Pilot Survey Results from the Canadian Survey of Household Spending Redesign

Tremblay, J., Lynch, J. and Dubreuil, G.
Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Ontario, K1A 0T6, Canada

Abstract

A major redesign of the Survey of Household Spending, conducted annually by Statistics Canada, has been undertaken over the last few years. A new data collection model, combining the use of a recall interview and a diary, was developed. The length of the recall period for certain types of expenditures was reduced and data collection was spread through the year. The content of the previous Food Expenditure Survey was also integrated. A pilot survey was conducted to evaluate this new model. This paper presents the results of the analysis on pilot data collection and data quality, including the impact of nonresponse, and the results of evaluation of the new collection methods developed for this model.

Key words: Expenditure survey, Collection methodology, Diary, Nonresponse, Reference period

1. Introduction

The Survey of Household Spending is undergoing a major redesign. A new continuous data collection model, combining the use of a recall interview and a diary, was developed to replace the one-year recall interview conducted in the first quarter of each year. The content of the periodic Food Expenditure Survey was also integrated.

A pilot survey was conducted over a one-year period to evaluate this new model. The objectives were to evaluate respondent reaction, test the operational logistics of continuous data collection, evaluate the new collection methodology and its potential impact on data quality, and develop processing and estimation methods suited to this new collection model.

The new components of the collection methodology were evaluated, with the main focus on continuous data collection, the reduced recall period for certain types of expenditures collected during the interview, and the quality of diary data. The diary evaluations also assessed the impact of estimating frequent expenditures from diary data and the use of receipts to reduce respondent burden.

This paper will present a summary of these evaluations. The objectives of the redesign are described in the next section followed by a description of the methodology of the pilot survey in section 3. The response rates and some results of the analysis on data collection are presented in sections 4 and 5 respectively. The complete and partial nonresponse to the diary is discussed in section 6 and a summary of the analysis of the data collection model is provided in section 7. Conclusions and challenges of the implementation of this new model are presented in section 8.

2. Background

The Expenditure Survey Program serves mainly as a data source for the System of National Accounts, the update of the basket used in the computation of the Consumer Price Index and a wide variety of social research applications, including the development of social policy simulation models. Prior to the redesign, the program was comprised of two surveys. The Survey of Household Spending (SHS) was conducted annually on a sample of approximately 21,000 households between January and March. Respondents were asked during a personal interview to report expenditures made over the previous calendar year for a comprehensive set of expenditure

categories. The Food Expenditure Survey (FES) was conducted on a periodic basis and asked selected households to report their detailed expenditures on food in a diary for a period of two consecutive weeks.

A redesign was undertaken to address various issues with these two surveys. One of the main objectives was to spread data collection over the entire year in order to reduce the considerable pressure on operational capacity in the field from January to March. The redesign was also expected to be more realistic in terms of what is required from the respondent so as to increase their cooperation and reduce response errors. Reference periods better suited to the respondent capacity to provide the information and a reduction of the interview length were therefore targeted. A third objective was to integrate the content of the FES into the redesigned SHS to increase the frequency and improve the reliability of detailed food data without increasing the costs. In past years, the frequency and sample size of the FES had been reduced because of budget considerations. As a result, the survey was no longer able to meet the requirements of the main users.

A new continuous data collection model, combining the use of a recall interview and a diary, was developed. Under this model, the annual sample is divided into twelve sub-samples that match monthly collection cycles. A questionnaire is administered to sampled households 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. Sampled households are also asked to report all of their expenditures, including detailed food expenditures, in a diary for a period of two weeks. Households are requested to include all their expenditures in the diary; however, diary data are mainly used to estimate the most frequent expenditures such as food and personal care, which are difficult to recall even for a short reference period.

This collection methodology, which reflects international practices, meets the main objectives of the redesign. However, the international model is quite expensive (in terms of unit cost), since three personal visits are generally made to each sampled household. The overall burden is also large, each household having to complete both the interview and the diary. Some measures were therefore taken to reduce the response burden. The content of the interview was adjusted under the constraint that the interview length should be no more than 60 minutes and the respondent's consent to use tax data on income was sought in order to reduce the interview content. Additionally, respondents were allowed to provide receipts instead of transcribing all expenditure information in the diary. To reduce collection costs, the personal visit generally made in the middle of the diary two-week reporting period was replaced by a telephone call and the verification procedures made by the interviewer during diary pick-up were reduced considerably.

The new collection methodology was evaluated in a pilot survey conducted from November 2007 to October 2008 with the objective of fully implementing the new design in 2010. A parallel run of SHS and the redesign version was also part of a bridging strategy to measure the impact of the new approach on survey products. The SHS-redesign data were collected from January to December 2009 in the ten provinces. Once processing completed, the data will be compared with the SHS 2009 data.

