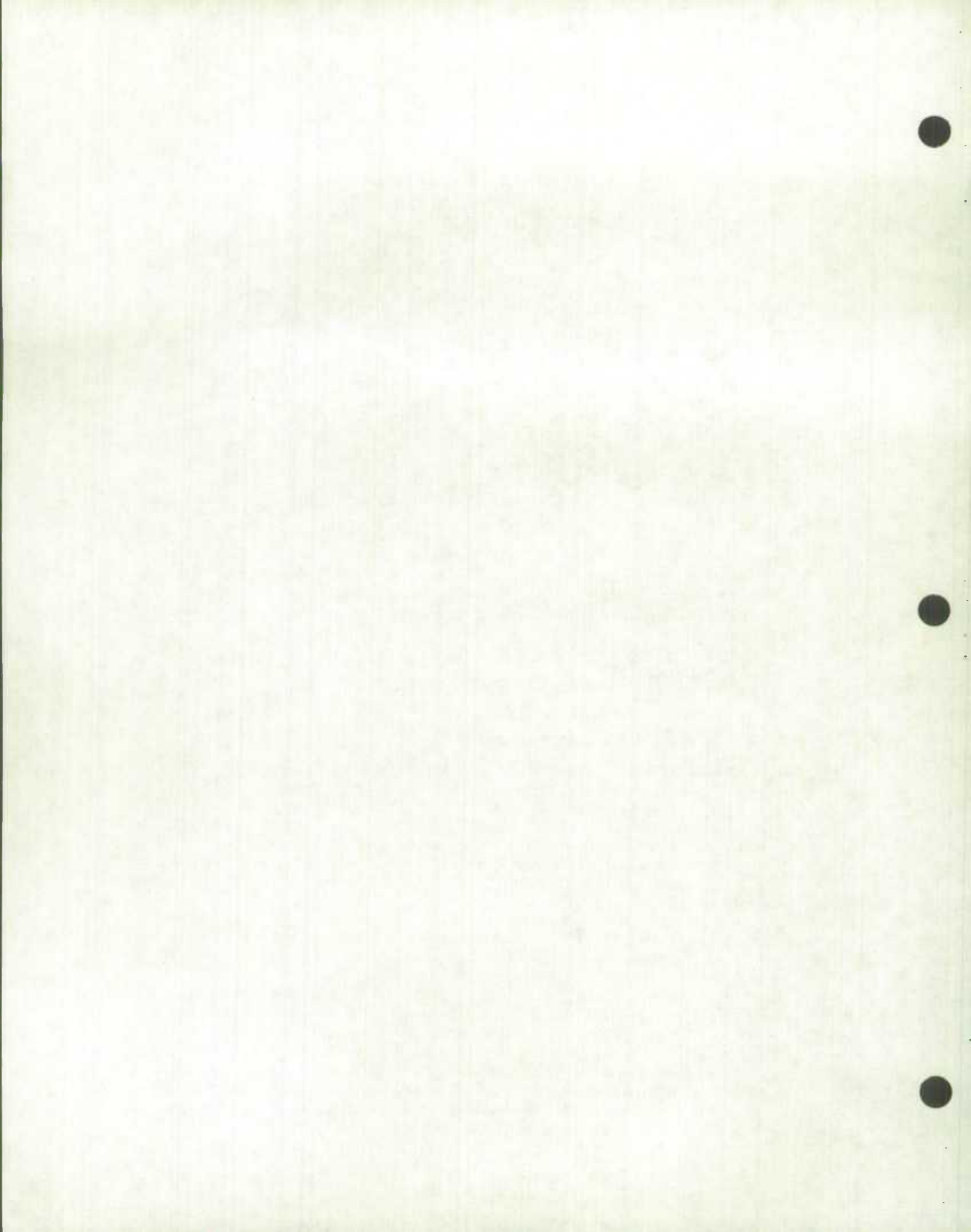
The decision to extend telephone interviewing had the following principal implications on the design of the NSR sample:

(i) Increase in assignment sizes: In the old design, NSR assignment sizes averaged 50 dwellings. Evidence that per unit costs were lower for larger assignments (Lemaître 1984), and the reduction in travelling under telephoning, both supported increasing the design yield per NSR PSU to 55 - 60 dwellings.

(ii) Level of assignment of rotation numbers: Unlike the old design, in the new design, all dwellings within secondaries will receive the same rotation number, which will cut down on the number of visits to the secondaries in month 2 - 6 in the sample.

6.2 Elimination of Stage of Sampling in Rural Areas

In the old design, the rural sample within PSUs was selected in three stages: secondaries (Census Enumeration Areas), clusters, and dwellings. The clusters corresponded to identifiable land areas containing up to 20 dwellings, which were delineated on the basis of field counts obtained whenever a new secondary entered the sample. Within secondaries, generally 5 - 6 clusters, with 3 - 4 dwellings per cluster, were selected.

The rural cluster stage was identified early in the redesign program as a possible candidate for elimination, on the grounds that (i) the sample variance would be reduced due to having one less stage of samplings, and (ii) the lead time required to introduce new secondaries into the sample could be shortened from 13 months to 7 months.

A field study was carried out on a sample of secondaries entering the LFS sample, in order to assess the feasibility of maintaining good quality dwelling lists for entire rural EAs, and to examine costs under such a procedure, with positive results on both counts. The variance implications of eliminating the cluster stage were also studied. Using 1971 Census data to simulate both the old and alternative design, components of variance were obtained for the Horwitz-Thompson estimator without ratio estimation. The percent reduction in total variance under the alternative design was found to range from 20 - 25% for major labour force characteristics (Choudhry, Lee, and Drew 1984).

On the basis of these findings, an early decision was taken to eliminate the rural cluster stage of sampling, and attention was turned to more global aspects of the NSR design.

6.3 Design with Urban/Rural Stratification

The old design featured implicit urban/rural stratification. PSUs were formed to have approximately the same ratio of urban to rural population as the stratum, and within selected PSUs the urban and rural portions were sampled independently. A premise underlying the design was that the PSU should correspond to an interviewer's assignment. However, in practice this correspondence was weakened since in order to attain the desired urban/rural ratio, frequently the urban and rural portions of PSUs were not contiguous.

