Sampling and Estimation Issues in the Redesign of the Canadian Survey of Household Spending

Christian Nadeau, Johanne Tremblay, Suzanne Lessard
Statistics Canada, 100, Tunney's Pasture Driveway, Ottawa, Ontario, K1A 0T6, Canada

Abstract

The Survey of Household Spending is conducted by Statistics Canada in the first quarter of each year. Respondents are asked to report their expenditures over the previous calendar year through a retrospective interview. A major redesign has recently been undertaken. The new model will allow the collection of frequent expenditure data from a diary, while less frequent expenditures will be covered by an interview using recall periods varying from one to twelve months. This approach will be tested via a pilot survey before being fully implemented in January 2010. A number of sampling and estimation issues that are related to the use of different reference periods are presented in this paper.

Keywords: expenditure survey, household budget survey, diary, recall interview, calendarization

1. Introduction

The Survey of Household Spending is undergoing a major redesign. The main objectives of this redesign are to review the one-year reference period currently being used and adapt it to the ability of the respondents to report their expenditures, to spread data collection over the year in order to reduce the impact on the agency's operational capacity, and to develop a strategy that would allow the integration of other surveys on related subjects.

A model based on continuous data collection that combines the use of a diary and an interview with recall periods of various lengths has been developed. This approach will be evaluated in a pilot survey from November 2007 to October 2008. This model, which is considered more realistic in terms of respondent ability to provide the information, is expected to reduce response error and burden. However, it increases the complexity of data processing and has major impacts on how the various outputs required by data users are produced. A large number of methodological issues need to be examined in order to develop this design and evaluate its impact on the products and their quality. This paper will focus on the issues related to the sample design and the estimation of annual expenditure in presence of different reference periods.

The current expenditure survey program is described in the next section along with the justification for its redesign. The proposed survey design is summarized in section 3. The sample design and estimation issues are treated in sections 4 and 5 respectively while the problem of reference period of annual estimates is discussed in section 6. Future plans in terms of development and evaluation for the new design in 2010 are finally presented.

2. Background

Household expenditures data are currently collected through two surveys, the annual Survey of Household Spending (SHS) and the periodic Food Expenditure Survey (FES). The SHS collects information from a sample of approximately 21,000 households. The survey has several objectives. It serves as a data source for the System of National Accounts (SNA) estimates, for the update of the basket used in the computation of the Consumer Price Index (CPI), for the development of social policy simulation models, as well as for a wide variety of social indicators and social research applications. The current SHS is conducted between January and March via a personal interview. Respondents are asked to retrospectively report their spending for an exhaustive set of expenditure categories for the previous calendar year. Information on household composition, dwelling characteristics and personal income is also collected.

The survey is viewed as difficult from a respondent's point-of-view as well as from an operational point-of-view. The response burden is high considering that the interview lasts almost two hours on average. The impact of the twelve-month reference period on response error is also a major concern. In addition, the current approach puts enormous pressure on operational capacity in the field since all the interviews have to be conducted from January to March. Each year, new interviewers have to be hired for a short period of time in order to satisfy this increased demand. These operational constraints translate into higher costs and potentially lower data quality.

In addition to the current SHS data, information on detailed food expenditures is also required about every four years for the update of the CPI basket. These data used to be collected through the FES from a sample of

15,000 households that are asked to complete a diary over a period of two weeks. Such a collection methodology involves two to three personal contacts for each household, which results in high collection costs. Due to budget constraints, the sample size of the 2001 FES was reduced to 8,500 households. This had a major negative impact on the precision of provincial estimates. Furthermore, the 2005 FES was cancelled.