3. Methodology of the Pilot Survey

The pilot survey was conducted over twelve monthly collection periods in two provinces, Quebec and Ontario, on a sample of 4,200 households. The pilot sample was selected according to a stratified (mainly) two-stage design from the Labour Force Survey area frame. The annual sample allocation and dwelling selection were done similarly to the regular SHS. This sample was spread

over the monthly collection periods in such a way that an entire primary sampling unit would be assigned to a single collection month and the sizes of the monthly samples were similar (Nadeau et al., 2007).

The questionnaire was administrated via a computer assisted personal interview. The twelve-month, three-month and one-month recall periods were based on calendar months, i.e. a twelve-month reference period covered the twelve calendar months preceding the collection month. At the end of the interview, respondents were asked to record in a diary daily expenditures made by all members of the household for a period of two weeks, starting the day following the interview. They were given the option of providing the receipts of their expenses to reduce the amount of transcription in the diary, and were asked to provide additional information directly on the receipt to clarify the receipt descriptions when needed.

The decision on which expenditure categories should be estimated from the two-week diary as well as the length of the recall period for each expenditure question in the interview had to be made *a priori*. The results of qualitative tests and expertise from international expenditure surveys with similar approaches were used. The potential increases in variance were also approximated for various categories of expenditures based on data from the Consumer Expenditure Survey of the U.S. Bureau of Labor Statistics. The final distribution of expenditure categories (largely equivalent to SHS questions) by collection reference period[1] is provided in Table 1. The proportion of total consumption expenditures to be collected according to each reference period is also included.

Table 1: Distribution of expenditure categories (SHS questions) and consumption expenditures by collection reference period in the pilot survey

Percentage (%)	Interview Recall Periods					Two-week diary
	Twelve months	One year (calendar)	Three months	Last payment	One month	
Expenditure category	35.1%	3.5%	3.9%	8.5%	1.8%	47.2%
Consumption expenditures	25.8%	0.5%	2.3%	28.3%	0.8%	42.3%

To address the issue of low diary response rate and the feedback from the field indicating that the diary appeared overly complicated, the diary format was modified during data collection. The two one-week diaries were amalgamated into one two-week diary. The content to be reported was reduced, with information such as quantity and type of food (fresh, frozen, etc.) removed. The number of sections was reduced from 4 to 2, one for food from stores and other goods and services, and one for food and beverages purchased from restaurants. The format by day was replaced by a list format. The grid provided on the front page of the diary to indicate the days with and without expenses was removed; respondents were required to indicate the days without expenses within the diary. Finally, the layout of the diary was improved significantly and the instructions simplified. The new diary format was used during the last two months of the pilot.

Diaries received from the field were scanned and data were captured and auto-coded to the level of detail required by SHS users through a process where the description of an item was matched to a regularly updated look-up dictionary. A rule had to be defined to identify usable diaries since

---

[1] In the paper, recall period refers to the reference period used in the interview; reporting period refers to the reference period used for the diary.

the diaries received from data collection showed that not all households filled in their diaries over the 14 days required. Usable diaries were defined based on the number of responded days, with a minimum criterion of five responded days over the two week period.

Donor imputation was performed on usable diaries to fill in missing data. An example of missing data is a reported item where the cost is unknown. Donor imputation was also used to increase the level of detail in the coding. When the coders were unable to assign a detailed code, they were often able to at least code the item to a more general code such as bakery products. Imputation was used to code these general codes to more specific codes, such as bread. For the interview, the edit and imputation methods used in SHS were reviewed to take into consideration the continuous data collection, the various recall periods and the use of tax data for income.

The high nonresponse rate of the diary in the pilot, combined with the possibility that diary data would be collected only for a subsample of the interview respondents once the new design fully implemented, led to the production of two different sets of weights, one for interview data and one for diary data. Interview nonresponse adjustments were applied to the sampling weights. Analysis was done on frame data, paradata and dwelling type in order to find a model explaining the interview nonresponse. Nonresponse classes were then created using the score method (Alavi and Beaumont, 2004) to produce the interview nonresponse weight adjustments. A second phase of nonresponse adjustment was applied to the diary weights. A similar approach was used with the advantage that all the information provided by the interview respondents was available to build the nonresponse model. Lastly, both interview and diary weights were calibrated to some household and population demographic estimates used in SHS (Lessard, 2005).

Annualization factors were applied to address the variation among expenditures of the length of the reference period and to produce annual expenditure aggregates. These factors were calculated as the ratio of the length of a one-year reference period over the length of the reference period used for data collection. Average estimates of annual expenditures as well as estimates of variance were produced for a certain number of expenditure categories to evaluate the new collection methodology. Variance estimates were produced using the bootstrap method.

4. Collection Response Rates

The collection response rate of the pilot survey interview was 60.0%. This rate is slightly lower than the collection response rate of SHS 2007 conducted in the first quarter of 2008. The combined SHS 2007 collection response rate for Quebec and Ontario was 62.9%. The response rate of the pilot for the three months overlapping with SHS collection was 55.8%.

The low response rate of the pilot survey follows a decreasing trend observed with SHS over the last few years. The response rates of SHS 2007 and SHS 2008, conducted in the first quarter of 2008 and 2009 respectively, are lower than those of SHS 2003 to 2006 for both Quebec and Ontario.

