

DATA QUALITY OF INCOME DATA USING COMPUTER ASSISTED INTERVIEWING: THE EXPERIENCE OF THE CANADIAN SURVEY OF LABOUR AND INCOME DYNAMICS

Chantal Grondin, Sylvie Michaud

Chantal Grondin, Statistics Canada, Jean Talon Bldg. 11-D3, Ottawa, Ont., Canada, K1A 0T6

KEYWORDS: Longitudinal survey, Statistical record linkage, Income survey, Tax data.

Computer-assisted interviewing (CAI) is the wave of the future. This relatively new technology is being used at Statistics Canada for the Survey of Labour and Income Dynamics (SLID).

SLID is a longitudinal survey that started in January 1994. SLID consists of two interviews a year: one in January to collect labour information and one in May for income. A major goal of this survey is to measure the impact that changes in family composition have on labour market behaviour and income.

The questions for the labour component of the survey are fairly linear. Hence programming the CAI application was not a problem. For income, however, it was not clear how the information should be collected nor how the application should be programmed. Furthermore, the possibility of using interactive edits with CAI was of great interest.

This paper will start by giving a description of SLID. Next, it will describe how the income application was programmed. It will then give some of the differences CAI brought to the income component of the survey. The evaluation procedures that were used will be reviewed briefly. This will be followed by the results of micro-comparisons between SLID test data (collected using CAI) and tax data for the same year, as well as parallel micro-comparisons between a paper and pencil survey (P&P) and tax data. We will end with some recommendations.

1. DESCRIPTION OF SLID

SLID is designed in such a way that people in the longitudinal sample are interviewed for a period of six years. If they move at any time during this period, procedures have been implemented to trace them and interview them in their new home (as long as it is within Canada or the United States).

As mentioned earlier, respondents are interviewed twice each year. The January interview collects detailed information about labour market activities in the previous calendar year (reference year). The May interview collects information on various sources of income for the reference year. The May interview can

be seen as a deferred interview from January since it concerns the same reference period. The income interview is carried out in May because, in Canada, income tax returns are required every year around the end of April. It was felt that better data quality would be obtained if the interview was done in May. In fact, SLID mails out a paper copy of the questionnaire prior to the interview and encourages respondents to refer to their tax documents and fill out the questionnaire in advance in order to reduce interview time.

The SLID sample is selected one year in advance of the first wave (at the beginning of the first reference year) from another survey, the Labour Force Survey (LFS). A preliminary interview is done for SLID as a supplement to LFS. The information gathered there is fed back to the respondent one year later to reduce recall errors during the labour interview.

In 1993, a test was done to simulate the entire collection process for SLID using CAI for a sample of about 1500 households. For the test, all respondents were also part of another LFS supplement: the Survey of Consumer Finances (SCF). SCF is a P&P income survey with content very similar to that of SLID.

One objective of the SLID test was to determine the best way to collect income data when using CAI.

2. DIFFERENCES BETWEEN CAI AND P&P INCOME INTERVIEW

One difference between CAI and P&P is that when using CAI, it is possible to tailor the questionnaire to various respondent groups. For the test, SLID developed three different income questionnaires (also referred to as paths or approaches). Prior to the test, each respondent was sent a notebook containing all the questions with references to tax documents when applicable. At the time of the interview, the respondent was asked if his/her notebook had been completed. If so, the interview was short. The respondent just had to indicate the lines in the notebook where amounts were recorded. The interviewer could scroll down the screen and enter the amounts on the corresponding lines.

Respondents who had not completed the notebook were asked if their tax form was handy. If so, they were sent through the tax approach. In this approach,

similar income types were grouped together and a more global question was asked first to determine if any of the income sources in the block applied. For example, for self-employment income, the respondent was asked if he/she had received any income from self-employment during the reference year, and if so, was asked more specifically what type it was (business, professional, commission, farm, fishing or other). If the respondent had not received income from self-employment, he/she would have skipped all questions related to self-employment.

Respondents who did not have their tax forms handy were sent through the block approach. This approach is similar to the tax approach, except that there were no tax line references.

Another major difference between CAI and P&P interviews is that the use of computers in interviewing allows interactive edits. For example, each time an amount is entered into the computer, a simple range edit can be done immediately while the respondent is still present, permitting instant verification of the amount, rather than having the same edit at the capture step, with no easy means of verifying the validity of the amount. As well, an edit for total income can be programmed to check that the sum of all items reported matches the reported total.

Another difference between CAI and P&P is the possibility to carry over information from a previous interview. For example, based on the January labour interview, SLID derives flags corresponding to four income items: wages and salaries, Unemployment Insurance, Social Assistance and Worker's Compensation. Hence, the wages and salaries flag is set if the respondent reported being a paid worker in January. At the time of the income interview, an edit pops up on the screen if the respondent does not report an amount for wages and salaries and the interviewer can prompt for an amount.

