SURVEY METHODOLOGY/TECHNIQUES D'ENQUÊTE 1981, VOL. 7 NO. 1

REDESIGNING CONTINUOUS SURVEYS IN A CHANGING ENVIRONMENT

M.P. Singh and J.D. Drew[1]

Survey organizations undertake periodic redesigns of continuous surveys. Reasons for such redesigns related to changes in information needs to be satisfied by the survey and changes in public awareness and attitudes towards surveys are discussed in the context of the redesign of the Canadian Labour Force Survey following the 1981 Census. In particular, the importance of close dialogue between users of the survey data and design statisticians at the early stages of the redesign process in order to establish survey objectives is stressed.

1. INTRODUCTION

Data from decennial censuses in addition to serving the need of their primary users, serve as one of the frequently used tools in designing new surveys and by far the most important tool for redesigning (designing) large scale continuous surveys of population and housing. For instance, the Canadian Labour Force Survey (CLFS), a monthly survey of 55,000 households across Canada [12], has been redesigned following each decennial Census. Two of the primary reasons for these post-censal redesigns are to update the sample design to reflect changes in population characteristics and boundaries of Census units, and to incorporate improved methodologies such as sample selection and estimation procedures. Also the redesigns provide a unique opportunity to respond to a) changes in information needs to be

---

[1] M.P. Singh and J.D. Drew, Census and Household Survey Methods Division, Statistics Canada.

satisfied by the survey, and b) changes in public awareness and attitudes towards surveys and other factors affecting data collection. In this paper, the discussions are focused on items a) and b) in the context of the redesign of the CLFS following the 1981 Census. They may be found relevant for other similar surveys.

With regard to information needs, it should be mentioned that at the time of the revision of the CLFS carried out during the 1970's [10], a great deal of emphasis was placed on more data and increased reliability of data at the provincial level. However, new and important uses of CLFS data have emerged since that time, such as the legislated use of CLFS estimates in determining the eligibility for benefits under the Unemployment Insurance Program, administered by Employment and Immigration Canada. The redesign currently being planned for will represent the first occasion to consider such new data requirements. Along with the uses of labour force data, uses of the Labour Force Survey vehicle itself for obtaining other socio-economic data have greatly increased in recent years. After briefly discussing the process of identifying the survey objectives in general terms, the discussion in section 2 will cover the following three specific situations for meeting data demands:

i) Improved monthly data at sub-provincial levels through reallocation of sample within the provinces.

ii) Reliable data on quarterly and annual basis at smaller levels of aggregations through alternate rotation patterns.

iii) Increased demand for socio-economic data vis-à-vis current survey capacity.

The third section of the paper discusses the impact of the data collection method-item (b)- on development of the design of the survey, and emphasizes the importance of timely decisions on the procedure to be

adopted. The seventies have witnessed significant changes in the general conditions in which surveys have been undertaken, including increased respondent resistance and sensitivity to response burden, emphasis on voluntary surveys, increased incidence of proxy response, relatively higher travel costs, and as a result increased and more refined techniques of using telephone in conducting surveys, mail surveys, etc. In such an environment, it is essential for a continuous survey to maintain a capacity for testing various new options (or to study the effects of changing conditions) with the view to developing and maintaining a cost-efficient design. Basic requirements for such a capacity together with the planning for a telephone experiment and its implications on the current design are discussed in this section.

Lastly, some projects related to updating of the sample and development of improved methodologies [16] are highlighted in the final section.

It should be noted that research into the areas of alternative sample allocations and rotation patterns discussed in sections 2.2 and 2.3 are in very initial stages. For that reason, the tables presented are the results of preliminary investigations only, but they do indicate various possibilities depending on the requirements and priorities of users for the survey data.

2. DATA REQUIREMENTS

2.1 Establishment of Information Needs

It has now become a standard practice when designing a new survey or redesigning an existing survey to determine the information needs for establishing survey objectives at an early stage in the project. For a continuous survey, failure to determine these needs on the assumption that general objectives have remained unchanged would defeat one of the most important purposes of the redesign exercise, namely evaluation of the survey from the viewpoint of its uses and effectiveness.

During the period between successive redesigns, information needs of users who participated in the earlier setting of objectives may have changed, and in addition new uses of the survey data other than those considered in the design of the survey may have emerged. It is incumbent upon the methodologists and the sponsors to discuss jointly with survey users the detailed information needs and priorities and to set up objecives for the redesign in clear and specific terms.

It is only when specific objectives have been agreed upon, that the design statisticians can properly discharge their responsibility of developing the most efficient survey design taking account of the operational constraints and the cost specified for the survey. The importance of this close dialogue at the initial stages cannot therefore be overemphasized. This is particularly so in the case of large-scale continuous surveys where major changes cannot usually be incorporated in midstream due to such factors as continuity of the time series, complexities of operations, and cost benefit considerations. Hence any failure at the outset in establishing survey objectives may continue to affect the survey results for the life of the design.