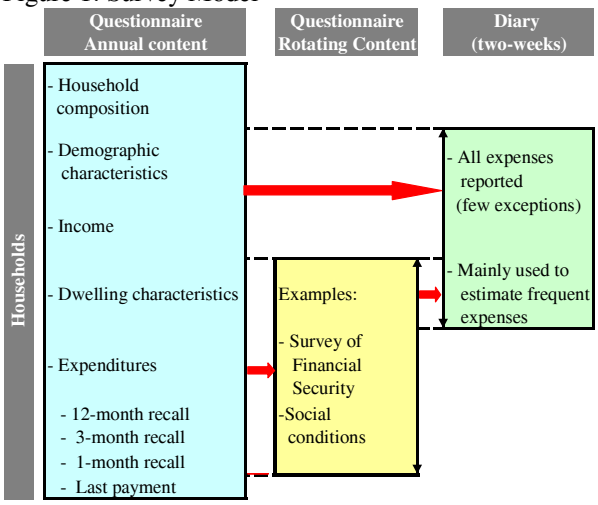
Changes to SHS data collection and questionnaire design are also required to achieve two major objectives of the New Household Survey Strategy Project (Statistics Canada, 2007) recently initiated by the agency. The first objective is to improve the distribution of the overall interviewer workload over time, while the second one consists of the integration of surveys with a high proportion of overlapping content. All these factors led to a fundamental re-examination of the design of expenditure surveys.

3. Proposed Collection Methodology

In order to spread the SHS collection through the year, the annual sample will be divided into twelve sub-samples corresponding to the monthly collection cycles. A questionnaire will be administered to the whole sample to collect data on regular expenditures such as rent and electricity using a last payment approach, and on less frequent expenditures using recall periods of one, three or twelve months. Data on all expenditure categories will also be collected through a two-week diary, which will be left with a portion of the households in each monthly sub-sample. These data will, however, be used mainly for the estimation of the most frequent expenditures such as those for food, clothing and personal care. Such a collection methodology is in accordance with international practices.

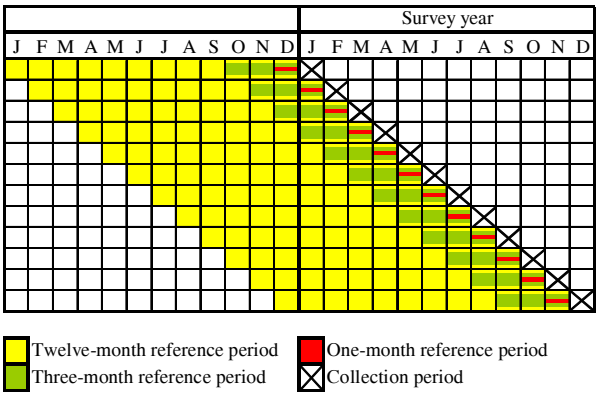
It is also expected that the content of some periodic surveys such as the Survey of Financial Security will be integrated as a rotating content supplement. Depending on the survey requirements and the impact on response burden, the supplement could be conducted on the complete sample or on a portion of the sample only. The overall survey model is summarized in Figure 1.

Figure 1: Survey Model



Under this monthly data collection scheme, the reference period will be staggered according to when the data are collected, as illustrated in Figure 2. The most common practice in household budget surveys is the use of a moving reference periods of fixed length extending backwards from the time of interview. However, the results of focus group consultations indicated that a "moving calendar period" would be seen as a good option for the respondents so it was decided to evaluate this approach for the recall periods in the pilot survey. This means that items collected over a one-month reference period will be asked for the calendar month preceding the month of the interview. Similarly, a twelve-month reference period will cover the twelve calendar months preceding the collection month.

Figure 2: Illustration of the Different Collection and Reference Periods



The use of a "moving calendar period" approach to define the recall periods in the interview allows the reference period of one specific item to be the same for all households assigned to a given collection month. This will not be the case for expenditure data collected

through the last payment approach since the period and even its length may vary from one respondent to the other. For data collected through the diary, the reference period will also vary between respondents since a moving period corresponding to the fourteen days following the interview will be used.

4. Sample Design

The size of the current SHS sample is approximately 21,000 households. It is selected according to a stratified multi-stage design (mainly two stages) from the Labour Force Survey (LFS) area frame. The sample is allocated provincially to obtain a compromise between an optimal precision of national estimates and the need for comparable precision of estimates from each province. Within each province, the allocation of the sample among the LFS strata is usually done proportionally to their population size. Some "high income" strata containing a larger proportion of high income households are oversampled (Mitchell and Nadeau, 2006). The type of design used for the annual sample selection as well as the sample allocation methodology will remain unchanged after the redesign.