The refusal rate and the non-contact rate of the pilot survey were very similar to the rates of SHS 2007 as indicated in Table 2. Larger differences were observed with residual nonresponse where language barriers and unusual circumstances were the major contributors to these differences.

Table 2: Comparison of response and nonresponse rates of the pilot interview to SHS 2007

Rates (%)	Quebec		Ontario		Quebec and Ontario	
	Pilot	SHS 2007	Pilot	SHS 2007	Pilot	SHS 2007
Response to the interview	67.0	66.4	54.9	59.9	60.0	62.9
Refusal	21.9	22.3	25.2	24.4	23.8	23.4
Non contact	7.3	9.2	13.6	12.1	10.9	10.7
Residual nonresponse	3.8	2.2	6.4	3.7	5.3	3.0

About 70% of the interview respondent households agreed to fill in the diary at the end of the interview, but the rate of returned diaries was only 64% of the interview respondent households. The combined effect of interview nonresponse and diary nonresponse led to a final diary collection response rate of 38%.

The low response rate to the diary observed in the first months of the pilot was the main motivation for major modifications to the diary content and format during the pilot. The new diary format used in the last two months had a positive impact on the response rate as shown in Table 3. The percentage of respondent households who agreed to complete the diary at the end of the interview was 6 percentage points higher than with the original diary. However, the percentage of diaries returned increased by only 3 percentage points. The new diary was easier to sell to the respondents; however, the percentage of households that agreed to complete it but did not return the diary was larger.

The new diary was also used for the 2009 collection of the redesigned SHS (SHSR 2009) and higher diary response rates were observed. For the ten provinces, 83% of the respondents to the interview agreed to fill in the diary, and 71% of the interview respondents returned their diary, for a final diary collection response rate of 44%.

Table 3: Comparison of diary response rate for the original and modified diary formats

Diary response rate (%)	Pilot Survey (Quebec and Ontario)		Redesigned SHS 2009 (10 provinces)	
	Original diary	Modified diary	Quebec and Ontario	All provinces
Diary accepted by interview respondents	69%	75%	80%	83%
Diary returned by interview respondents	63%	66%	69%	71%
Final diary collection response rate	38%	41%	41%	44%

5. Monitoring of Data Collection

The pilot survey was assigned to the pool of experienced interviewers working on continuous computer-assisted personal interview surveys such as the Labour Force Survey and the Canadian Community Health Survey. Collection effort was adapted to the difficulties encountered in the field throughout the various monthly collection periods. On average, 5.1 attempts per dwelling were made during data collection and an analysis of the paradata (Lynch et al, 2010) showed that the average number of contacts made by interviewers was higher for the months when respondents were more difficult to contact such as December and March.

The short length of the monthly collection period was a major concern for the field operations. With the purpose of giving more time to the interviewers to reach the non-contact cases, the data collection period was extended from four to six weeks beginning in June 2008. The analysis indicated a decrease of the average non-contact rate for the months with a longer data collection period; however, the collection month and other factors may also have had an impact.

At the beginning of the pilot, interviewers were asked to try to spread the data collection of the monthly sample over the month. The objective was to try to have a distribution of the diary reporting periods as uniform as possible over the month. This idea was abandoned early in the pilot. Interviewer feedback combined with the results of the analysis of paradata showed the importance of a first contact with the respondent early in the month in order to obtain a better monthly response rate. In the pilot, all sampled households were generally contacted at least once within the first 15 days of the month. It was recommended to target a first contact with the household within the first 10 days of the month.

Throughout the pilot, data collection monitoring, interviewer debriefing and comments, in-depth interviews and observation of interviews were all useful tools that helped the whole team better understand what was working well or not, especially the problems related to the diary. Better instructions were then given to the interviewers and the respondents, for example, to improve their understanding of the importance of the diary and to decrease the respondent burden.

6. Diary Nonresponse

Three types of nonresponse to the pilot diary were defined: total nonresponse, incomplete diaries and incomplete diary items. Various analyses were done to try to understand the impact of these types of nonresponse on diary estimates. In the context of these evaluations, diary total nonresponse corresponds to a household for which an interview was not completed, a diary was not returned or the returned diary was determined to be non usable. Incomplete diaries are usable diaries with a minimum of five responded days and containing a certain number of nonresponded days. Finally, incomplete diary items correspond to items with missing information or with insufficient information to be coded at the level of detail required for SHS users.

6.1 Total Nonresponse

About 36% of the respondents to the interview did not complete the diary and 11% of the received diaries were determined to be non usable. The final response rate for the diary component in the pilot was therefore 34%. This low response rate raises concerns about the representativity of the usable diary data: does the sample of households who returned usable diaries remain representative of the population? Although it is not possible to answer this question, information from the interview was available to evaluate whether the diary respondents were over- or under-represented in terms of socio-demographic characteristics compared to the interview respondents.