In a longitudinal survey, the previous year's information can be used to improve the edits. For example, the total income edit can be improved based on the previous year's total income.

3. EVALUATION OF QUALITY: linkage to tax data

Each year, Canadians fill in their income tax return form near the end of April. The income component of SLID is collected at the beginning of May, a time when respondents are more likely to remember everything about their income for the previous calendar year. Also, experience from a previous survey shows that, for self-employed people, it is much harder to respond to income questions at other times

during the year, yielding more non-response than when the interview is close to the tax deadline.

Even though there are content differences between household income questionnaires and the tax form, much of the current data quality evaluation done for income surveys compares survey data to tax data. For our data quality evaluation, we felt that linking to tax data would be appropriate as it would allow a direct evaluation of the data collected, at least for those income sources comparable to the tax form.

A direct link to tax data is not possible without the Social Insurance Number (SIN) of each respondent. In the absence of that number, we used other information such as the first and last name of the respondent, date of birth, marital status and postal code. These variables were used in a statistical record linkage with the CANLINK system.

In a few words, statistical record linkage will link records that have the "highest probability" of belonging to the same individual, based on the comparison of certain key variables.

To avoid getting non-matches due to spelling errors in the last name, we used the New York State Intelligence and Identification System (NYSIIS). This system encodes the name from its phonetic. Hence, two names that sound approximately the same will have the same NYSIIS code.

In this process, the matching variables used were the NYSIIS code for last name, sex, first and last name, date of birth, marital status and postal code.

Following the statistical linkage, we retrieved from the tax file, for each linked person, his/her spouse's name and SIN (if there was one). We were then able to link a few more persons, ending up with a match rate of 84%.

As for non-matches, we found by looking at their income data from SCF in 1991 that fifty percent of them reported an income of zero. Since the match was done using the 1991 tax file (the 1992 file was not available at the time), it is likely that these people were not on the 1991 tax file to begin with.

We examined the characteristics of the non-matches. Most were between 15 and 19 years of age and single (41% of non-matches) or aged 35 to 54 and married (12%). In all, 62% of the non-matches were in households where at least one member was matched, the rest were in completely unmatched households. Again, non-matches in partially matched households were usually 15 to 19 years old or married. Therefore, the hypothesis that these people were not on the tax file but rather were declared as dependents on their spouse's or parent's tax form is very plausible for at least a good portion of them.

4. MICRO-COMPARISONS

We were able to compare, for a sample of respondents and at a micro-level, survey data collected using P&P with tax data for 1991 versus survey data collected using CAI with tax data for 1992. We concentrate here on micro-comparisons to illustrate how CAI responses differ from P&P.

The micro-comparisons presented in this section are limited to three variables: wages and salaries (W&S), unemployment insurance benefits (UI), and interest and dividends (I&D). The choice of these variables was motivated by the fact that there was an interactive edit on CAI for W&S and UI. As mentioned earlier, based on our labour interview in January, a flag was set if a person was a paid worker, and another if a person was unemployed during the reference year. An edit would thus pop up on the screen if the respondent failed to declare W&S or UI when we expected him/her to. We hoped that this edit would solve some of the under-reporting problems, especially for UI which is believed to be under-reported in SCF. As for choosing I&D for this study, even though there was no such edit, this income category also has under-reporting problems.

Before reviewing the results of the micro-comparisons, we need to emphasize the limitations of this study. First, respondents to SLID were part of the SCF sample in the previous year. For that reason there might be a conditioning effect. Secondly, even though the income questionnaires for SLID and SCF were roughly similar, SLID had more detailed questions for certain income sources; the interview also included questions on wealth. This could have put a greater burden on the respondent. Furthermore, the material that was sent to the respondent prior to the interview was different for both surveys. SLID made more references to the tax form than the SCF. Finally, the response rate for the CAI survey was about 67%, while that of the P&P survey was about 83%. Based on our experience, there is a new non-response factor that is introduced in CAI surveys: transmission failures (we realize that this may happen in decentralized CAI surveys only). Since cases are transmitted between computers, some cases may be lost or receipt may be delayed, so they end up as non-response cases.

Despite the above limitations, it was felt that the evaluation could still give an idea of the quality of CAI data.

Interesting results were observed in the micro-comparisons. Table 1 shows the percentage of people who reported, for each variable under study, either no income in both the survey and the tax file, the same

amount in both, or a different amount.