As an alternative to the old design, D_0, (with the rural cluster stage eliminated), a design, D_1, featuring explicit urban/rural stratification was studied. Like D_0, the alternative design D_1 consisted of 3 stages of sampling in both urban and rural areas. In urban strata, the stages were: PSUs (consisting of individual or nearby urban centers), clusters, and dwellings. In rural strata, the stages were: PSUs, (consisting of collections of nearby rural EAs), secondaries (EAs), and dwellings. Under D_1, both urban PSUs and rural PSUs were designed independently to yield samples corresponding to interviewer assignments.

The two design alternatives were evaluated, from the point of view of variance and cost (Choudhry, Drew, Lee 1984). In the variance study both designs were simulated for the case of 2 PSUs per stratum using design counts based on the 1971 Census, and study variables based on 1976 Census data.

In terms of costs, a simple model was developed for D_0, the old design under telephone interviewing, and components were estimated using results from a detailed Time and Cost Study (Lemaître 1983). Relative costs for the travel components between designs D_0 and D_1 were estimated by means of a simulation study, in which average dispersion of the sample under the two designs was obtained up to the second stage of sampling using the population centroids of the EAs.
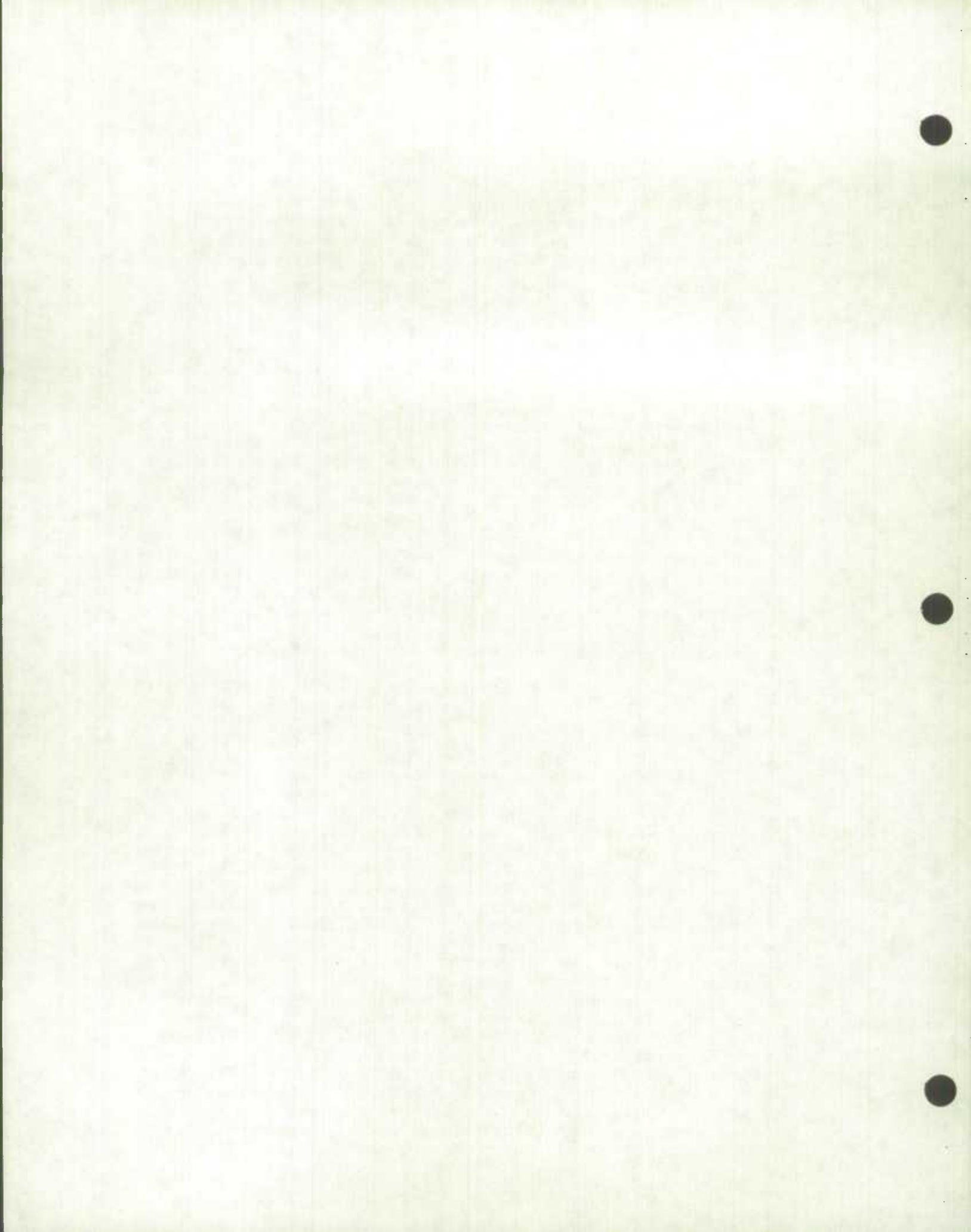
Findings were that the design D_1 was 1.09 times as cost efficient as D_0, and that from the combined perspective of cost and variance, D_1 outperformed D_0 with overall efficiencies of 1.25 for employed and 1.05 for unemployed.

Based on these findings, design D_1 was adopted in 70% of Economic Regions with sufficient urban and rural population to yield separate strata. In the remaining Economic Regions, with the exception of Prince Edward Island, design D_0 was adopted (see Section 6.6).

6.4 Number of PSUs Selected Per Stratum

In the LFS design, since the sample yield per PSU is fixed, the number of PSUs selected per stratum also determines the number of strata. In over two thirds of cases, the urban, rural or combined strata within ERs yielded only enough sample for 2 or 3 PSUs. Further stratification in these cases was ruled out on the grounds that there should be at least 2 PSUs per stratum to permit unbiased estimation of variance.

The remaining ERs were stratified to the point of 2 or 3 PSU's per stratum. Estimated first stage variance reductions over the situation under the old design of from 3 - 6 selected PSU's per stratum were up to 14% for employed (Choudhry, Lee, and Drew 1984). The stratification was carried out using the clustering algorithm described in Section 5.

6.5 Use of Clustering Algorithm in Formation of NSR PSUs

In both the old and new LFS, stratification is carried out prior to formation of NSR Primary Sampling Units. PSUs are delineated within the stratum to be as similar as possible with respect to stratification variables, while being as geographically compact as possible. PSU delineation which was carried out using the clustering algorithm noted earlier, required minimization of geographic and maximization of the non-geographic variables.