4.1 Sample Size

Given that sub-annual amounts will be collected for several items under the new design, the period-to-period variation in spending represents an additional source of variability, and will contribute to an increase in sampling error. The impact of the length of the reference period on the precision of estimates has been assessed using the American Consumer Expenditure Survey (CES) data. The study showed that the use of a one-month reference period instead of a twelve-month reference period leads to a predicted variance increase of 47% for telephone expenses, 33% for electricity, 21% for natural gas and 14% for rent. Such results clearly advocate for an increase in the overall sample size. This increase will however be limited by budgetary constraints.

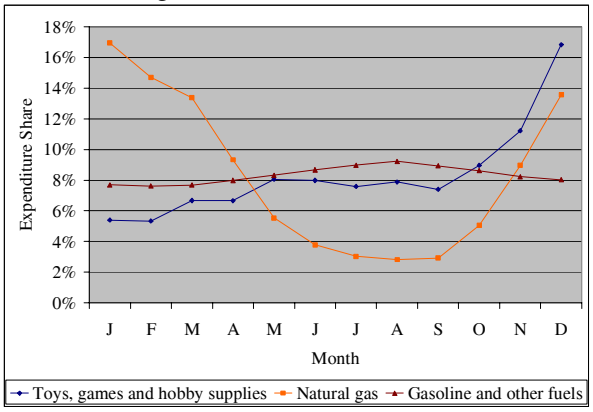
The proposed design involves an additional personal contact for each respondent who will have to complete the diary. This should increase the travel costs. On the other hand, the continuous data collection will provide the opportunity to better integrate the data collection of SHS with that of other household surveys using continuous collection through personal interviews. It will also reduce the costs of hiring and training new interviewers. The overall impact of these factors on collection costs is difficult to predict. The redesign is currently planned with an overall sample of 24,000 households and an expected sub-sample of 50% for the diary component. However, these numbers will be reviewed when more information on the collection costs of this new design become available from the pilot survey.

4.2 Size of the Monthly Sub-samples

From a collection point-of-view, balancing sample sizes from month to month seems suitable to avoid any major variation in the monthly workload. On the other hand, it might be desirable to allocate the annual sample between the months proportionally to the monthly expenditure levels in order to reduce the sampling variability of the annual estimates.

The month-to-month variation patterns differ by expenditure category as illustrated in Figure 3 for three categories of expenditure data provided by SNA. The adoption of an optimal monthly allocation for one expenditure category could therefore have a negative impact on the sampling variability for another one. As an example, the monthly distribution of the sample that minimizes the variance of the annual estimate of "Toys, games and hobby supplies" would lead to only a 5% relative reduction in CVs for that category but to an overall relative increase of 1% in the CVs for other spending categories. Such results, combined with the operational desirability of stable monthly workload, advocate for keeping the same sample size every month.

Figure 3: Monthly Proportion of Annual Expenditures for Three Categories Based on SNA Data



4.3 Creation of the Monthly Sub-samples

There might be some advantages to spreading the sample selected in each stratum over the year given that the period-to-period variation pattern of expenditures may differ from stratum to stratum. But to minimize the impact on collection costs, it would be preferable to assign an entire primary sampling unit (PSU) to a single collection month. A problem,

however, is that the small number of PSUs selected per stratum, only three on average, prevents each stratum from being represented each month, or even each quarter.

For the pilot test that will be conducted in November 2007, the strata were regrouped into superstrata to facilitate the creation of monthly sub-samples and ensure that each superstratum was represented each quarter. Simulation studies showed that the formation of superstrata made of geographically close strata and of "high income" strata helped to control the sampling variability when forming sub-samples from the SHS annual sample.

The monthly samples are formed in a way to ensure that, for each superstratum, at least one sample PSU is assigned to each quarter and that the number of sample PSUs per quarter is as uniform as possible. In addition, the selected PSUs of a given stratum are assigned to the maximum number of quarters and months.

4.4 Sample Coordination

The continuous data collection will provide a good opportunity to integrate the collection activities of household surveys conducted through personal interviews. In the LFS, the first of the six monthly interviews is generally done through personal interview. Furthermore, the Canadian Community Health Survey (CCHS) has recently been redesigned such that its data collection could be done on a continuous basis using two-month collection cycles (Thomas, 2007). About 50% of the sample for that survey is selected from the LFS frame, with data collection through personal interviews. Since LFS, CCHS and SHS make use of a similar design, reduction in travel costs could be expected if the monthly sample collection of these surveys could be coordinated.