Table 1. Micro-comparisons between survey and tax data for P&P versus CAI

W&S	P&P	CAI
survey & tax = 0	24.7%	29.6%
survey = tax	32.7%	32.9%
survey diff. from tax	42.5%	37.5%
UI		
survey & tax = 0	69.1%	70.5%
survey = tax	13.7%	11.9%
survey diff. from tax	17.2%	17.6%
I&D		
survey & tax = 0	50.7%	58.8%
survey = tax	16.2%	17.9%
survey diff. from tax	33.0%	23.2%

The percentage of people declaring no wages both in the survey and tax was a little higher for the CAI sample. This is due more to the two different survey years than to the collection method. In fact, the same pattern can be observed when looking at cross-tabulations of tax data from 1991 and 1992: there is an increase of 5% in the percentage of people who reported no wages in salary in 1992 compared to 1991. The same can be observed for I&D, but with greater differences between the two years. On the other hand, results for UI are very similar for both years, and were about similar for both years from the tax data.

Table 2, on the following page, looks at non-zero amounts in either the survey or tax data.

There seems to be more agreement between survey and tax data for CAI than for P&P, except for UI where the differences are not significant.

Also, there seems to be less under-reporting with CAI than with P&P for all three sources. It is possible that the flags for W&S and UI have helped the respondent to remember the different types of income received during the year. As for I&D, it is possible that the wealth questions in the notebook, especially the ones about money in bank accounts and savings, made the respondent think more about all of his/her income sources, resulting in less under-reporting of I&D.

There is an increase in the percentage of "don't know or refusal" answers for the I&D category with

CAI, but it is stable for the other sources. There is also an increase in the percentage reporting I&D in the survey only when using CAI.

Table 2. Non-zero amount in either the survey or tax

W&S	P&P	CAI
agreement within 5%	61.9%	67.6%
diff. by more than 5%	21.8%	24.0%
amount in tax only	10.4%	3.4%
amount = Dk in survey 3.8%	3.7%	
amount in survey only	2.1%	1.4%
UI		
agreement within 5%	51.3%	50.3%
diff. by more than 5%	21.8%	29.5%
amount in tax only	16.0%	10.7%
amount = Dk in survey 8.3%	8.1%	
amount in survey only 2.6%	1.3%	
I&D		
agreement within 5%	36.4%	47.4%
diff. by more than 5%	18.1%	20.2%
amount in tax only	41.4%	21.2%
amount = Dk in survey 1.8%	5.8%	
amount in survey only 2.4%	5.5%	

For SLID, the same results were cross-tabulated by the variable "path", to see if the quality of the data was different for the different approaches. Table 3, next, shows these results.

It becomes evident that the notebook and tax approaches yield better results than the block approach. This is probably because people who are willing to take the time to fill in their notebook use their tax form when doing so. On the contrary, people who do not have access to their tax documents give more approximate amounts, and probably forget some items. In fact, a study was done comparing people who used the three different approaches. It showed that people who went through the notebook or tax approach reported more income sources than people who went through the block approach.

The results from Table 3 show very good agreement between the survey and tax data for the notebook and tax approaches. Also, the percentage giving no amount in the survey for a particular source, but having an amount on their tax file is much higher among those who go through the block approach.

Table 3. CAI results by approach for non-zero amounts

W&S	Nbook	TAX	BLOCK
agreement within 5%	83.2%	85.8%	36.2%
diff. by more than 5%	11.8%	11.8%	47.2%
amount in tax only	3.4%	1.18%	4.88%
amount = Dk in survey 0.0%	0.0%	0.0%	10.6%
amount in survey only 1.7%	1.2%	1.2%	
UI			
agreement within 5%	73.7%	80.4%	16.41%
diff. by more than 5%	15.8%	10.7%	50.0%
amount in tax only	7.9%	7.1%	14.8%
amount = DK in survey 0.0%	0.0%	1.8%	18.0%
amount in survey only 2.6%	0.0%	0.8%	
I&D			
agreement within 5%	66.3%	70.3%	9.5%
diff. by more than 5%	16.0%	17.6%	29.26%
amount in tax only	13.1%	8.8%	36.7%
amount = DK in survey 0.0%	0.0%	1.1%	15.7%
amount in survey only 4.6%	2.2%	8.8%	

Next, table 4 relates to people with a flag indicating the presence or absence of income from W&S or UI. We kept only people who responded to the labour interview, which is where the flags come from.