6.6 Two Stage Design for Prince Edward Island

For Canada's smallest province, Prince Edward Island, sampling rates are 4% in order to produce monthly LF estimates with required levels of data reliability. In view of the high sampling rates, a less clustered design consisting of a two stage sample of EAs and dwellings, with deep geographic stratification was adopted. It was found to have marginally higher costs than D_0, however from the overall perspective of cost and variance, it came out well ahead with efficiencies of 2.21 and 1.11 for employed and unemployed relative to D_0 (Choudhry, Lee, and Drew 1984).

7. ESTIMATION

7.1 Final Stage Ratio Estimation

In the old LFS, a final stage ratio estimation was carried out by detailed province/age/sex cells. With the development within Statistics Canada of improved and more timely subprovincial population estimates (Verma, Basavarajappa, and Bender 1982), an intermediate ratio estimation step was studied in which survey estimates of population 15 + for subprovincial areas are ratio adjusted to external estimates prior to the usual final ratio estimation. Findings were that the procedure, while not impacting on the variances of provincial level data, resulted in variance reductions for sub-provincial areas ranging from close to 70% for employed to 7% for unemployed (Earwaker and Bélanger 1981). In practice a raking ratio procedure in which the two ratio estimation steps are iterated until both marginal controls are satisfied was adopted, beginning in 1983.

7.2 Improved Estimates for Household and Family Units

Paul and Lawes (1982) used LFS longitudinal data files, which link households over the six months in the sample, to demonstrate that non-response rates are higher amongst households with fewer members. For the old LFS, non-response adjustment consisted of re-weighting at local area levels. This was done without regard to household size, hence the resulting estimates of households and families by size had 1 - 3% biases. Another problem related to the inconsistency
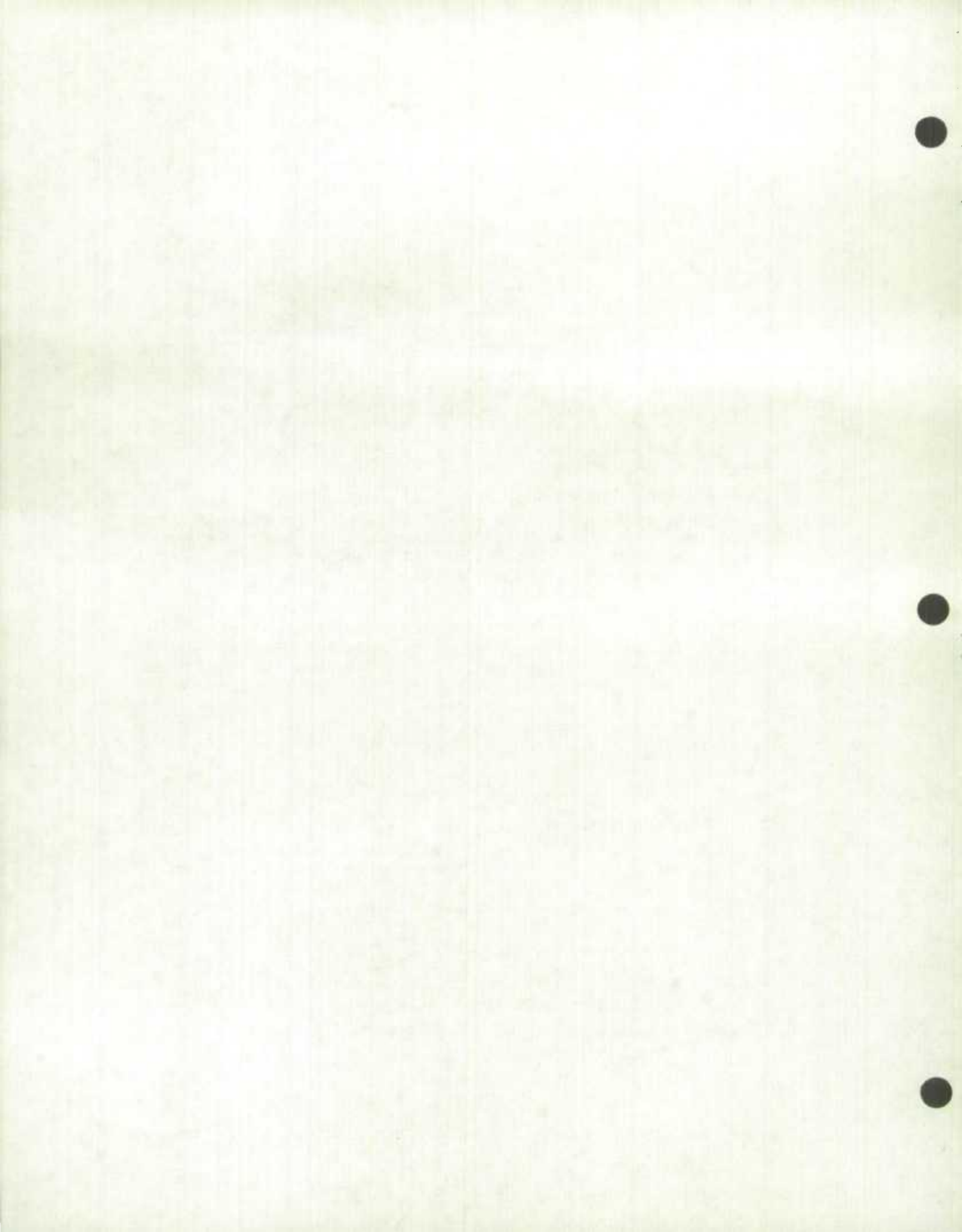
of family and individual based statistics (Macredie 1983). When demographic estimates of families by size, currently under development by Statistics Canada's Demography Division, become available, it is intended to incorporate them as an extra dimension in the final stage raking ratio estimation procedure, to address both problems.

As an interim measure, the use of LFS longitudinal data is being studied as a mean to derive household size distributions based on both respondents and non-respondents, prior to the final raking ratio estimation (Ghangurde 1984).

7.3 Small Area Estimation

Demand for Labour Force estimates for small areas (domains) such as Federal Electoral Districts (FEDs) and Census Divisions (CDs), both of which number over 250 units across Canada, has increased in recent years. Since it was not possible to respect the boundaries of such areas in the design of the survey, various alternative small area estimation methodologies were evaluated. A sample dependent estimator was proposed as a combination of post-stratified and synthetic estimators, which relies entirely on the post-stratified estimator whenever the sample size in the domain is sufficient according to certain criteria, and which otherwise introduces a synthetic component whose relative weight depends on the deficiency of the sample in the domain. Based on study findings, it was recommended that the sample dependent approach be developed as a means of providing annual or multi-year average estimates for areas such as FEDs and CDs (Drew, Singh, and Choudhry 1982). Implementation and further research and developmental work is proceeding under Statistics Canada's Small Area Data Program.