In an environment of fiscal restraint such as currently exists, a seemingly legitimate concern on the part of the survey sponsoring agency may be that initiation of discussions with the users might spark the type and degree of demands which could reach well beyond the scope of the survey. However,as long as the importance of budgetary constraints are clearly realized by both, there should be a definite advantage to such discussions. Not only would they enable the statisticians to take stock of and prioritize demands, but also they would serve to identify and inform users of those requirements which cannot be met by the survey and alert them to the consequences of misuse of the survey data. For fuller discussions on the role of user consultations in analysis of requirements, and on identification of feasibility, priority and method, we refer the reader to a paper by Fellegi and Ryten [5].

Another point to be emphasized is the lead time required in redesigning a continuous survey and the importance of input from survey users at the early stages of this process. In contrast with adhoc surveys where normally design activities may be completed in a couple of months to about a year, the lead time needed for continuous surveys is much longer. By the same count, payoffs as well as stakes are higher. For the CLFS redesign, while the detailed research plans are currently being formalized, some initial studies have already been in progress since the middle of 1980 and the introduction of a redesigned sample is scheduled for 1985.

As a beginning step in the process of determining information needs for the coming CLFS redesign, members of a recently established Evaluation Program For the LFS [15] will meet with all the major users of CLFS data and a sample of other users, for the purposes of identifying users' needs and how well the existing LFS satisfies these needs. Based on these findings, the design statisticians will intensively follow up cases where new information needs have emerged or earlier information needs have changed. Success of the redesign program thus becomes heavily dependent upon the timely specification of the requirements so as to provide the survey designers sufficient time to evaluate alternatives and choose the most appropriate strategy for a given situation.

The specification of information needs should include a specification of; characteristics of interest; the types of estimates required- rates, levels, changes in rates, or changes in levels; required frequency of estimates; cross classifications (if any) desired for the characteristic at different area levels of interest; and finally associated data reliability requirements.

The specifications should also include a description of the uses of the data, and their bearing on decision-making processes or allocation of funds. Equally essential is an assessment by the users of the importance

of the survey data for their program. Where information needs of a user are diverse, the user should also be asked to indicate priorities for them. Having received this input from users, overall priorities would be established and, subject to budgetary and other restrictions, would be translated into specific survey objectives. It should be emphasized that a primary responsibility of the design statistician in the process of user consultation is to provide assistance to the users in understanding what input is required of them and to provide technical guidance, for instance in the determination of reliability requirements, and identification of possible means of meeting their requirements.

It is worth drawing attention to a note by Platek [11], to papers by Cahoon, Kniceley and Shapiro [1], in which the importance of establishing clear survey ofjectives at an early stage of the survey has been emphasized.

To illustrate the importance of precise specification of survey objectives in deciding upon the choice of survey strategy, we present below three alternate means of meeting demands for more data. The choice of a particular strategy or combination of strategies should depend upon the type of data needed and the priorities set out for them. In the following sections alternatives will be presented with respect to reallocation of the sample (sect. 2.2), use of alternate rotation patterns (sect. 2.3) and lastly, expansion of the scope of the survey (sect. 2.4).

2.2 Sample Reallocation for Improved Monthly Subprovincial Estimates

Before discussing the implications of sample reallocation on data reliability, the expression for the coefficient of variation used in calculations is briefly discussed with relation to sample size, frequency of a characteristic, and design effect.

For the LFS, the coefficient of variation for monthly estimates for the characteristec unemployed (u), and for an area (a) of interest, can be expressed as

$$CV_a(u)\% = 100 \frac{SD_a(u)}{u_a}$$

where $$SD_a(u) = \left(\sum_{t \varepsilon a} F_t \ (W_t - 1) \ P_t \ p_t \ q_t \right)$$

where $\sum_t$ = sum over strata or collection of strata for which sample design and sampling rate are the same.

and
- P = estimated persons 15 years of age or over
- W = inverse sample rate ($= P/n$, where n = sample size)
- p = proportion unemployed
- q = $(1 - p)$
- F = design effect
- $U_a = \sum_{t \varepsilon a} P_t \ p_t$

From the above formulation, it can be seen that the reliability of estimates of level for a characteristic are primarily dependent on three factors:

i) Sample size: since $W = P/n$, other factors being constant, the CV% decreases proportionately to increases in the square root of the sample size. That is to reduce the CV in half, the sample size would have to be quadrupled.

ii) Frequency of the characteristic: the coefficient of variation is approximately inversely proportional to the square root of the proportion of persons having the characteristic. Thus for unemployed, the lower the unemployment rate, the larger the sample size required to obtain reliable estimates.

iii) Design effect: The design effect provides an overall comprehensive measure of the combined effect of all the design features, such as

stratification, multistage sampling and estimation. It is defined by the variance estimate obtained from the survey divided by the variance that would have resulted had the sample been taken in the form of a simple random sample of persons. The interpretation of a design effect of 2 for unemployed, would imply that, cost consideration aside, the sample design was only half as effective for measuring the characteristic unemployed as a simple random sample would have been. For the LFS, design effects are generally greater than one for most characteristics, due to the need for concentrating the sample in a relatively few selected areas as a means of reducing data collection costs.