Estimates of the variables collected during the interview were calculated from the interview respondents and from the diary respondents. For this comparison, the diary nonresponse adjustment was simply an inflating factor at the stratum level, and no calibration was used for both the interview and the diary weights. The differences for the main socio-demographic characteristics collected in the interview are presented for Quebec and Ontario in Figure 1 (referred to as "without adjustment"). Although in most cases the differences were not significant, large differences were found for some characteristics such as household type where single parent

families and, to a lesser extent, couples with and without children are over-represented in comparison to the other types of households among the diary respondents. The elderly also seemed more inclined to fill in the diary. On the other hand, persons living alone, renters and households with incomes in the first quartile (Q1) are under-represented compared to interview respondents.

Figure 1: Differences (%) between socio-demographic estimates based on interview respondents and diary respondents, with and without adjustments for diary nonresponse and calibration
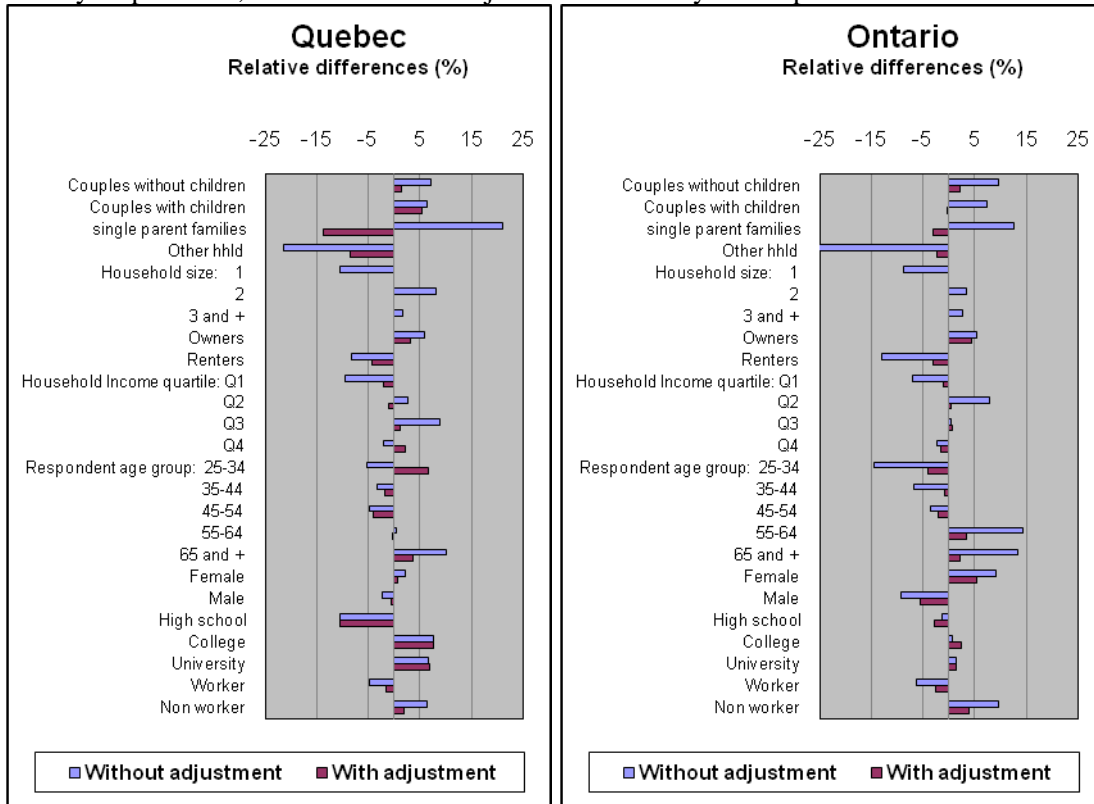
Evaluations were done to develop diary nonresponse weight adjustments based on the information collected from the interview, and to develop a calibration strategy based on population counts by age and household counts by size from household and population demographic estimates. The analysis indicated that the best approach to reduce the problem of representativity of the diary respondents is to use information on household type and household income for the nonresponse adjustment, and to use the same control totals in the calibration of interview and diary weights. The differences between socio-demographic estimates produced from the interview respondents and the diary respondents when this strategy is used are provided in Figure 1 (referred to as "with adjustment").

Despite the improvement of the representativity of many socio-economic characteristics shown in Figure 1, differences that might have an impact on the expenditure estimates still remain. Table 4 shows some provincial summary results. Most of the differences are low in Quebec. In Ontario, the estimates from the diary respondents are slightly lower than estimates from the interview respondents. These expenditure estimates were also produced for several domains based on the socio-demographic characteristics described earlier. On average, the absolute relative differences are at 6.0% for Quebec and 4.7% for Ontario.

