



SSSSSSSSSSSSSSSS

**FIRST WAVE ROLLS IN**

\$ There is still much to do before the first wave of data are "on the street", but a major step is now behind us: the collection of the first full year of annual data has been completed and the data are now here being processed. It is still too early to talk in detail about the results but we have some indicators on how things went in the field, and can share these. This was our first full field data collection, so even though the results are great, we expect them to improve.

**Response information**

Final, exact response rates will not be known until processing has been completed—for example, we may convert someone from "respondent" to "non-respondent" because not enough real information was provided during the interview. In a panel survey, response rates are never a simple matter. But SLID may make other panel surveys seem simple by comparison because of its two-phase approach, with labour information collected in January and income in May.

**Editor's Note**

Although the summer months tend to be devoted more to leisure activities, the SLID message must be getting around. We are now receiving several requests weekly from people wishing to subscribe to *Dynamics*. Many are also purchasing subscriptions to our Research Paper Series either on paper (\$50/year) or diskette (\$15/year). Welcome to our new subscribers; we hope that SLID will meet your expectations.

Those who have been with us longer have been patiently waiting for data. Our first SLID data product will soon be available. More details are in this issue of *Dynamics*. We are excited about reaching this next important "kilometre-stone".

Philip Giles

The sample for the labour interview consisted of about 42,000 persons, counting both adults and children. This total included 40,000 longitudinal respondents—people living in dwellings selected for the first panel in January 1993. The remaining 2,000 or 5% were *cohabitants*. This is the term given to people now living with longitudinal respondents who were not doing so when the sample was originally selected. The response rate is about 86%. In 6% of the cases, there was a definite non-response (for example, cases that were "unable to trace" or refusals). The remaining 8% were not completed or not obtained because of technical problems in the computer-assisted interviewing system. Some of these cases may yet be retrieved through the back-up diskettes used by the interviewers.

If the May income interview is considered in isolation, a response rate of about 83% was obtained. The number of cases not received due to technical problems dropped to 3%; again some of these may be retrieved. Refusals and "unable to trace" cases represented about 9% of the total. There was an additional 3% non-response for other reasons. Under 1% were not eligible for the income interview—for example, people who had moved into an institution or abroad, or who had died.

The labour and income interviews deal with related but distinct subject matter, and share the same reference period. They are separated into two interviews only to optimize data quality. Specifically, we believe that the labour information is best collected close to the end of the reference year to minimize recall error, while income is best collected at tax time when respondents can refer to their income tax records. The labour and income interview together make up one wave.

A respondent who completes an interview for labour but not income, or vice versa, is a partial respondent to the wave as a whole. Our best estimate currently is that, of the 42,000 longitudinal respondents and cohabitants, about 33,000 were old enough to complete a labour and income interview (that is, they were aged 16 or over on January 1, 1994). Of this total, roughly 72% completed both a labour and an income interview, 14% responded to labour but not income, 6% responded to income but

not labour, and 8% responded to neither. Thus our *wave response rate* is about 92%. It must be stressed that these are very early results and could change somewhat as data cleaning progresses.

**Looking ahead**

Some 500 joiners (new cohabitants) were identified in May. Because these people were not present in the household at the beginning of the year, they were not eligible for this year's wave of labour and income information. However, they have been identified as new household members and, assuming they are still sharing a dwelling with a longitudinal respondent next year, their labour and income information will be collected for the second wave.

A final note: The data collection operation is concentrated into two short periods each year, but there is quite a lot of field work that goes on between. For example, respondents will be sent a bulletin before the second wave, providing them with some feedback on data collected in the first wave. Also a letter was sent thanking respondents for their participation in the first wave. We included a change of address card with this letter, and have received about 250 of these cards back from respondents indicating that they had or would be moving. This is an encouraging and unexpected level of response, and will no doubt be very useful in the second wave of data collection.

\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$

**FIRST DATA PUBLICATION:  
PRELIMINARY INTERVIEW  
RESULTS**

§ This fall, SLID will release its first data publication, in anticipation of the full first-wave microdata file and extensive analytical results in the spring of 1995. Its scope will therefore be modest, highlighting the key types of information collected in the preliminary interview in January 1993. The 1993 preliminary interview recorded the background characteristics of our first panel of respondents, so the upcoming publication will provide users with a chance to "get to know" the SLID respondents, particularly their human capital and demographic profiles, before taking on the wide range of activity and income data that will become available later.

In addition, we hope the publication's few short articles will whet users' appetites for reading more or developing their own studies later. In some cases, the background classification variables contained in the preliminary interview are unique from those covered in other surveys, for example accumulated work experience, parents' educational background and number of children born/raised. For the more standard variables, this exploratory publication will refer to benchmark findings from other surveys to assess SLID data quality and the limitations resulting from sample size. These are the topics we plan to cover:

1. Gender wage gap

This study will cross the current (January 1993) work information of the preliminary interview with background classification variables to examine the male-female wage gap while referring to some existing research on the subject. It will take advantage of SLID's coverage of a wide range of variables important in controlling for "legitimate" reasons for male-female wage differences, such as age, schooling, occupation, hours and tenure. Feeding into this study will be results on an additional control variable not usually available...(read on!)

2. Work history

Accumulated work experience is not determined simply by age and years of schooling, but that is how researchers usually have had to estimate it (age - years of schooling - 5). In other words, they assume continuous employment immediately after school onward. How do the results of the work experience questions in SLID differ from this approximation, with obvious implications for research? Are there population subgroups (apart from the obvious one of women with children) for whom the formula is somewhat misleading? By shedding some light on this aspect of human capital, it is hoped that the work experience variable will not only enhance the quality of studies based on SLID data but also increase options for other research.