Historically for the LFS, the characteristic unemployed has usually been considered of primary importance, and the total size and allocation of the sample have been determined to achieve specified reliability requirements for monthly estimates of unemployed. Prior to a sample size increase during the 1970's, the sample of 36,000 households was allocated with the primary objective of providing good monthly estimates for unemployed at the national level. When the sample was increased to 55,000 households, the additional sample was allocated on the basis of achieving more uniform reliability between provinces for unemployed. Because the increase was carried out after the redesign of the sample there was an additional restriction imposed by the sample design, namely that in Self Representing (SR) strata (i.e. larger cities) the sample could only be increased by multiples of the existing sample size, and in remaining (NSR) areas, increases had to be half-multiples of the existing sample size (i.e. 50%, 100%, 150%, etc.)

In increasing the sample size, uniform sampling rates by type of area (NSR and SR) within provinces were retained, as this provided an effective allocation scheme for improving provincial estimates for unemployed. Table 2.1 illustrates the impact of the increase on monthly CV's for the characteristic unemployed for the period Jan 75 to Dec 1980. The uniform sampling rates have the additional advantage of providing a good general purpose allocation considering the broad range of characteristics

on which information is collected, not only for the LFS, but by other surveys utilizing the LFS capacity.

Table 2.1 Pre-Increase and Post-Increase CV% for Monthly Estimates of Unemployed

Province	Post Increase Sample Size (households) (2)	% Increase (3)	CV% for unemployed	
			pre-increase (4)	post-increase (5)
Newfoundland	3056	70.30	8.44	6.23
Prince Edward Island	1418	200.00	18.12	9.61
Nova Scotia	4021	29.80	6.55	5.34
New Brunswick	4217	67.78	8.23	5.44
Quebec	8541	17.06	4.56	3.54
Ontario	10850	14.24	4.31	3.65
Manitoba	4719	141.34	11.13	6.55
Saskatchewan	5724	200.00	14.42	7.56
Alberta	6709	100.00	8.65	6.49
British Columbia	6124	42.20	5.76	4.99
Canada	55379	52.12	2.33	1.88

With provincial CV's currently at acceptable levels, there has been an increased demand for more reliable subprovincial data. In the remainder of this section we examine how the reliability levels for subprovincial monthly estimates of unemployed could be improved by means of within province sample reallocations.

A disadvantage of the self-weighting design (uniform sampling ratio) is that for subprovincial regions variability in population sizes translate

into variations in the reliability of estimates. Currently monthly estimates of unemployed are published separately for 47 LFS Economic Regions for which the CV's are 25% or less. The remaining 19 ER's have been collapsed into groups of 2-4 to ensure that the reliability levels for the groups meet publication criteria.

It has been recently determined [17] that an additional sample of approximately 3000 dwellings would be required to achieve a 25% CV for each of the individual ER's where collapsing is carried out. It was also shown that these dwellings could be achieved by reallocating samples from the larger CMA's in the respective provinces. Refinements on such within province reallocations are currently being investigated using more months of survey data in the calculations and also taking into consideration reliabilities of estimates for other subprovincial areas as discussed in section 2.3 .

For illustration purposes, below we consider what could be achieved by within-province sample reallocations for the province of Manitoba. Present reliability levels for Manitoba's 8 Economic Regions based on data for the period Feb 78 to May 79 are shown in column 5 of Table 2.2. It might be noted that currently ER's 65 and 68, and ER's 61-64 are collapsed for publication purposes. The sample was reallocated to NSR and SR portions of individual Economic Regions so as to minimize data collection costs while achieving a fixed CV (22%) for unemployed following a general approach suggested by Fellegi et al [4], for all the ER's except 67 (Winnipeg), for which the sample size had to be reduced by 288 households. It should be noted that under the sample reallocation the provincial CV remains virtually unchanged, although costs would increase somewhat due to heavier sampling in NSR areas.

Table 2.2 Within Province Sample Reallocation for Manitoba Economic Regions (period Feb 78 - May 79)

	existing sample allocation				reallocation of sample			
Economic Regions (1)	W_{NSR} (2)	W_{SR} (3)	Hhlds (4)	CV(u)% (5)	W'_{NSR} (6)	W'_{SR} (7)	Hhlds (8)	CV(u)% (9)
61	41.67	-	477	25.98	30.16	-	659	22.00
62	41.67	-	308	25.89	30.33	-	423	22.00
63	41.67	90.00	690	17.55	92.55	74.54	399	22.00
64	41.67	90.00	242	30.84	25.48	36.37	435	22.00
65	41.67	-	404	22.36	40.40	-	417	22.00
66	41.67	90.00	426	20.18	52.69	65.30	353	22.00
67	-	90.00	2030	7.73	-	104.88	1742	8.35
68	124.00[1]	90.00	142	36.44	33.37	92.70	288	22.00
Manitoba	43.27	90.00	4719	6.15	39.37	100.52	4716	6.19

[1] remote area sample

There are some potential problems with an allocation scheme optimized for the subprovincial estimates for the characteristec unemployed, that have yet to be fully addressed, however. For instance, it may be less efficient for other surveys utilizing the LFS capacity. While for other surveys the desired allocations could be achieved by sub-sampling the LFS selections, this would nevertheless reduce the sample size available to such surveys. Additionally the robustness of such an allocation against changes in the unemployment levels over time would have to be studied further.