Table 4: Differences (%) between interview expenditure estimates based on interview respondents and diary respondents with adjustments for diary nonresponse and calibration

Expenditure category	Relative differences (%)	
	Quebec	Ontario
Housing	-0.9	-4.5
Household operations	-2.9	-5.5
Household furnishing and equipment	2.2	-7.2
Transportation	0.2	-5.6
Health services	-1.8	1.2
Recreation	5.3	5.4
Education	-3.2	-0.7
Other expenditures	12.9	6.5
Total expenditures	0.2	-3.0

6.2 Incomplete Diaries

The diaries received from the field were not necessarily complete. Some households did not fill in the diary over the 14-day reporting period required. During the pilot, information was collected in order to distinguish nonresponded diary days from days with no expenses. Based on this information, almost half of the usable diaries contain nonresponded days. Overall the average number of nonresponded days for the usable diaries is 1.7 or 11.8% of the required days. These nonresponded days could therefore contribute to an important underestimation of the diary expenditures.

Some methods of adjustment to compensate for this type of nonresponse were investigated with data from the first nine months of the pilot. It is known that household expenses are not uniformly distributed over the seven days of a week. The pilot results also indicated that the propensity of response for a given day was correlated with the probability of having expenses on this day. An adjustment method taking into consideration the day of the week was then studied and compared to a uniform adjustment based only on the number of responded days.

When all data from the pilot were received, it was found that the average number of nonresponded days was 2.9 for data collected from the modified diary in comparison to 1.3 days for the original diary format. In the modified diary, the respondent had to indicate «no purchases» with the date for each day without expenses. During the SHSR 2009 data collection and with the modified diary still in use, a verification question was added at the diary pick-up visit. The results showed that 37% of households did not indicate their days without expenses in the diary.

The grid approach used to collect information on days with and without expenses in the first ten months of the pilot had shown some quality problems. The results from the approach used with the modified diary were however more worrying, indicating that days without expenses could be seriously under-reported. No adjustment was thus made for nonresponded days in the pilot.

6.3 Incomplete Diary Item

On average, 86 items were reported in a usable diary, of which 64% were food items from stores, 29% other goods and services, and 7% food and beverages purchased from restaurants. The expenses from restaurants had to be reported in a separate section with additional information such as type of restaurant and number of meals.

In general, the information provided by the respondent was complete enough to assign a code at the level of detail required. Of the 111,500 reported food items from stores and other goods and services, 8.1% were not detailed enough to be assigned a detailed code and donor imputation was applied. About 79% of these items were food items for which a very high level of detail is necessary. For 53.4% of the imputed items, at least the food category, such as meat, dairy product or fruit, was known. In some cases, the type of fruit or type of meat had to be imputed, for others the imputation was at a higher level of detail. As an example, the respondent indicated that he bought milk but the type of milk had to be imputed.

For the expenses from restaurants, the imputation rates for type of restaurant and number of meals were 19.3% and 18.7% respectively. These high imputation rates were partially attributable to the fact that some respondents reported these expenses in the wrong section, mainly under the modified diary format.

Finally, some respondents provided a total amount of expenditures, such as the total for a grocery store, without information on what was bought. A list of specific items with their costs had to be imputed. These imputed items represent 3.5% of the total of all expenses reported in the diaries.

7. Evaluation of the new collection methodology

The interview and diary components of the new collection methodology were evaluated. The impact of the modifications made to the interview to reduce its length and to reduce the recall period of some categories of expenditures were measured. The impact of the use of receipts was assessed. Diary data quality was evaluated through a comparison with SHS estimates and potential sources of underestimation from diary data were investigated. The respondent fatigue and the impact of the telephone follow-up were analyzed.

7.1 The interview

The median length of the pilot interview was 67 minutes. This represents a reduction of 36% in comparison to the 105 minutes of the SHS 2006 interview for Quebec and Ontario. A reduction of 55% was observed for the expenditure sections of the questionnaire due to the large number of expenditure categories only collected from the diary under the new model. However an additional 9 minutes was needed for the introduction of the diary at the end of the interview.

A large proportion (85%) of the household members aged 15 or more provided their consent for the use of tax data for personal income. This contributed to a reduction of 30% in the median length of the income section of the questionnaire, but a reduction of only 1.5 minutes in terms of interview time. For the persons who consented to the use of their tax data, the match rate to the income tax file was 91%. The income of the unmatched persons was imputed.

The last payment questions of the pilot interview provided annual estimates similar to those produced from the one-year recall questions of SHS for the 2008 reference year. The overall difference for the sum of all last payment questions was only 3% and no statistically significant differences were found for many components including mortgage, electricity, natural gas and home insurance. The reduction of the reference period also had no major impact on the variances and CVs of the annual estimates for the expenditures collected using this approach. These results on variances confirm what was expected since the last payment approach was used for expenditures made with regular payments.

The three-month recall questions, although not frequently used in the pilot interview, were more problematic. Large differences were observed with the corresponding one-year recall of SHS 2008. Although it is difficult to determine which of the two recall periods provides the better estimates, it seems that some respondents provided annual values in the pilot. These results indicate that the sequence of questions in the questionnaire must be reviewed to help respondents make the transition between twelve-month and three-month recall questions.