On the short term, it would have been difficult to modify the sample design of these surveys to allow better coordination in the selection of PSUs. For the pilot sample, the PSUs were thus selected as usual. However, the selected PSUs were assigned to each collection month in such a way as to maximise the coordination of collection for PSUs in common with CCHS or LFS first interviews, while respecting the representativity constraints of the survey described in 4.3. Although not optimal, this approach has allowed an overlap, for collection, of about 75% of the SHS selected PSUs with CCHS or LFS PSUs.

The pilot will be conducted on a sample of 4,200 households in two of the ten provinces. The size of the monthly samples will then be small in comparison to LFS births and CCHS. However, this is viewed as a first step in evaluating the operational suitability of a coordinated collection approach and in trying to better understand its impact on travel costs.

5. Estimation

The SHS data are currently collected for a fixed period of one year corresponding to the previous calendar year. Therefore, the annual estimates of total expenditures are simply calculated as the weighted sum of the observations. The final weight is the product of the inverse of the probability of selection, a nonresponse adjustment factor, an influential data adjustment factor and a calibration factor.

The SHS national response rate generally varies from 70% to 75%. The response rate varies by province and by region within province. The nonresponse adjustment is performed within predefined nonresponse adjustment groups to compensate for the loss of nonresponding households and to somewhat correct for nonresponse bias. The nonresponse adjustment groups are based on stratification information such as urbanization levels. The lack of information on nonrespondents however restricts the performance of the nonresponse adjustments. The nonresponse adjustment factors correspond to the inverse of the weighted response rate within the nonresponse adjustment groups.

Expenditure estimates can be highly affected by the presence of influential observations. A procedure is incorporated at the weighting step to identify observations that contribute excessively to aggregate estimates and to reduce the impact of the very extreme cases.

Calibration uses variables correlated with expenditures to correct for the potential lack of representativity of the respondents and to reduce the variability of expenditure estimates. The correlations between the expenditure variables and potential auxiliary variables have been evaluated. The impacts of different sets of controls on the calibration factors and on the variance of the estimates have also been assessed. This led to the use, for calibration, of the number of persons in 8 different age groups, the number of households of size 1, 2 and 3+ and the number of earners in six different earning classes, defined at the provincial level (Lessard, 2005).

The proposed collection approach for the redesign will further complicate the estimation process. The estimation of annual expenditures will combine data

collected for different reference periods as illustrated in Figure 2. These reference periods depend on the collection periods and their lengths vary by expenditure item. These differences combined with the year-to-year and month-to-month variation pattern of expenditure data need to be accounted for at the estimation stage in order to minimize the impact on estimate variability and bias.

5.1 Adjusting for Reference Period Length

Annualization factors will be applied to address the item-to-item variation in the length of the reference period and produce annual expenditure aggregates. These factors will simply be calculated as the ratio of the length of a one-year reference period over the length of the reference period used for data collection. They will be equal to 12/3 for three-month recall items, to 12 for one-month recall items and to 365/14 for expenditures collected from the two-week diary. In general, these factors will be the same for all households. One exception will be for the items collected through the last payment approach where the duration of the period to which the payment is applied will be provided by the respondent and may vary from one household to the other.

5.2 Accounting for the Variation of Representativity across Periods

To even-out the effects of seasonal and other temporal variations, the reference periods associated with the different monthly sub-samples should all be similarly represented in the calculation of the annual estimates. To avoid any bias, the weighting should also preserve, as far as possible, the geographical representativity of these sub-samples. Even with a design where each superstratum is uniformly represented over the quarters, monthly and quarterly variations in the achieved sample size and representativity cannot be avoided. Monthly variation of nonresponse and dwelling vacancy will contribute to such variations. The overall calibration as it currently stands may also disturb an originally balanced sample. The magnitude and the impact of these variations will have to be evaluated to determine the type and extent of adjustment required.