2.3 Alternate Rotation Patterns

In a rotating panel survey such as the LFS, the monthly sample size determines the reliability of monthly estimates of levels and rates; however, it is primarily the rotation pattern which determines:

i) the reliability of estimates of change, whether month to month, quarter to quarter, or for a calendar month from one year to the next and

ii) the reliability of estimates obtained by combining monthly data to arrive at quarterly, semi-annual or annual estimates.

In general, rotation patterns which are better for i) are not as good for ii) and vice versa. Thus the choice of a rotation pattern should be governed by the relative priorities attached to these types of estimates.

At the time of earlier redesigns of the LFS, there was little demand for estimates of type ii) and therefore the choice of the current LFS rotation pattern has reflected a predominant importance for estimates of month to month change. Under the current LFS rotation pattern, households remain in the sample for six consecutive months, and each month one-sixth of the households rotate out of the sample and are replaced by new ones. This scheme is very efficient for measuring month to month changes as the 5/6 households in common from one month to the next results in moderate to high correlations between successive months' samples for most characteristics.

The same correlations between successive months' samples which are advantageous for estimates of change are disadvantageous for average estimates of level. As a result, the current LFS rotation pattern is not as efficient for quarterly, semi-annual or annual estimates of rates or level as some other schemes. It is of interest to compare the performance of the LFS rotation pattern for combining data over months and for estimates of month to month change with that of the Current Population Survey (CPS), the counterpart of the LFS in the United States. In the

CPS households remain in the sample for 4 consecutive months, are out for 8 months, and then rotate back in again for 4 more months. Thus each month 1/4 of the households rotate.

If we denote V_m as the variance of estimate for a given month m and V_{cm} as the corresponding variance for estimates obtained by combining data for c months, then the variance reduction factor due to combining data, K, is defined as:

$$K = \frac{V_{cm}}{V_m}$$

Similarly, if we let $V_{(m,\ m+1)}$ denote the variance for estimates of change between months m and m+1, then the variance reduction factor for month to month change, K', is defined as

$$K' = \frac{V_{(m,\ m+1)}}{V_m + V_{m+1}}$$

It should be noted that K' is approximately equal to (1 - the correlation coefficient between the months' estimates). Table 2.3 presents values of K and K' for the two rotation schemes for the characteristic unemployed. The smaller value for K for the CPS rotation scheme indicates that it is more efficient for combining data, while the smaller value of K' for the LFS rotation scheme indicates it performs better for estimates of month to month change. It should be noted that the figures presented in the table for the LFS are the result of preliminary investigations only [7], and results for the CPS are taken from [18].

Table 2.3 Comparison of LFS and CPS Rotation Schemes for Unemployed

	Variance reduction factor due to combining months data (K)			Variance reduction factor for month to month change (K')
	3 mo	6 mo	12 mo	
CPS	.50	.31	.20	.50
LFS	.67	.48	.29	.44

Research studies are planned to confirm the results of Table 2.3 for the LFS, to consider similar variance factors for a broader range of characteristics, and also to consider the implications on combined estimates and estimates of change for other rotation patterns such as 3 - 9 - 3 (three months in the sample, 9 months out, and 3 months in) and 1 - 2 - 1 - 2 - 1, (one month in, 2 months out, one month in, etc.)

To further illustrate the impact of rotation pattern on average estimates of level, Table 2.4 presents the sample sizes necessary to achieve 25% CV's for annual estimates of unemployed for individual Census Divisions for the LFS and CPS rotation patterns under two different allocation schemes.

The augmentation allocation is based on retaining the present sample allocation and adding to it whenever necessary to produce the required reliability level for individual Census Divisions (CD's). The reallocation strategy on the other hand is based on a complete reallocation of the sample to achieve the required reliability levels for CD's. On practical considerations, both of these are extreme options and are used only for illustrative purposes. The reallocation strategy in some cases would result in a deterioration of monthly provincial estimates, while the augmentation strategy is clearly too expensive. A comprehensive strategy taking into consideration annual reliability levels for Census

Divisions, simultaneously with monthly reliability levels at the Province, Metropolitan Area, and Economic levels is currently under investigation.

Table 2.4 Additional Dwellings Required Monthly for 25% CV for Annual Estimates for Unemployed for Census Divisions

Province	LFS ROTATION PATTERN		CPS ROTATION PATTERN	
	Augmentation	Reallocation	Augmentation	Reallocation
NFLS	--	--	--	--
PEI	--	--	--	--
NS	90	--	12	--
NB	7	--	--	--
QUE	2,520	--	1,055	--
ONT	2,300	--	952	--
MAN	3,578	1.411	1,887	--
SASK	2,080	--	877	--
ALTA	1,357	--	647	--
B.C.	370	--	184	--
Canada	12,302	1,411	5,614	--

It should be noted that the calculations are based on the assumption of the current LFS design and the unemployment level at the time of the 1976 Census. Further, assumed density factors of 1.5 and 3 are used for SR and NSR areas, and the variance reduction factors used are those given in table 2.3. Thus the figures in Table 2.4 should be considered only for the purpose of illustration and relative comparisons, as changes in any of the above factors including the design, will result in changed allocations.