7.2 The diary

Use of receipts

The use of receipts considerably reduces the response burden associated with completing the diary. In the pilot, 82% of the respondent households with a usable diary provided receipts. Overall, more than half of diary items reported were obtained from receipts: the proportion is as high as 67% for food items bought in store, 16% for expenses from restaurants, and 47% for other goods and services. However, the percentage of items requiring imputation of a code at a higher level of detail was greater for the items reported from a receipt (5.8%) than for items reported by transcription (2.3%). The experience with the processing of the pilot data indicated that processing methods could be improved to reduce this gap.

Additional analyses indicated that 24% of households provided only receipts: they did not record any purchases in the diary. This finding, combined with feedback from the data collection and processing teams, indicates that diary procedures on the use of receipts were not well understood by all households or by all interviewers. Not only did some households provide only receipts, but other households provided their receipts and also transcribed the information in the diary. These double counts had to be cleaned up during data processing. Diary instructions on how to report expenditures when a receipt is available and when it is not available need to be improved.

Comparison of estimates

Various analyses were done to try to understand the overall quality of diary data. The estimates of average annual expenditures produced from the pilot diary data were compared to the SHS 2008 estimates for the categories of expenditures to be estimated from the diary data under the new collection methodology. It was found that a large proportion of diary estimates was significantly different from SHS 2008 estimates, and was generally lower than SHS. The pilot aggregated estimate for all diary categories of expenditures was 15% lower than the sum of the corresponding expenditure categories in SHS. Some differences were expected due to the changes in collection methodology and the difference in the survey reference for the two sets of estimates. However, the magnitude of these differences and the large number of lower diary estimates led to some concerns about the impact of diary nonresponse and diary response errors.

Some findings about the impact of total nonresponse were presented earlier in section 6.1. Although the response rate to the diary was very low in the pilot, the results of nonresponse analyses using various scenarios of nonresponse adjustment and calibration have not revealed any significant systematic underestimation of expenditures due to a representativity problem with the diary respondents[2]. The impact of incomplete diaries described in section 6.2 is however a major source of underestimation of expenditures from the diary data since all nonresponded days are considered as days without expenses in the aggregate estimates. The approach of deriving

---

[2] Diary weights used in these estimates were not produced with the strategy described in section 6.1, but differences in provincial expenditure estimates have not indicated systematic underestimation of expenditures.

information on the number of responded days by requiring information on days without expenses has not proved to be as good as expected. Quality issues have been raised with the approaches used in the original and the modified diary of the pilot. Considering the potential magnitude of this source of underestimation, efforts should be pursued to improve the quality of this information under the modified diary approach or to develop another approach to obtain this information.

Other potential sources of underestimation from the pilot diary data were investigated. Some analysis of the households providing only receipts showed that their average expenditures ($636) were significantly lower than the expenditures of households providing both receipts and purchases transcribed in the diary ($1040). This contributes to an underestimation of expenditures, particularly those for which receipts are rarely available or for which respondents prefer to keep the receipt for possible exchange or warranty purposes. In addition to this problem, it was found that the diary estimates were much lower than the SHS 2008 estimates for some expenditures, such as cigarettes (68%), lotteries (33%), hair services (51%) and laundromats (99%). Certain types of expenditures seem to be more easily forgotten by respondents.

The problem of underreporting of some specific categories of expenditures has to be addressed. Some measures are the improvement of the instructions on the use of receipts and the verifications during the diary telephone follow-up call or the diary pick-up visit on whether the respondent has included various types of expenditures frequently forgotten. The pilot is a good source of information to identify which categories of expenditures should be prioritized in these verifications.

### *Respondent fatigue and impact of the telephone follow-up*

The expenses reported in the diaries were analyzed by diary day, i.e. for each of the 14 reporting days. The results showed significant differences between the reporting days and a significant linear decrease over the reporting period. Respondent fatigue in a diary survey is a problem previously documented and observed in most diary expenditure surveys (Silberstein and Scott, 1991). The pilot results indicated that the decrease over time was less for food expenses than for non-food expenses, as illustrated in Figure 2.