The nonresponse adjustment methodology could be modified to account for the month or quarter of collection. This would limit the bias that could be introduced by the reduction of representativity for a given reference period because of nonresponse. One way to achieve that would be, for example, to include the month of collection as an independent variable when modelling the propensity of response. Nonresponse adjustment groups would then be formed from respondents and nonrespondents showing similar predicted propensities of response based on such a model. If the response rates are lower for a given month, the nonresponse adjustment groups that include the respondents for this collection month should show higher nonresponse adjustment factors.

To better control for temporal representativity, constraints could be added to the calibration for the weights of each monthly sub-sample to sum to one-twelfth of the provincial population, or for each quarterly sub-sample weights to sum to one quarter of the population. The choice of controls for calibration would then have to be balanced between the need for good representativity over space (i.e. geographically and socio-economically) and over time. For example, the total number of households, persons and earners could be controlled quarterly in addition to the calibration constraints enumerated at the beginning of section 5. More detailed constraints could be added for the quarters, but annual calibration constraints would need to be relaxed to keep the constraints at a manageable level. The best compromise will have to be identified when the pilot data become available.

Another concern with the representativity of the reference periods is with respect to the expenditure items having a reference period shorter than a month. Food expenditures, collected through a two-week diary, are one such case. The interviewers will be instructed to distribute their interviews evenly between the weeks of the month but the sample size per week will not be controlled at the design stage. A design that allocates sample by week would have too much impact on the response rates and on field operations. The magnitude of week-to-week variations and their impact on the estimates for the items collected over a two-week period will have to be evaluated and temporal adjustments may be needed.

5.3 Accounting for the Reduced Length of Recall Period and the Reduced Sample Size of the Diary

As already mentioned in 4.1, the reduction of the length of the recall period will generally lead to an increase in the variance of the annual estimates. Although this increase is expected to be smaller for very frequent expenditures, the reduced sample size of the diary component will also have a negative impact on the variability of the annual aggregate estimates.

For some users and most particularly for the CPI basket where very detailed information is required, this will need to be resolved by cumulating data over two, three or even four years. Different issues will need to

be considered in combining data over more than one year.

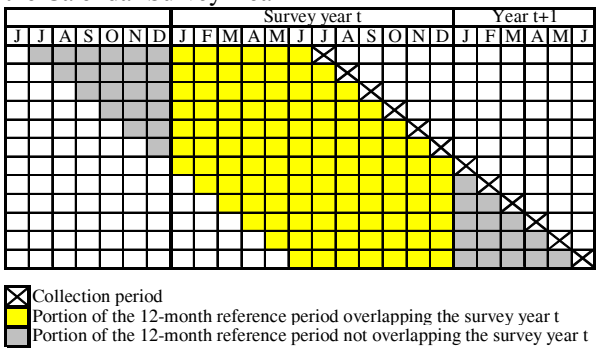
Populations and their characteristics may change over time and these changes may not be ignorable if a large number of years are used. Changes to the questionnaire over the years may also have an impact. Furthermore, under the current design, the same PSUs can be selected in successive years. The potential effect of positive covariance when combining data over several years will have to be evaluated. Sample design minimizing this impact may have to be considered.

6. Reference Period of Aggregate Estimates

Continuous data collection combined with the use of "moving calendar reference periods" will have an impact on the reference period of aggregate estimates. The reference period of the estimates produced will no longer correspond to the survey calendar year as previously illustrated in Figure 2. For items reported over a one-month period in a survey conducted in year t, data will cover expenditures made from December of year t-1 for the January sample, to November of year t for the December sample. For a twelve-month reference period, the period covered by the reported data will include expenditures made from January t-1 to November t. It can also be seen from Figure 2 that annual aggregate estimate for an item collected using a twelve-month reference period would be centered around the middle of December t-1 instead of the end of June t.

The July *t* to June *t+1* samples would better cover the expenditures for calendar year *t* for these items as illustrated in Figure 4 but it would delay the release of survey results.

Figure 4: Overlap between the Twelve-month Reference Periods of July t to June t+1 Samples and the Calendar Survey Year



Although many countries use a similar collection methodology, they do not adjust the data to better cover the survey calendar year. Whether an adjustment to the calendar year, referred as calendarization, is needed or not depends on survey data use. Calendar year data are not needed for the update of the CPI basket. However SHS data are required on an annual basis in the Canadian SNA. There may then be some types of expenditure for which inflation is important and calendarization would be preferable. This suggests that estimators that cover the calendar survey year better should be investigated and could be applied for selected expenditures.