The point clearly illustrated by Table 2.4 is that if sufficient priority is attached to improved quarterly, semi-annual or annual average estimates, then for cost-efficiency reasons, there would be a strong case for changes in the rotation pattern. In this event, apart from the theoretical investigations of various rotation patterns, including the study of the impact of rotation group biases on them, a detailed examination of response burden and other operational aspects would have to be tested through the field experimental capacity described in section 2.4. If, on the other hand, higher priorities are given to the estimates of month to month change, then the LFS rotation pattern should remain unchanged.

2.4 Current Survey Capacity

Recent years have witnessed an increased demand for more detailed labour force data and data on a wide variety of characteristics influencing the labour market situation. During the 1970's Statistics Canada successfully responded to these demands by undertaking a major survey revision [10] which included an expanded capacity for use of the LFS as a vehicle for conducting other surveys. The current redesign will provide the opportunity to re-evaluate the role of the LFS in this regard.

Since the LFS is the only continuous household survey program carried out by Statistics Canada, integration of other household surveys with the LFS is desirable in the sense that these surveys can take advantage of the investment the LFS represents in terms of sample frame, design, data collection, and processing systems to obtain data more quickly, at less cost and greater reliability than would be possible through independent surveys. With the increased flexibility and capacity of the LFS achieved through the revision and through methodological improvements made at the last redesign, demands for use of the LFS as a vehicle for conducting household surveys have continued to increase in recent years. Examples of such surveys include: Survey of Consumer Finances, Asset and Debt, Family Expenditure, Annual Work Pattern, Household Facilities and

Equipment, Student Identification, Job Opportunity, Travel, Education, Smoking Habits, and Leisure Time Activities. Integration of these occasional surveys takes three different forms.

First, in the majority of cases, these surveys are conducted as supplements to the LFS due to cost and timeliness considerations. In such cases the most commonly adopted procedure is to collect data during the same visit, immediately after the LFS interview. In the case of enquiries with longer questionnaires, such methods of data collection as drop-off/pick-up are also utilized.

A second level at which the LFS has been utilized by other surveys is to select a different set of households in the same sampled areas as the LFS and to utilize labour force interviewers but at a different time period from the LFS. This is somewhat more costly than a supplement, but nevertheless represents a considerable saving over an independent survey. Examples of such use include; the Survey of Consumer Finances in odd years, when the survey content is expanded to include in depth questioning, for instance on Assets and Debts, and the program of Family Expenditure Surveys which consists of a recall survey and a diary survey.

The other situation in which LFS frame has been used is to select an 'independent sample' from the LFS frame, but in areas not currently being sampled by the LFS. The advantages over a totally independent sample are saving in sample design and implementation costs and also the control to avoid overlapping with the LFS and surveys associated with it. The Canada Health Survey for instance followed this approach in its survey design in cities, although in the remainder of the country a separate design was called for due to unique operational constraints.

Currently along with the LFS redesign activities, methodological aspects of other major surveys are also being researched. Just as it is important for the primary subject of enquiry using the continuing survey vehicle to

re-evaluate and re-establish its own objectives, it is equally incumbent on the other major users of the vehicle to follow the same course of action.

This will provide an opportunity for such surveys to maximize to the extent possible their benefits from the redesigned capacity of the vehicle, by determining optimal designs for their surveys, by being aware of implications of redesign alternatives, and by providing input to the decision processes.

Sponsors of each such major survey and the associated methodologists have recently begun this undertaking. Major studies in the optimization process would include stratification, sampling stages, allocation at various level of aggregations, determination of sampling and sub-sampling fractions, rotation patterns, response rates and their adjustments and other factors in the estimation process. It is not unlikely that these studies would result in a collection of optimal designs differing to a varying degree for different surveys.

Depending upon the importance attached to the major surveys using LFS vehicle and the degree to which the optimal designs differ, one of the three options may be followed, namely:

a) to redesign the current vehicle as a continuing household survey primarily for the LFS taking account of other surveys to the extent possible,
b) to redesign the current LFS vehicle as a general purpose survey or
c) to redesign the vehicle only for the LFS and develop a separate vehicle for conducting other major socio-economic surveys

The current situation is somewhere between a) and b); design features are optimized for the LFS, particularly in Non Self Representing Areas, (sect. 4.2); nevertheless the capacity is used extensively in a general purpose sense, as has already been described. It will be noticed that there is a very fine distinction between the options a) and b) and that

the difference mainly lies in the degree of importance associated to various subjects of enquiry using the redesigned vehicle.

In order to illustrate the distinction between the two approaches let us consider the problem of allocation of sample at a given level of aggregation (say R). Suppose there are m enquiries (m = 1, 2, ..., M) with the corresponding optimum allocation as n_m at level R, with the LFS allocation being denoted by n_ℓ. Say their magnitudes are as follows:

$$n_1 \leq n_2 \leq \cdots n_\ell \cdots \leq n_m \cdots \leq n_M$$

indicating that the subject M requires largest sample at level R. Note that this may happen at level R even if the total sample size for subject M may be smaller than that of the LFS.