7.4 Variance Estimation

The methodology for variance estimation for the redesigned sample will continue to be based on Keyfitz's (1957) method, although it will be further modified to the case of a two step final stage ratio estimation, i.e., to a single iteration in the raking ratio estimation procedure. As subsequent iterations exert only a very small influence on estimates, they are being ignored in variance estimation. Some further refinements of the current variance estimation procedure are under study, such as adopting clusters as replicates in SRUs, as opposed to the current practice of grouping clusters into two psuedo-replicates.

It should be noted that variance estimators given by Rao, Hartley, and Cochran (1962), and by Rao (1975) were evaluated as alternatives to the current method in SRUs, where the RHC design is followed (Choudhry, Lee, and Sida 1984). The current method and the alternatives were studied both with and without ratio adjustment. The current method without ratio adjustment was found to overestimate the variance for certain characteristics (e.g., 20% for employed), however with ratio estimation, biases were negligible. Estimated biases were also negligible for the alternatives. The principal advantage of the alternatives was that they were more stable. The current method was retained however, due to its simplicity and also because of the complications in estimating variances of change or averages under the alternative methods.

7.5 Composite Estimation

In the LFS, moderate to high month-to-month correlations exist for most characteristics due to the 5/6'th common sample. Different composite estimators were studied by Kumar and Lee (1983), which take advantage of these correlations by use of data from previous samples to improve the current month's estimates. Their studies focussed on a class of AK composite estimators studied recently by Huang and Ernst (1981) and others in the context of the U.S. Bureau of the Census' Current Population Survey.

Findings under the assumption that the ratio estimator is unbiased, were that from the perspective of mean square error, a compromise choice of the A and K weights yielded up to 5% gains for monthly estimates of level for unemployed and employed, and from 5% – 16% gains for corresponding month-to-month change estimates. A decision on implementation of composite estimation was delayed pending further studies on the impact on composite estimators of any changes in rotation group bias, stemming from modifications in non-response adjustment and ratio estimation procedures, and pending closer examination of its operational implications.
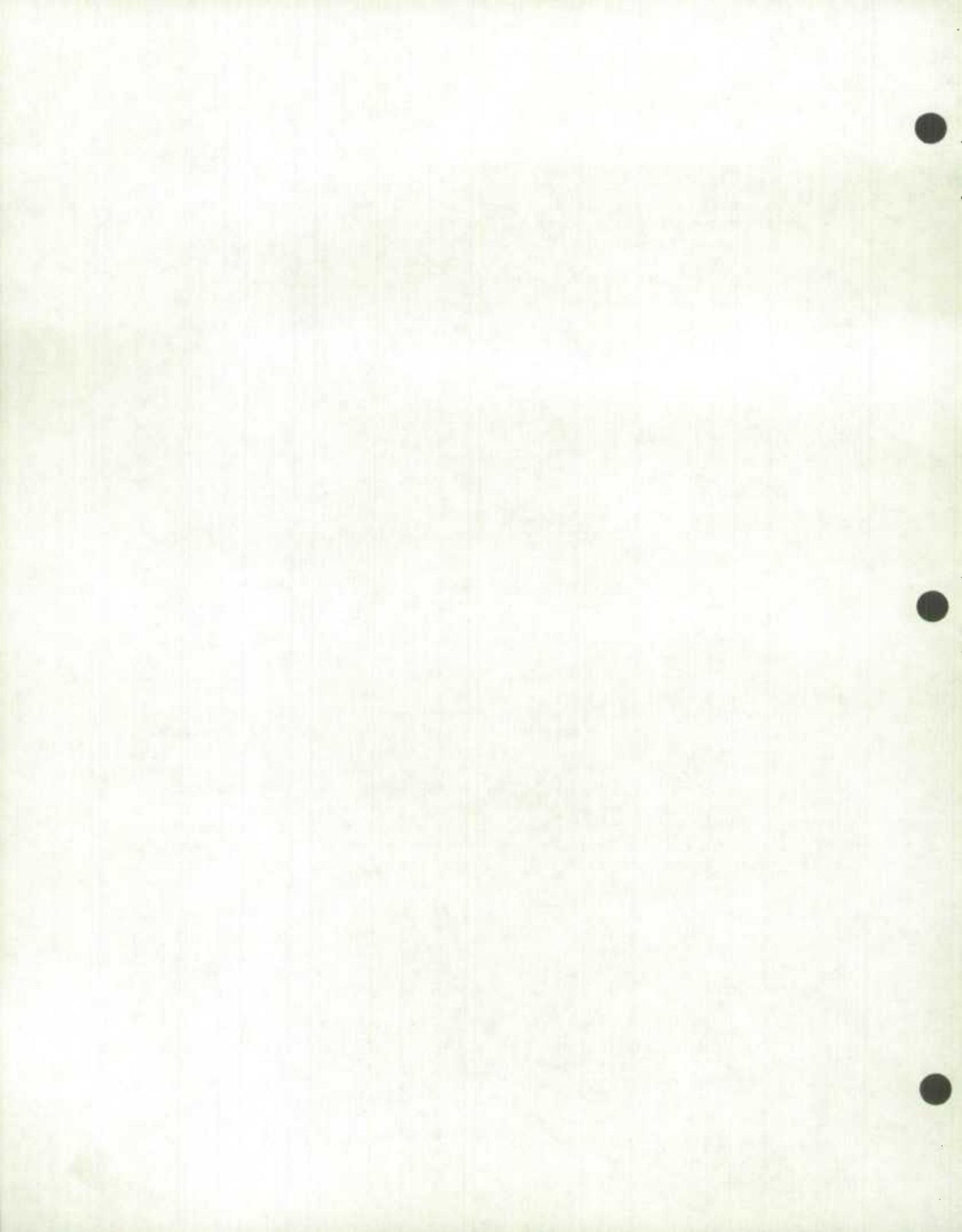
7.6 New Rounding and Release Policy

In the old LFS estimates of level were rounded to thousands and released if greater than 4 thousand. This policy was applied uniformly in all provinces for all estimates, with the intent that released data should have a coefficient of variation of 33.3% or less.