Figure 2: Reported expenses by diary reporting day
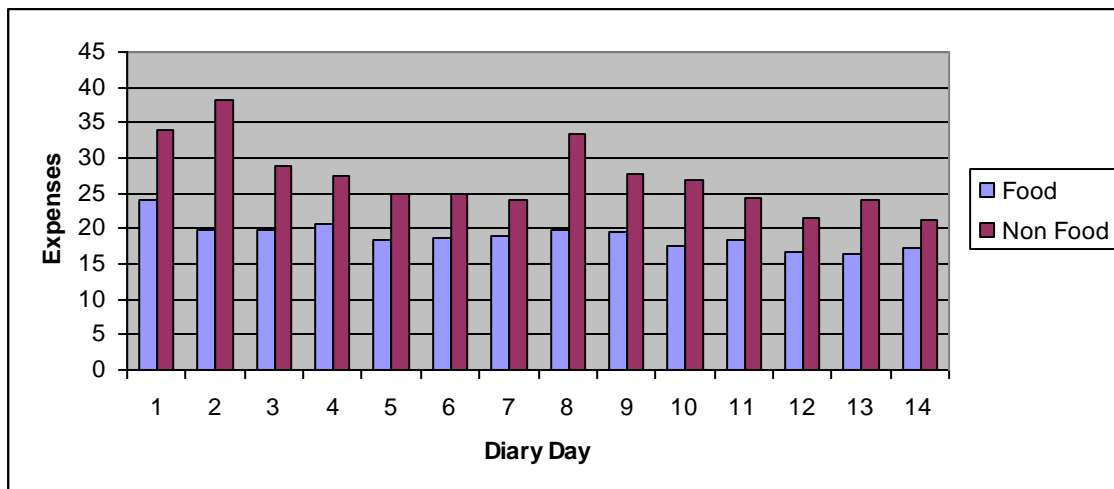
The impact of the telephone follow-up on respondent fatigue in the pilot was also evaluated. The interviewer had to call the household seven days after the beginning of the reporting period, but, in practice, the phone call was not necessarily done that day. The diary reporting days were then rescaled and centered on the day of the follow-up call for each household. The results of the analysis indicated a significant increase in total reported expenditures between the day before the follow-up and the day of the follow-up. The results also indicated than the follow-up had a larger impact on non-food expenses than on food expenses as illustrated in Figure 3.

Figure 3: Reported expenses by rescaled day (centered on the follow-up day)
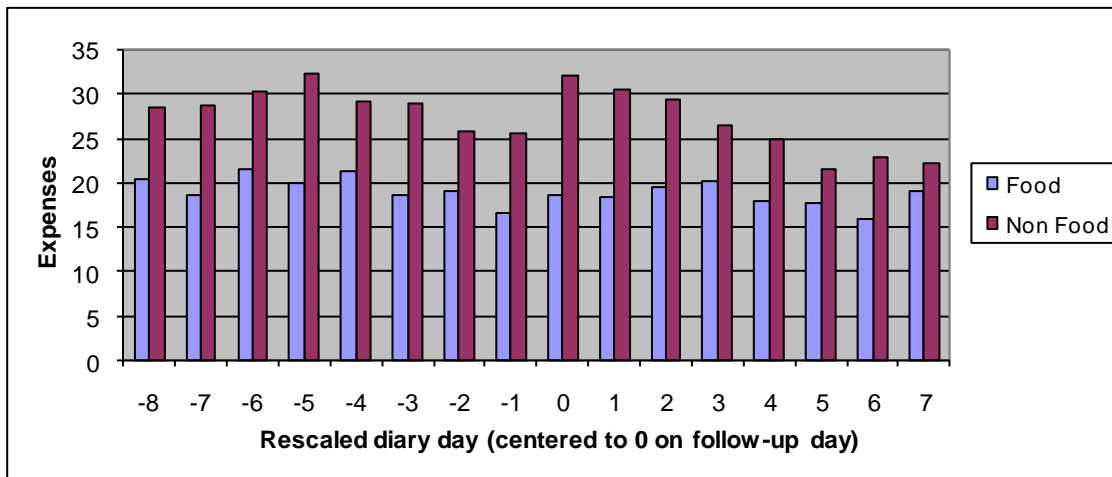
Impact on precision of annual expenditure estimates

The use of a diary to collect a large proportion of the expenditure categories has an impact on the precision of annual expenditure estimates because of the reduction of the reference period from one year to two weeks. The additional source of variability with the use of sub-annual periods is the period-to-period variation in spending. It was therefore expected that the increase in sampling error would be less important for expenses made frequently by the household over the various sub-annual periods of the year. The results of the pilot generally confirm this assumption.

The coefficients of variation (CV) of the average of expenditure categories estimated from diary data under the new design were produced from the pilot. They were compared to the SHS 2008 CV estimates of Quebec and Ontario combined, after being adjusted for differences in sample size. The ratios of the CVs and standard errors (SE) of the adjusted pilot estimates to SHS 2008 estimates are presented in Table 5 for a few categories of expenditures.

The results indicate that the impact of the reduction of the reference period is not very large for frequent expenses such as food, restaurants and fuel for automobiles. Even for expenses made by a smaller proportion of the population where the CV is generally larger, such as pet food, only a small increase is observed.