In the case of option b) the approach would be to aim at a compromise allocation (say n*) such that $n_1 \leq n^* \leq n_m$. In case of option a) however, the allocation would always be n_ℓ determined to be optimum for the LFS, and in order to accommodate other surveys, flexibility would have to be introduced into the vehicle to allow for over-sampling or sub-sampling as required. Option c) while having some technical merit, suffers from operational problems, such as co-ordination of the two vehicles to prevent overlapping samples. Even if such problems are taken care of, this option as such can be ruled out on the grounds of being very expensive unless some enquiries equally important and complex as the LFS come along.

Discussions and studies are being carried out in order to make a final decision on these options. It seems at this stage that the requirements of most surveys currently using the vehicle would be met under option a) by incorporating minor changes in various aspects of the LFS design and increasing the capacity of the survey vehicle as described below.

In order to meet the data requirements, studies are being undertaken to develop an alternate small scale survey capacity in addition to increasing the capacity for the current LFS vehicle. One component of this program would be a collection of statistics on new subject matters in anticipation of future requirements. This would thus serve as a 'pilot' for full scale enquiries for more detailed data. As well the small scale survey capacity would provide an opportunity for analytical studies to examine inter-relationships between various social and economic phenomena. The capacity may frequently be used for surveys where it is necessary to react quickly in response to data associated with policy concerns of the federal government. The third area where this capacity would be useful is the development of new techniques through well designed field experimentation. This last aspect is discussed in more detail in the following section along with data collection.

3. DATA COLLECTION

In a large scale survey, a single or a combination of data collection methods such as personal interview, telephone and mail may be used, depending upon the type of enquiry, available facilities, respondents' attitudes, resource situation and timing constraints. Whatever be the method adopted at the initial phase of a continuous survey, it requires regular review as changes in the environment and conditions under which data are collected will directly affect its quality. Over time, respondents' attitudes towards surveys may change due to changing life style or increased respondent burden; new tools and techniques may be developed; legal requirements, the resource situation or quality of interviewers may have changed. All these affect the quality of data collected and hence the choice of method. Although certain changes in the data collection procedure may be introduced at any time during the life of the survey design, major changes affecting the cost and quality are usually introduced along with the redesign of the vehicle. This is

because the procedure adopted for data collection affects both the type of sampling design as well as the estimation procedure, and hence to be cost effective the method of interview must be decided upon well in advance of the sampling plan. It should be noted that for a survey vehicle like the LFS, which is used by various types of enquiries, the effect of any change in the procedure of collecting data needs to be investigated, including testing in the field, for as many of the major enquiries as possible. In this respect, just as in the case of establishment of the survey objectives, close discussion and coordination among sponsor, field staff and methodologist are important at the very early stages of planning.

As mentioned in the previous section, a small scale capacity is being considered to meet the current needs of social statistics. One of the primary purposes of establishing this capacity is to provide an opportunity for testing and developing new operational and methodological procedures. Testing of alternatives will focus on the data quality through operational measures such as response rates, slippage, error rates, etc., and also the effect on the cost of the survey. This methods test capacity may be utilized depending on the purpose of the test in any of the following manners: use of same households as the LFS, separate set of households in the same area as the LFS, or a completely different sample. Also the purpose of a particular test will determine its duration, location and the spread of the sample.

It is expected that certain new methods and procedures will be tested in the field with a view to examining suitability for the ongoing LFS. As an example, one such test which deals with the extension of telephone interviewing in the rural areas and smaller urban centres is briefly discussed below.

After a period of testing, the use of the telephone interview was expanded at the time of the last redesign to cover all Self-Representing Units primarily to reduce the cost of data collection. Currently in the LFS, households are interviewed in person the first month they are in the sample. In Self-Representing Units, if the respondent agrees to the telephone, the interviews are as a rule conducted via telephone in the second through sixth month the household remains in the sample. In other areas interviews are conducted in person. A similar telephone interviewing procedure will be tested for NSRU's. However, due to concern over the confidentiality of the data, telephone interviewing will be restricted to areas with a very low incidence of party lines.

Objectives of testing telephone interviewing in NSR areas will be to determine for the LFS and other surveys using the vehicle:

i) Effect on data collection costs and sample design implications. Reductions in the travel component of collection costs and the potential for interviewers to handle larger assignments, could permit designs with less concentration of the sample, and hence a reduction in sampling variance. For instance, it might be possible to eliminate one or more stages of sampling.

ii) Data quality. Acceptance of telephone interviewing, effects on non response rates, and if possible on survey estimates would be examined.

The test would be conducted on a sub-set of the ongoing LFS interviewer assignments, augmented in some cases by 10 - 20 percent to study the effects of larger assignment sizes and different concentrations of the sample.

4. OTHER DESIGN RESEARCH

In this section, we briefly highlight some of the redesign projects related to updating the sample and introduction of methodological improvements in the sample design and estimation procedure.