More rigorous, provincially based rounding and release criteria were developed for the redesigned sample to satisfy the conditions that the CV of unrounded estimates should be 33.3% or less, and that the rounding error should not exceed 20% of the standard error of the unrounded estimate. Findings were that release criteria could be dropped to 2 - 3 thousand, for all provinces except Quebec and Ontario, and that estimates for subprovincial areas should de rounded to hundreds instead of thousands (Kumar 1982).

8.   SAMPLE REALLOCATION

Specific data reliability objectives of the redesign having particular emphasis on better data at subprovincial level are noted in Section 2. In addition to the general improvements in the data reliability levels through the refinements in the methods and procedures, it became necessary to consider reallocation of the sample within provinces to meet the reliability levels noted in objectives (ii), (iii) and (iv). Sample size increases were needed in 13 out of 66 Economic Regions, 6 out of 24 CMAs and 27 out of 42 non-CMA cities. An average 28% reduction in the CV's for unemployed was obtained for these cases. In addition, for the 30 ERs with old CV's in the range of 15 - 25%, the reallocations achieved an average 12% reduction in CV's. As a rule of thumb, under the redesigned sample, monthly data for ERs and CMAs and quarterly data for other cities will be based on minimum monthly sample sizes of 300 and 120 households per month respectively.

It is worth noting that the redesign objectives did not directly consider two important uses of LFS data by federal government departments. These are the use of 3 month moving average unemployment rates for subprovincial Unemployment Insurance (UI) Regions in determining the regionally variable number of weeks worked to qualify for UI benefits, and the use of 3 year average data for 180 - 200 areas consisting of individual or combined Census Divisions, for use in allocating federal funds to assist new industrial initiatives. However, the re-distributions of the sample will indirectly benefit both of these applications.

In determining sample size requirements to meet the objectives, average unemployment rates for the period 1980 - 82 were used, in view of medium term forecasts for sustained high unemployment during the 80's.

A general implication of these reallocations was the movement of a significant proportion of the sample from larger CMA's and Economic Regions to smaller ones. This had an adverse impact on the provincial and national estimates due to the departure from the usual proportional allocations. This decrease in reliability at higher levels was however more than compensated by the general increase in the reliability achieved through the structural improvements in the methods and procedures as a result of research investigations.

A study was also carried out using the cost-variance model suggested by Fellegi, Platek and Gray (1967) to arrive at optimum sampling rates in the NSR and SR areas. This resulted in a shift of sample from NSR to SR areas . This was most pronounced in Quebec and Ontario where the proportion of SR sample increased from .60 to .72, as compared with .78 of the frame, yielding gains for provincial estimates of unemployed equivalent to a 5% variance reduction, assuming a fixed sample size. In addition, this optimization helped in achieving objectives (ii) and (iv), and it benefitted the Survey of Consumer Finances and Rent Survey.

It is worth noting that the structural improvements in the design achieved through factors such as improved stratification, reduction in a stage of sampling in the NSR areas, incorporation of subprovincial controls in the estimation procedures, and more refined sample allocations resulted in better than current reliability levels for national and provincial estimates while meeting the objectives for the subprovincial data. This opened up the possibility of reducing
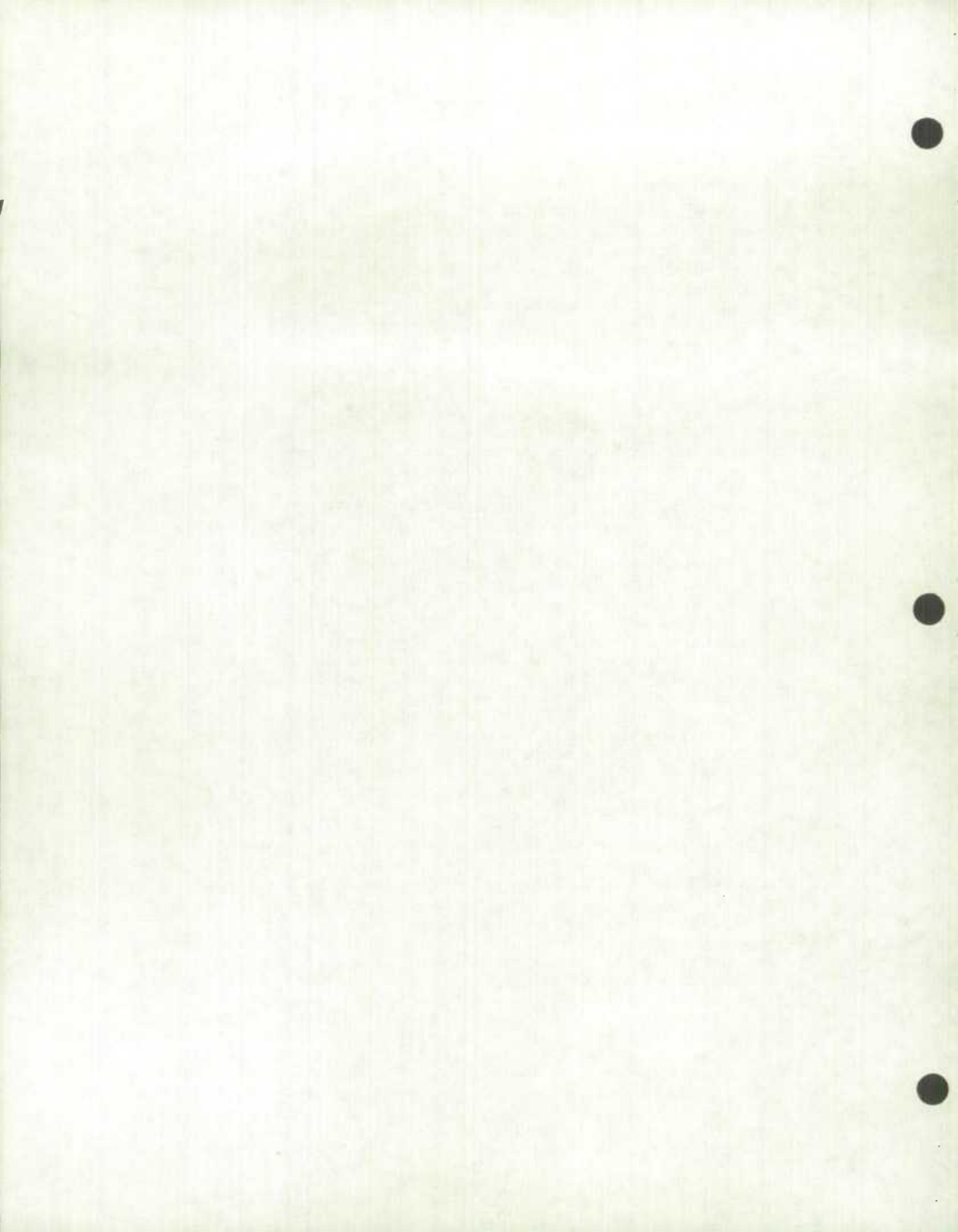
the overall sample of the LFS in order to release funds for the collection of data on certain other socio economic issues from time to time. The size of the redesigned sample was therefore fixed as 51,500 households per month down from 55,000. This overall reduction of 6 - 7% was achieved through a uniform reduction in all provinces with the exception of Prince Edward Island. In addition, per unit data collection costs will be reduced due to increased telephone interviewing.