Table 5: Comparison of SHS 2008 CV (Quebec and Ontario) and Pilot CV (adjusted for sample size) for some categories of diary expenditures

Expenditure category	Proportion of diary expenditures	SHS 2008 CV	Pilot adjusted CV	Ratio Pilot/SHS 2008	
				CV	SE
Food	28%	1.1	1.7	1.6	1.3
Restaurants	11%	2.5	4.7	1.9	2.0
Fuel for automobiles	11%	2.1	3.1	1.4	1.3
Alcohol from stores	4%	3.4	5.1	1.5	1.6
Pet food	1%	4.6	7.6	1.6	1.3
Clothing	8%	2.2	6.0	2.7	2.3
Books	1%	4.4	10.1	2.3	2.6
Museums	0.2%	5.4	19.6	3.6	3.2
Auto repair and maintenance	4%	3.2	15.2	4.8	5.8

The impact of the reduction of the reference period on expenses made less frequently by the household over the year, such as repairs and maintenance of automobiles or admissions to museums, is greater. The reliability of these estimates under the new model could not be good enough to be released at the provincial level. These results were expected. When the content of the questionnaire was developed, the decision to collect data only in the diary on some categories of expenditures made less frequently was mainly to control the length of the interview. The proposed strategy was to combine more than one year of data to produce reliable provincial estimates for these expenditures. With an interview length of 67 minutes, it is felt that this approach should be generally maintained to avoid potential negative impact of an increase of the interview length on the response rate of the diary. However, there might be a few categories of expenditures for which a change will be required to meet some specific user needs.

Another factor that has an impact on the precision of annual expenditure estimates is influential data. A respondent may report an abnormally large but real expense for a category of expenditures estimated from the diary. When the annualization factor is applied, this household may have a large contribution to the estimates and to the variance estimate. Methods to detect these influential data were applied to the pilot data and adjustment methods are in development. From the analysis of pilot data, it was also found that some influential data detected were a miscoded expense falling in a category where expenses are generally much lower. The influential data detection was also a useful tool to detect some coding errors and to help to improve the coding.

8. Conclusion and remaining challenges

The pilot survey conducted to evaluate a new collection methodology has provided relevant information on respondent reaction and data quality. In addition to the final results presented in this paper, preliminary results were produced to provide some input on time for the development of the interview, diary and collection procedures of SHSR 2009 and SHSR 2010. Some of the improvements suggested in section 7 have already been (partially) incorporated. The main pilot results and some of these changes are summarized below.

The response rate of the pilot survey was lower than expected, primarily for the diary component. Slight improvements were observed for 2009 but preliminary results from SHSR 2010 indicate a more significant increase. For the first six months of data collection, the interview response rate was 67% and the rate of returned diaries from the interview respondents was 76%, an increase of

7 and 12 percentage points respectively, in comparison to the pilot. The collection response rate of the diary for the first six months is 50%.

The continuous monthly data collection with the new model allows the interviews to be assigned to a pool of experienced interviewers. The major modifications to the interview are quite positive. The length of the interview was reduced considerably. The reduced recall periods are mainly last payment questions, and have generally shown good results. A few modifications were made to the sequence of three-month questions in SHSR 2010 to help respondents make the transition between twelve-month and three-month recall questions.

The quality of diary data has to improve. Although the diary is considered a more realistic mode to collect frequent expenses, the various sources of underestimation identified during the pilot need to be addressed. The use of receipts introduced to reduce the response burden produced good results, but some households did not record any purchases in the diary. This contributes to an underestimation of expenditures for which receipts are rarely available. Furthermore, certain types of expenditures seem to have been frequently forgotten by the households. Diary instructions were improved and verifications on potential forgotten expenses were added in SHSR 2009 and 2010. The impact of these modifications will have to be assessed.

Respondent fatigue is another source of underestimation. Diary follow-up has been shown to have an impact on reducing respondent fatigue and should be maintained. However the best time for this follow-up is still unknown. A follow-up early during the reporting period could help to reduce the percentage of households that agree to complete the diary at the end of the interview but do not return the diary. On the other hand, an early follow-up could reduce the positive impact on the linear decrease of reported expenses over the 14-day period. For SHSR 2009, the follow-up was done earlier with the objective to improve the diary response rate. The impact on data quality will have to be evaluated.

Finally, one major source of underestimation identified from the pilot results is the incomplete diaries, i.e., diaries with nonresponded days. Even if follow-up procedures are improved, there will always be households not filling in the diary for the complete period of 14 days. The investigation based on responded days from the pilot data has shown the importance of taking this factor into consideration in the adjustment procedures. However, the quality of the information collected in the field with respect to days without expenses has to improve.

In addition to the evaluation and improvement of the new collection methodology, future work will include an evaluation of the impact of the new methodology on the various types and levels of estimates required by the users, based on the SHSR 2009 and SHS 2009.

Acknowledgements

References

Alavi, A. and Beaumont, J.-F. (2004). Nonresponse Adjustment Plans for the Labour Force Survey, Paper presented to the Statistics Canada Advisory Committee on Statistical Methods, Meeting no. 38.

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

Lynch, J., Houle, A.-M. and Chabot-Hallé, D. (2010). Rapport d'évaluation de la collecte de l'enquête pilote 2007-2008 de l'EDM-R, internal document.

Nadeau, C., Tremblay, J. and Lessard, S. (2007). Sampling and Estimation Issues in the Redesign of the Canadian Survey of Household Spending, Joint Statistical Meetings 2007.

Silberstein A. and Scott S. (1991). Expenditure Diary Surveys and their Associated Errors, Measurement Errors in Surveys, Wiley & Sons, pp303-326.