Table 4. Distribution of amounts by presence or absence of flag

W&S	Presence	Absence
agreement within 5%	68.3%	7.9%
diff. by more than 5%	24.4%	1.2%
amount in tax only	1.3%	4.3%
amount = Dk in survey 3.7%		0.6%
amount in survey only 0.5%		2.1%
survey & tax = 0 1.9%		83.8%
UI		
agreement within 5%	53.9%	5.6%
diff. by more than 5%	31.4%	2.7%
amount in tax only	3.7%	2.9%
amount = Dk in survey 7.9%		1.2%
amount in survey only 0.0%		0.5%
survey & tax = 0 3.1%		87.1%

Interestingly, in most cases where the flag is present (indicating that there should be an amount), an amount was generally reported during the income interview (93.2% for W&S and 85.3% for UI, corresponding to the sum of percentages from the first, second and fifth categories in the table). Also, when there is an amount in the survey, tax data confirms the accuracy of the flag in almost all cases for W&S and in all cases for UI. As well, when the flag is present but there is no amount reported, tax data indicates in a majority of cases that there should have been an amount (we confirmed that cases where there was a "don't know" or "refusal" answer in the survey have an amount in tax, except for one case for UI). Also, there are only a few cases where the flag is present when it should not be. Therefore, it seems that the presence of the flag is in general a good predictor for the presence of an amount.

On the other hand, when the flag is absent but there is an amount reported in the income survey, tax data confirms that an amount was expected in a majority of cases. This indicates that the respondents did not report being paid workers (or UI beneficiaries) in the labour interview when in fact they should have. Hence, there may be a need to impute the flag in such cases, as well as all the information that goes with it in the labour interview. As for cases where there is no flag and no amount reported in the survey, a small percentage do have an amount in the tax file (4.3% for W&S and 2.9% for UI). There is no way we could have known that an amount was expected for these cases.

Furthermore, we can consider imputing an amount when there is a refusal to an item since we observed that there was almost always one on tax.

We also looked at these results by approach to see if there were any differences in the pattern of answers. We noticed that even though the flag seems to help in reporting an amount independently of the approach, the precision of the amount varies from one approach to another. For example, among people taking the block approach and with a W&S flag present, only 37% reported an amount that is within 5% of the amount from the tax file. For the notebook and tax approaches, this percentage is around 85%. The same effect is observed with people for whom a UI flag is present.

5. RECOMMENDATIONS

The implementation of surveys using CAI is still fairly new at Statistics Canada and more testing is needed before firm conclusions can be made. Because proper tests could not be done between P&P

and CAI, our conclusions are limited and certain effects may be confounded. However, the experience gained by this test allowed us to make certain decisions that have an impact on the design that will be implemented for production.

There is a definite difference in data quality depending on how people report their data. People who used the notebook or tax approach gave better quality data. It is not clear however if the three different approaches programmed to accommodate respondents and interviewers have helped or not. For P&P, it is usually estimated for income surveys that 40% of people fill in a questionnaire before hand (the equivalent of the notebook approach). Therefore, the improvement in data quality obtained through CAI may be partially due to the fact that the "tax driven" application motivated interviewers to ask respondents to refer to their tax documents.

Even though data quality seems to be better with both the notebook and tax approach, only one application will be programmed in production, one which is similar to the notebook approach. The reasons for this are first, to simplify the collection instrument as much as possible for interviewers, and second, because response rates were much lower with the CAI test than what is usually observed with P&P for income surveys.

The lower response rate could be due to the fact that to get income sources similar to tax meant that they had to be broken down. At the same time, we also tried to collect wealth data in the test. The combination of the two meant twice the number of questions for the CAI application than for the P&P interview. It is felt that reducing the number of questions will probably bring the response rate up. However, since the tax approach gave good results in terms of data quality, interviewers should be trained to encourage respondents to use records whenever possible.

Dependent interviewing was introduced by setting flags in the computer environment. The flags derived in the January interview indicated whether an amount was expected for a specific item in the May interview. When the interview was conducted, if no amount was entered for those fields, the flag triggered an edit to ask the respondent if the item had been forgotten. Analysis of the flags showed three types of results.

First, when a flag was set to true (i.e. an amount was expected), an amount was reported on the survey, or marked as "don't know" in most cases. This greatly decreased the proportion of cases where no amount was reported in the survey but an amount was reported in tax. When the flag was set but no amount

was reported, half of the time it looked like the flag was set by error (there was no amount on the tax file). Second, when the flag was set to false (which meant that nothing in January led us to believe that an amount should be expected), and an amount was reported for that source, the tax data seemed to indicate that we should believe the amount (there was also an amount reported on the tax form). Third, even if reporting of items has improved with the addition of the flags, there are still some data quality issues for the amounts for respondents who were interviewed through the block approach. The amounts are often rounded or a don't know to the amount is given.

Hence, the application used in production will use the flags to help the interview. As for data quality, we plan on continuing our evaluation of the test data, as well as doing evaluation on a continual basis.