6.1 Investigation of Calendarization Methods

To illustrate the problem of calendarization for expenditure items collected for a twelve-month reference period, two variables are defined below and illustrated in Figure 5:

i) the monthly expenditure values $y_{i,m}$ where $i = 1, 2, ..., N$ represents a unit of the population and $m = 1, ..., 24$ corresponds to the months of January of year t-1 to December of year t;

ii) the twelve-month expenditure values $Y_{i,p} = \sum_{m=p}^{p+11} y_{i,m}$ covering expenditures from month p to $p+11$.

Figure 5: Illustration of the Monthly Expenditure Value and Twelve-month Expenditure Value Concepts and Notation

	t-1									t								
m	1	2	...	7	8	9	10	11	12	13	14	15	16	17	18	19	...	24
	J	F		J	A	S	O	N	D	J	F	M	A	M	J	J		D
1																		
...																		
i				$y_{i,7}$	$y_{i,8}$	$y_{i,9}$	$y_{i,10}$	$y_{i,11}$	$y_{i,12}$	$y_{i,13}$	$y_{i,14}$	$y_{i,15}$	$y_{i,16}$	$y_{i,17}$	$y_{i,18}$			
...																		
N						$Y_{i,7}$												

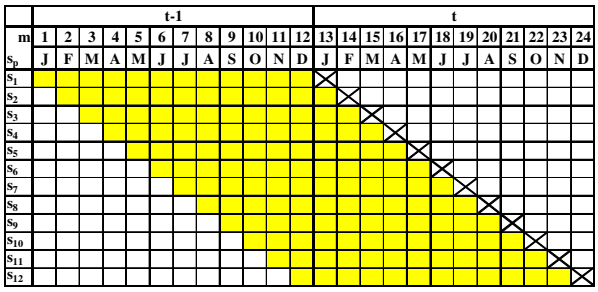
Under the proposed design, the sample s is selected from the population and allocated to twelve monthly sub-samples s_p , $p = 1,..., 12$ for collection purposes. For each $i \in s_p$, $Y_{i,p}$ is collected as illustrated in Figure 6 but the monthly values $y_{i,m}$ are unknown. The problem is that the reference period of the estimates produced by the simple estimator

$$\hat{Y} = \sum_{p=1}^{12} \sum_{i \in s_p} w_i Y_{i,p} = \sum_{p=1}^{12} \sum_{i \in s_p} w_i \sum_{m=p}^{p+11} y_{i,m}$$

where the w_i are the survey weights, will not correspond to the total expenditure for the calendar year t, which is

$$Y_t = \sum_{i=1}^{N} Y_{i,13} = \sum_{i=1}^{N} \sum_{m=13}^{24} y_{i,m} .$$

Figure 6: Period Covered by the Twelve Monthly Samples for Survey Year t Under a Twelve-month Reference Period



Estimators that could better fit Y_t will have to be model-based. Different models are being investigated. A set of estimators derived from a model defined at the unit level $y_{i,m} = \beta_m + \varepsilon_{i,m}$, where $i = 1, 2, ..., N$ and $m = 1, ..., 24$ has been identified (Nadeau and Tremblay, 2006). However the twelve-month reference periods being the same for all sample units in a collection month, it was recommended to define the model at an aggregated level (Statistics Canada, 2006). Such a model could be defined as:

$\bar{Y}_p = X_p \beta$, where $\beta = (\beta_1, ..., \beta_m, ..., \beta_{23})^T$

with $\beta_m = N^{-1} \sum_{i=1}^{N} y_{i,m}$ and $\bar{Y}_p = N^{-1} \sum_{i=1}^{N} Y_{i,p}$

and where X_p is a vector of dimension 23, with 1 in positions p to $p+11$ and 0 elsewhere, which indicates the period referred by each sub-sample p.

Given that samples s_1 to s_{12} do not provide any data for month 24, the calendarized estimator

$$\hat{Y}_t = \sum_{m=12}^{23} \hat{\beta}_m \quad ,$$

which covers months 12 to 23, will be studied.