4.1 Redesign of Self-Representing Units

Current LFS Self-Representing Units (SRU's) correspond to those cities which were sufficiently large to yield a sample capable of supporting at least one interviewer. Minimum SRU sizes vary from a population of 10,000 in the Atlantic Region to 25,000 in Quebec and Ontario. A first step then will be a re-definition of the SR universe taking into consideration impact of the 1976-77 sample size increase, population shifts, and changes in boundaries of Census Metropolitan and Census Agglomeration areas.

Larger SRU's are divided into sub-units and within sub-units, first stage sampling units (i.e. clusters), are delineated on the basis of field counts obtained in 1973. The clusters correspond approximately to city blocks. A two stage sample of clusters, and dwellings (3 - 5 per selected cluster) is selected following a pps method based on random groups of clusters [14] . Using census data to simulate the LFS design, research is being carried out to investigate the effects on sampling effciency and operational suitability of alternative first stage sampling units - such as census enumeration areas, blocks or block faces - and of alternative allocations of the sample between and within first stage units.

Another focus is on alternative means of achieving and maintaining an up-to-date sample in SRU areas. Due to the rapid and uneven growth which occurs in these areas, without regular updating, the variance of estimates can increase substantially [2]. Under the present sample updating program [13], [3], for sub-units being updated, revised dwelling counts for individual clusters are obtained on the basis of complete field counts. As an alternative to independently obtained field counts, the use of census units, dwelling counts and maps in the redesign of the sample is being investigated. Discussions are also in progress with Post

Canada concerning possible use of Post Canada maps and dwelling counts to provide a future means of updating the LFS sample without incurring the expense of field counts. The key to this would be the planned linking of Postal Codes to 1981 Census units, and hence to LFS sampling units.

4.2 Redesign of Non Self-Representing Units

Non Self-Representing Units correspond to the smaller urban centres and rural areas. In the present design, 1 - 5 geographically contiguous, approximately equi-sized strata are formed within the NSR portions of individual Economic Regions. Industry classifications were taken into consideration in forming strata. Within strata, approximately 15 Primary Sampling Units (PSU's) were delineated so as to be similar to the stratum with respect to stratification variables and rural to urban population ratios. To satisfy this last constraint, frequently urban centres had to be shared amongst several PSU's within the stratum, often resulting in discontiguity between rural and urban portions of PSU's.

Initially two PSU's per stratum were selected following the randomized pps systematic method [8]. The sample was increased by selecting additional PSU's [6], and at present 3 - 6 PSU's are selected per stratum. It is felt that the sample increase strategy adopted may have led to a reduction in the efficiency per unit cost of survey, although the circumstances of the increase occuring in midstream ruled out more technically desirable alternatives such as re-stratification to form an increased number of strata, each with two selected PSU's.

As data requirements and design constraints, both technical and operational, vary from province to province alternative designs will be investigated by provinces or groups of provinces taken together as opposed to seeking a uniform national design.

The NSR design is very much dependent not only on whether telephone interviewing is adopted as discussed in section 3, but also on the survey objectives. For instance, if an increased importance is attached to annual estimates for Census Divisions or to the estimates from other major surveys using the vehicle, then a design in which CD's were taken as primary strata would be seriously considered. In such a case, the design would likely feature rural/urban sub-stratification within CD's and utilization of either Census Sub Divisions or Census Enumeration Areas as first stage sampling units. Studies would be required to determine whether any loss in sampling efficiency for the LFS would be incurred under such a design, due to the reduced amount of optimal type stratification.

Additional studies in the NSR design which are planned, primarly to improve the design efficiency and facilitate updating include:

Buffer Areas

Generally growth in NSR areas is not large enough to warrant updating the sample between redesign. Exceptions, however, are the NSR areas close to the boundaries of certain Census Metreoplitan Areas. During the 10 year life cycle of the design, growth frequently reaches into these areas, where a more flexible design capable of being updated is therefore required.

Stages of Sampling

Studies will be conducted to determine the implications, both operational and theoretical of reducing the number of stages of sampling. This study would be closely linked to the study on telephone interviewing.

4.3 Estimation and Variance Estimation

A number of studies are planned into estimation and variance estimation procedures used by the LFS and other household surveys. Some of these

are briefly highlighted below:

Final Ratio Adjustment

The current estimation procedure for individuals incorporates ratio estimation at the province level, using official population estimates by age-sex categories, adjusted for out of scope population (military and institutional). Research will be conducted into determining optimal age-sex post-strata, applying the ratio estimation at sub-provincial levels, and adjustment of LFS data for census undercount.

Estimation for Family Units

In the past, post censal estimates of numbers of family units have been unavailable, with the result that there has been no standard procedure from one survey to the next for producing family based estimates. This project will address both of these problems, as well as attempting to ensure consistency between family and individual based estimates.

Variance Estimation

Research will be carried out to compare alternative estimators with the Keyfitz [9] estimator currently being used from a point of view stability and extent of bias in the current estimator due to the violation of the sampling with replacement assumption. This will be examined for both seasonally adjusted as well as unadjusted sample estimates.