9. IMPLICATION OF CHANGES ON LFS ASSOCIATED SURVEYS

Most of the household surveys conducted by Statistics Canada take advantage of the investment the LFS represents in terms of sample frame and design, data collection vehicle and processing systems to obtain data more quickly, at less cost and greater reliability than would be possible through independent surveys. The design and operations of these surveys are integrated with those of the LFS to varying degrees.

Most common are supplementary surveys consisting of additional questions to LFS respondent, which incur only incremental costs of the extra questions. Surveys which due to the sensitivity of the subject matter due to the length of the questionnaire, might impact on the LFS are not done as supplements. Typically such surveys have been conducted by LFS interviewers on a separate sample of households selected in the same areas as the LFS. Less closely integrated with the LFS are surveys which select different areas from the LFS, but which benefit from use of the LFS sample design and from control of overlap with the LFS sample.

As noted earlier the main orientation of the redesign program was towards the LFS, but efforts were also made to enhance flexibility of the vehicle for general applications. In this light changes being adopted for the LFS that will benefit its associated surveys are briefly highlighted below:

The sample reallocation resulting in a shift of sample from NSR to SR areas will be more robust for general applications and in particular will improve estimation of income and rent changes from the SCF and Rent Survey. Also, the adoption of a 300 household minimum sample size for CMAs will benefit these surveys for which CMA estimates are produced.

The general multi-variate stratification using 15 variables adopted (in both NSR and SR areas) will also represent an improvement for non-LFS applications over the current industry specific or simple geographic stratification.

Three changes will specifically benefit surveys using different sets of households: (i) the elimination of a stage of sampling in rural areas will considerably shorten the lead time to 7 months from 13 months in the old design, (ii) PSUs will be more geographically compact due to the adoption of explicit rural and urban strata, which will benefit smaller surveys where greater correspondence between PSUs and interviewers assignments is needed, and (iii) the flexibility introduced through the refinements in the sample stabilization program will allow selection of subsamples of virtually any size at the national, provincial or subprovincial levels for surveys using the LFS vehicle.

Finally the modification in the method of estimation introduced in the redesign, in the use of a household size distribution in the ratio adjustment as an interim measure (with eventual incorporation of demographic estimates of families by size as an additional dimension in the final staged raking ratio estimation procedure) will improve the consistency amongst family and individual related labour market statistics, and will improve family statistics on expenditure and income.

10. SUMMARY OF CHANGES AND FUTURE RESEARCH PLANS

Most of the research investigations for the post-1981 Censal redesign of the LFS have been completed, with the implementation of the new sampling design underway and certain aspects of research in estimation methodology still to continue. It is worth noting that many of the

investigations carried out during this program have confirmed the soundness of the past methods and procedures used in the LFS such as those of sampling two PSUs per stratum in the NSR areas, use of the RHC method and existing density factor (4 - 5 households per cluster) in the SR areas and continuation of the six month rotation pattern. However, several investigations have as well lead to improvements both in the redesign process and the new survey design; primarily due to change in emphasis on the data reliability objectives (as noted in Section 2), availability of better and easily accessible information and technological advances.

Improvements in the redesign process included the use of 1981 Census data in place of independently obtained field counts for updating the SR sample, the reduction of the clustering operation, and automation of stratification and PSU formation. Also cost savings will result from phasing-in much of the redesigned sample in an "on-line" fashion. Under this approach, the new sample will be introduced and the old sample will be dropped one rotation group at a time over a 6 month period, as compared with the traditional method of keeping the old sample at full strength for a 3 - 4 month period while building up the new sample (Mayda, Drew and Lindeyer 1984). Process cost savings as compared with the previous redesign are estimated at $1.8 million (in 1983/84 dollars).

Principal improvements in the cost efficiency of the LFS survey design include the extension of telephone interviewing to months 2 - 6 in the sample in NSR areas, the adoption of an NSR design featuring explicit urban/rural stratification, elimination of a stage of sampling in rural areas, the general purpose stratification in both SR and NSR areas, the use of subprovincial population controls in the estimation procedure, and more refined sample allocation procedures. These improvements were sufficient to permit gains in the reliability of subprovincial data, while retaining the status quo for provincial level reliabilities, and while decreasing the overall sample size by 6 - 7%. Reliability gains averaged 14% for coefficients of variation of unemployed for the half of the Economic Regions and CMAs with poorest reliabilities under the old design, with for the most part, little or no change in remaining areas. Subprovincial gains for estimates of employed will be even greater. On the cost side, the sample size decrease, coupled with reduced costs due to the extension of telephone interviewing will result in estimated cost savings of $.7 million per year (1983/84 dollars).