Since $\bar{Y}_p$ can be estimated from survey data by $\hat{\bar{Y}}_p = \dfrac{\sum_{i \in s_p} w_i Y_{i,p}}{\sum_{i \in s_p} w_i}$, this become a standard estimation problem. However, defining β by least square estimation would be problematic using only twelve months of data.

It is proposed to try to specify a time series model for β. The $\hat{\bar{Y}}_p$ could be graphically analysed from data collected under the new design to determine the type of model for β. As an example, it would be useful to determine if a linear trend model allowing for inflation or other more complex models would be preferable. Even under these types of models, more than twelve months of data should be used in order to get an estimator that is more robust to model specification.

Calendarized estimates tend to be less precise than uncalendarized ones. An evaluation will be needed to measure the efficiency of the calendarization methods as well as their impact on the precision of the estimates.

6.2 Evaluation of Data on Dates of Purchases

The collection of the dates of important purchases has also been identified as a potential approach to produce calendar year estimates or to improve calendarization methods. In practice, this option would have a large impact on the interview length if applied to all items collected through a twelve-month reference period. In order to assess the potential gain of this information, the months of purchase will be collected in the pilot survey - but only for the expenditure questions related to repairs and renovation. These expenditure estimates have been identified as among the most important for SNA and they are more likely to be affected by inflation or other economic changes.

7. Conclusion and Future Plans

In conclusion, a new collection model has been developed for SHS and will be tested in a pilot survey from November 2007 to October 2008. The main features of this model include the use of a diary for the data collection of frequent expenditures, the reduction of the length of recall periods for some types of expenditures, the even distribution of data collection over the year and the integration of the content of the FES.

A sample design, adapted to the needs of continuous data collection, has been developed for the pilot. Some components of this design, such as the monthly sample sizes and the use of superstrata, will be evaluated. The pilot will also be used to evaluate the operational suitability of a coordinated collection approach with other surveys as well as the impact on costs. If the results show benefits in coordinating collection, attempts will be made to increase the overlap in the selection of SHS and CCHS PSUs as well as in the assignment of collection months between SHS, CCHS and LFS first interviews.

The new design will have a major impact on the estimation methods. The effect of seasonal and other temporal variations will have to be assessed from the pilot data to determine the type and extent of adjustments required in the estimation strategy. Considering the major importance of calibration variables in the current SHS, the choice of controls will have to be balanced between the need for good representativity over space and over time, and the best compromise will have to be determined.

Some estimators aiming to produce calendar year estimates will need further investigation. In addition to calendarized estimators, the options of using information on dates of purchase and of producing revised estimates based on July t to June t+1 samples will be evaluated. These evaluations will be done taking into account the accuracy of estimates, the timeliness and the production cost and workload.

The objective is to fully implement the new design in 2010 with a bridging strategy in 2009. For the reference year 2009, data will be collected under the new and the current approaches but with smaller sample sizes. It is planned to conduct the current SHS with a sample of about 15,000 households and to release 2009 SHS data based on this sample. A sample of about 9,000 households would be used to collect the data under the new model. These data would be mainly used to evaluate the impact of the new design and provide information on these impacts to data users when data from the new design will be released for 2010 reference year.

Acknowledgements

The authors would like to thank Guylaine Dubreuil and Jean-Louis Tambay for their valuable comments.

References

Lessard, S. (2005). Revision of the Calibration Strategy of the Survey of Household Spending – Summer 2005, Internal Document.

Mitchell, C. and Nadeau, C. (2006). Using Income as an Auxiliary Variable to Improve the Design of Household Surveys, Joint Statistical Meetings 2006

Nadeau, C. and Tremblay, J. (2006). Redesign of the Survey of Household Spending. Paper presented to the Statistics Canada Advisory Committee on Statistical Methods, Meeting # 43.

Statistics Canada (2006). Minutes of Meeting #43 of the Advisory Committee on Statistical Methods, Internal Document.

Statistics Canada (2007). Statistics Canada's Household Survey Strategy, Internal Document.

Thomas, S. (2007). Review of the Weighting Methodology for the Canadian Community Health Survey. Federal Committee on Statistical Methodology Research Conference.