Small Area Estimation

Research will continue into estimation methods for non-standard areas cutting across design strata. Estimators being studied include synthetic, composite, and sample regression. Attention is also being given to the treatment of large growth clusters falling into the sample, particularly as they affect estimates for small areas.

5. SUMMARY

While redesigning continuous surveys, the importance of close discussions among users, sponsor and designers at an early stage of the program is emphasized. This will not only help to re-evaluate the effectiveness of the ongoing program but it will be a useful exercise in identifying and informing the users about the limitations of the survey. As a result of such discussions the survey objectives can be established in the light of current and future data requirements. To illustrate the importance of the precise specification of the objectives, three alternate means of meeting data demands are discussed, namely reallocation, rotation patterns and survey capacity. Choice of these or other alternative would clearly depend on the specification of the information needs.

Like the specification of survey objectives, data collection procedure plays a very significant role in deciding the survey strategy for a particular situation. Designers aim at developing the most efficient design per unit cost and since the major part of the cost relates to the data collection, an early decision in this respect is essential. A small scale capacity is being developed to list and develop new procedures and it is planned to use this capacity in examining the suitability of telephone interviews in rural and smaller urban centres.

At present steps are being taken to establish the information needs and also to decide upon the field methodology. Several research and evaluation projects in the above context have been started. In addition research related to other aspects of the design and estimation methodology has begun.

RESUME

Les organismes spécialisés révisent périodiquement leurs enquêtes permanentes. Ces révisions tiennent à l'évolution des besoins en information auxquels l'enquête doit répondre et à l'evolution de la perception et de l'attitude du public à l'égard des enquêtes; elles sont analysées dans le contexte de la révision de l'enquête sur la population active du Canada après le recensement de 1981. En particulier, les auteurs font ressortir l'importance dès le début du processus de révision du dialogue entre les utilisateurs des données de l'enquête et les statisticiens concepteurs afin de déterminer les objectifs de l'enquête.

REFERENCES

[1] Cahoon, L.J., Kniceley, R.M. Jr. and Shapiro, G.M. (1980), "Informational Needs for Current Demographic Survey Design with Discussion of Key Redesign Research Projects", ASA Proceedings of the Section on Survey Research Methods, pp. 96-112.

[2] Drew, J.D., Choudhry, G.H. and Gray, G.B. (1978), "Some Methods for Updating Sample Survey Frames and their Effects on Estimation", ASA Proceedings of the Section on Survey Research Methods, pp. 62-71.

[3] Drew, J.D. and Singh, M.P. (1978), "An integrated Program for Continuous Redesign and Sample Size Reduction - A proposal for Self-Representing Areas", Technical Memorandum, Census and Household Survey Methods Division, Statistics Canada.

[4] 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, pp. 421-453.

[5] Fellegi, I.P., Ryten J. (1977), "An application of Functional Analysis - Current Trends in Statistics Canada", paper presented at the 25th Plenary Session of the Conference of European Statisticians.

[6] Gray, G.B. (1973), "On increasing the Sample Size (no. of PSU's)", Internal Technical Memorandum, Census and Household Survey Methods Division, Statistics Canada.

[7] Gray, G.B., (1979), "Sampling Variance", Internal Technical Memorandum, Census and Household Survey Methods Division, Statistics Canada.

[8] Hartley, H.O. and Rao, J.N.K., (1962), "Sampling with Unequal Probabilities Without Replacement", Annals of Mathematical Statistics, 33, pp. 350-374.

[9] Keyfitz, N., (1957), "Estimates of Sampling Variance where Two Units are Selected from Each Stratum", Journal of the American Statistical Association, 52, pp. 503-510.

[10] Petrie, D.B., (1973), "Project Review: Assessment and Revision of the Canadian Labour Force Survey", discussion paper presented at Federal Provincial Conference.

[11] Platek, R. (1979), Lecture Notes for Survey Methodology Overview Course, Statistics Canada.

[12] Platek, R. and Singh, M.P. (1976), "Methodology of the Canadian Labour Force Survey", Catalogue No. 71-526, Statistics Canada.

[13] Platek, R. and Singh, M.P. (1978), "A Strategy for Updating Continuous Surveys", Metrika, Vol 25, pp. 1-7.

[14] Rao, J.N.K., Hartley, H.O. and Cochran, W.F., (1962), "On a Simple Procedure of Unequal Probability Sampling Without Replacement", Journal of the Royal Statistical Society, 24(2), pp 482-490.

[15] Sayant, G., Pold, H. and Macredie, I., (1981), "Project Proposal - Program Evaluation Labour Force Survey - Summary", Internal Report, Statistics Canada.

[16] Singh, M.P., and Drew, J.D., (1981), Research Plans for the Redesign of the Canadian Labour Force Survey", for presentation at 1981 ASA Meetings, Section on Survey Research Methods.

[17] Tarte, F., (1981), "Improvement of Reliability for Economic Regions", Internal Technical Memorandum, Census and Household Survey Methods Division, Statistics Canada.

[18] U.S. Department of Commerce, Bureau of the Census (1978), "The Current Population Survey Design and Methodology", Technical Paper 40, Washington, D.C., U.S. Government Printing Office, pp. 64, 96-99.