Following the completion of the sample redesign a principal focus of design related research and development for the LFS in coming years will be on investigation of a dual frame methodology, underwhich a portion of the sample would be converted to a telephone frame, using Random Digit Dialing (RDD) techniques. A multi-year program of RDD testing including research into implications of higher non-response to the RDD telephone sample, of research into dual frame estimation methodologies and of study of centralization versus decentralization of telephone interviewing is currently in the planning stages, as part of a newly established telephone survey development program (Hofmann, Drew, Catlin and Mayda 1984). Another design related initiative will be aimed at developing cost efficient means of intercensally updating the area sample in SR areas.
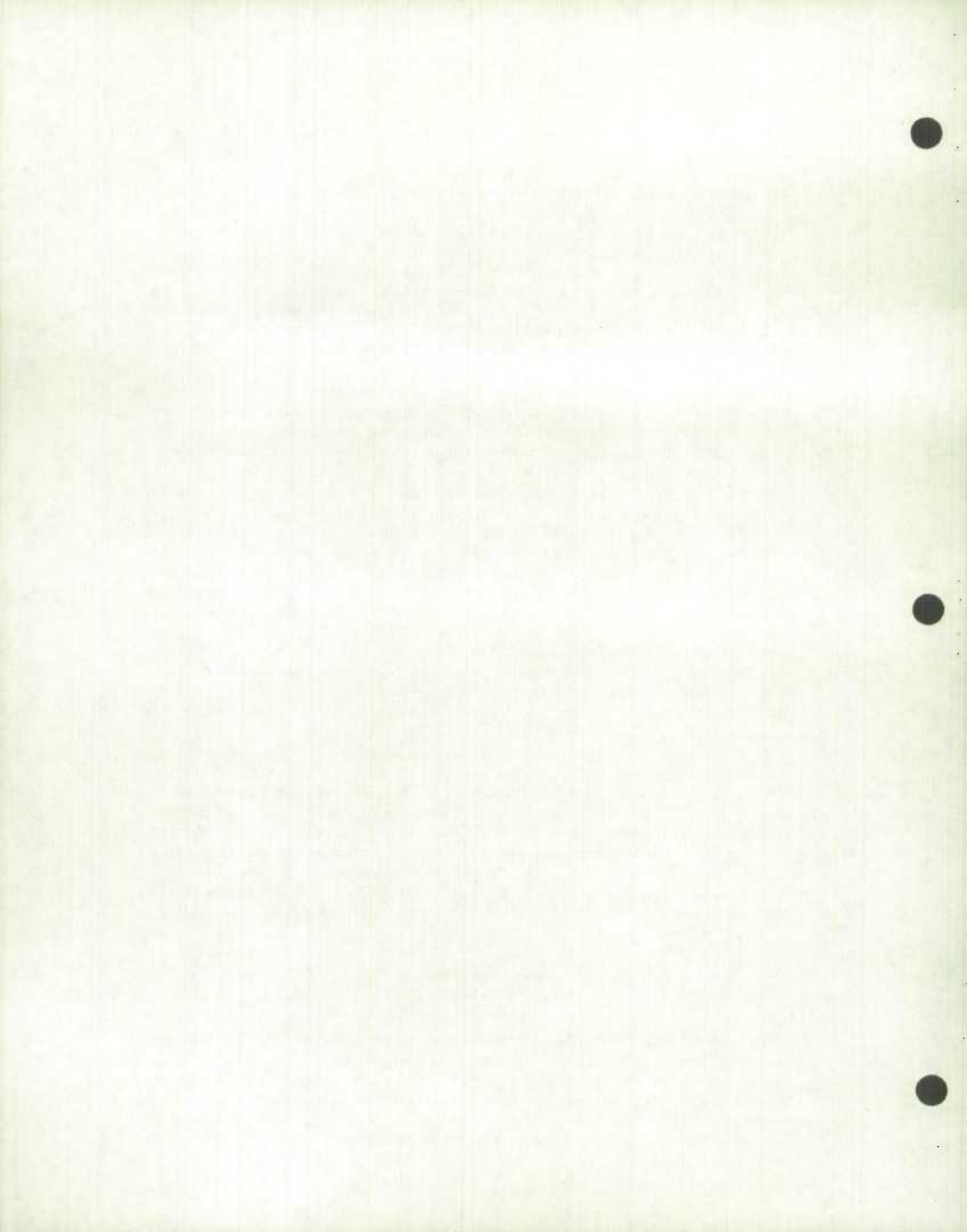
REFERENCES

CHOUDHRY, G.H., LEE, H. and DREW, J.D. (1984). Cost Variance Optimizations for the Canadian Labour Force Survey. Internal Technical Report (in preparation), Census and Household Survey Methods Division, Statistics Canada.

CHOUDHRY, G.H., LEE, H. and SIDA, R. (1984). Variance Estimation for the Redesigned Labour Force Survey. Technical Report, Census and Household Survey Methods Division, Statistics Canada.

DREW, J.D., CHOUDHRY G.H. and GRAY G.B. (1978). Some Methods for Updating Sample Survey Frames and Their Effects on Estimation. *Survey Methodology,* 4, 225-263.

DREW, J.D., SINGH, M.P. and CHOUDHRY, G.H. (1982). Evaluation of Small Area Estimations Techniques for the Canadian Labour Force Survey. *Survey Methodology,* 8, 17-47.

EARWAKER, S. and BÉLANGER, Y. (1981). Ratio Estimation at the Subprovincial Level for the Labour Force Survey. Technical Report, Census and Household Survey Methods Division, Statistics Canada.

FELLEGI, I.P., GRAY, G.B. and PLATEK, R. (1967). The New Design of the Canadian Labour Force Survey. *Journal of the American Statistical Association,* 62, 421-453.

FOY, P., BÉLANGER, Y., DREW, J.D. and JONCAS, M. (1984). Multivariate Clustering Algorithm for Stratifications and its Application to the Canadian Labour Force Survey. Technical Report (in preparation), Census and Household Survey Methods Division, Statistics Canada.

FRIEDMAN, H.P. and RUBIN, J. (1967). On some Invariant Criteria for Grouping data. *Journal of the American Statistical Association,* 62, 1159-1178.

GHANGURDE, P.D. (1984). Evaluation of LFS Non-Response Adjustment in Household Size Cells. Technical Report, Census and Household Survey Methods Division, Statistics Canada.

GRAY, G.B. (1973). On Increasing the Sample Size (No. of PSUs). Technical Report, Census and Household Survey Methods Division, Statistics Canada.

HARTLEY, H.O. and RAO, J.N.K. (1962). Sampling with Unequal Probabilities without Replacement. *Annals of Mathematical Statistics,* 33, 350-374.

HOFMANN, H., DREW, D., CATLIN, G., and MAYDA, F. (1984). A Proposal for a Telephone Survey Development Program. Internal Statistics Canada Report.

HUANG, E. and ERNST, L. (1981). Comparison of an Alternate Estimator to the Current Composite Estimator in CPS. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 303-308.

JUDKINS, D.R. and SINGH, R.P. (1981). Using Clustering Algorithms to Stratify Primary Sampling Units. *Proceedings of the Section on Survey Research Methods, American Statistical Association Meetings,* 274-284.

KEYFITZ, N. (1951). Sampling with Probabilities Proportional to Size: Adjustment for Changes in the Probabilities. *Journal of the American Statistical Association,* 46, 105-109.

KEYFITZ, N. (1957). Estimates of Sampling Variance where two Units are Selected from each Stratum. *Journal of the American Statistical Association,* 52, 503-510.

KOSTANICH, D., JUDKIN, D., SINGH, R. and SCHANTZ, M. (1981). Modification of Friedman – Rubin's Clustering Algorithm for Use in Stratified PPS Sampling. *Proceedings of the Section on Survey Research Methods, American Statistical Association Meetings.*

KUMAR, S. (1982). Investigation of the Labour Force Survey Rounding and Release Criteria. Technical Report, Census and Household Survey Methods Division, Statistics Canada.

KUMAR, S. and LEE, H. (1983). Evaluation of Composite Estimation for the Canadian Labour Force Survey. *Survey Methodology,* 9, 178-201.

LEMAITRE, G. (1983). Some Results from the Time and Cost Study. Technical Report, Census and Household Survey Methods Division, Statistics Canada.

MACREDIE, I. (1983). Family Oriented Measures of Employment and Unemployment. Presented at OECD Working Party on Employment and Unemployment Statistics.
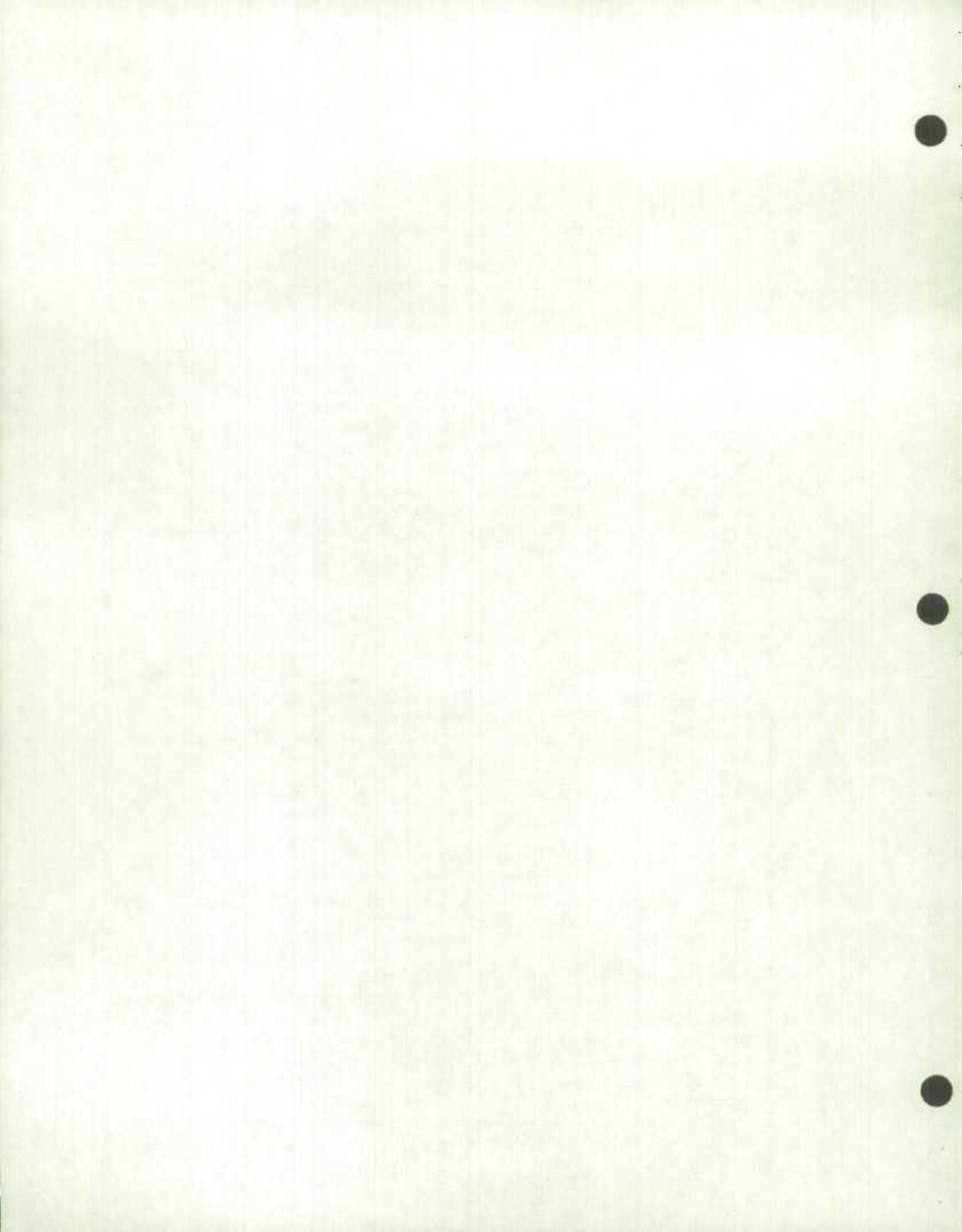
MAYDA, F., DREW, D., and LINDEYER, J. (1984). Phase-in of the Redesigned Labour Force Survey Sample. Technical Report, (in preparation), Census and Household Survey Methods Division, Statistics Canada.

PAUL, E.C., and LAWES, M. (1982). Characteristics of Respondent and Non-Respondent Households in the Canadian Labour Force Survey. *Survey Methodology*, 8, 48-85.

PLATEK, R. and SINGH, M.P. (1976). Methodology of the Canadian Labour Force Survey. Catalogue No. 71-526, Statistics Canada.

PLATEK, R. and SINGH, M.P. (1977). A Strategy for Updating Continuous Surveys. *Metrika*, 25, 1-7.

RAO, J.N.K. (1975). Unbiased Variance Estimation for Multi Stage Designs. *Sankhya*, Series C, 37, 133-139.

RAO, J.N.K., HARTLEY, H.O. and COCHRAN, W.G. (1962). On a Simple Procedure of Unequal Probability Sampling without Replacement. *Journal of the Royal Statistical Society, Series B*, 24, 482-490.

SINGH, M.P. and DREW, J.D. (1977). Sample Expansion in Self Representing Units of the Canadian Labour Force Survey. Technical Report, Census and Household Survey Methods Division, Statistics Canada.

SINGH, M.P., and DREW, J.D. (1981a). Research Plans for the Canadian Labour Force Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association Meetings*.

SINGH, M.P. and DREW, J.D. (1981b). Redesigning Continuous Surveys in a Changing Environment. *Survey Methodology*, 7, 44-73.

VERMA, R.B.P., BASAVARAJAPPA, K.G. and BENDER, R.K. (1983). The Regression Estimates of Population for Sub-Provincial Areas in Canada. *Survey Methodology*, 9, 219-240.