### 3. Intergenerational mobility with respect to education

Parents always hope their children will do better than they do, and a general rise in educational attainment suggests that in this respect they usually do. The SLID preliminary interview data will allow us to measure such intergenerational mobility (in both directions) more precisely. Furthermore, we'll be able to examine the characteristics of adults who have outdistanced their parents in terms of educational attainment and those who have lost ground relative to their parents (after controlling for the general rise in education).

### 4. Employment equity data

This article will describe the information available in SLID on two of the four employment equity "target groups": aboriginals and visible minorities. Researchers interested in employment equity will want to know about data quality and the potential of SLID with respect to other surveys. The article will examine data quality by comparing with benchmark data, and will discuss quality constraints such as survey design and sample size. It will also look at how data on minority groups from other longitudinal surveys have been analyzed.

So that SLID followers are aware of all the preliminary interview topics, two short notes will summarize the results on marital history and the number of children born/raised. Finally, we will include an overview of the content and expected uses and benefits of the data.

Readers may remember that the February issue of *Dynamics* foresaw the summer release of a microdata file from the preliminary interview. The upcoming analytical publication has replaced those plans, permitting us to stay on target for releasing the full microdata file and major results in the spring of 1995.

§§§§§§§§§§§§§§§§§§

## FEASIBILITY STUDY OF THE USE OF TAX FILE DATA

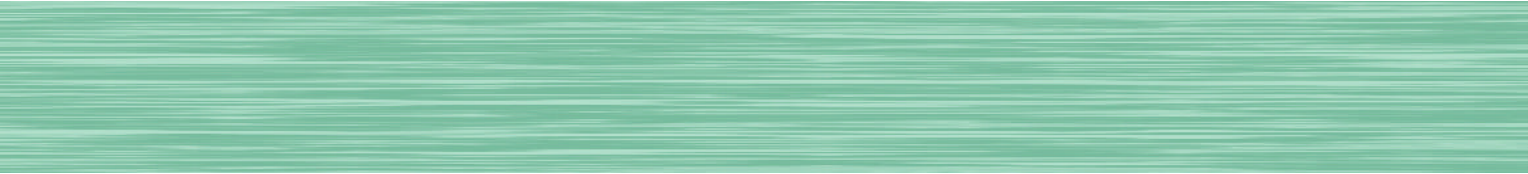
§ Several members of the SLID team have examined one potentially significant way to cut survey costs and reduce respondent burden while improving data quality. SLID Research Paper 94-11 **The use of tax file data in the Survey of Labour and Income Dynamics** examines the feasibility and benefits, in the case of SLID, of using income tax files as an alternative to collecting income data by interview each year.

A major reality in using the tax file source is that it would not completely replace the traditional interview approach for income topics. This is because not everyone files an income tax return and some respondents would not choose this method of providing information to the survey. At the outset then, only a portion of the data will be affected, and procedures will remain the same as regards many respondents, apart from adding the consent question for everyone.

Just how many respondents might give Statistics Canada their consent to refer to information already provided to Revenue Canada? This was tested at least in an exploratory way by posing a simple, hypothetical question to SLID's first panel of respondents in the May 1994 income interview. Just over half (53%) said they would agree to let Statistics Canada access their tax records to obtain income data. This result is fairly consistent with a similar test conducted in August 1993 during the Labour Force Survey interview. It is possible that the consent rate will rise when the question is no longer hypothetical, because they will receive more information stressing the benefits of responding to the survey this way.

The consent question is important since respondents must be fully informed of what is being requested, even when another household member is responding for them. We plan to send an information package prior to the income interview. Respondents would have the option to withdraw permission at any time.

What are the benefits of using tax file records to the extent possible? First, the quality of data from this source is considered better than respondent information for a number of reasons. Non-response to all or certain income questions could be reduced, and fewer respondents would be lost through attrition over the course



of the survey (six years). Also, it is known that income amounts tend to be underreported in household interviews relative to amounts recorded for income tax purposes.

Some studies within SLID have been undertaken to estimate the difference in data quality between the tax record source and direct survey information. For purposes of data quality evaluation only, tax file data as well as survey data (1993 SLID field test and 1992 SCF) were obtained for a small sample of respondents for two consecutive years. Longitudinal data quality was estimated by focussing on changes calculated from each two-year data source. The tax file source was considered the benchmark. It was found that the survey data tended to overestimate changes in the receipt of income, and that this is probably because sources of income tend to be omitted more often. Overall, it is felt that data quality would be improved by using tax file data as much as possible.

Certain other benefits of using tax data to obtain income information in the case of SLID are fairly clear and can be roughly quantified. The possibility of reducing respondent burden is very attractive in a longitudinal survey where the same people are interviewed many times over six years. For those respondents who give their consent to access tax data, the number of interviews over six years could be reduced from 13 to 7. The potential to reduce costs has of course been considered. For example, it is estimated that a 50 percent reduction in the number of households interviewed would result in a 40 percent cost savings with respect to the field costs of the income interview.

Related to data quality is the issue of mixing sources of information on income. First, conceptual or content differences exist. For example, household surveys like SLID typically use a "money income" concept which therefore excludes many fringe benefits as a component of wages and salaries, unlike tax data. SLID asks for some non-taxable items not included on income tax forms, but some of these can also be estimated. Second, even if the content differences are negligible or can be overcome, are there other problems with mixing data collected by different methods? Perhaps the direct survey data are not all that homogeneous - being a mixture of recall, reference to tax returns and proxy reporting - so that the results will not be seriously affected by the addition of

higher quality data generated purely from tax returns. However, the issue needs further consideration and feedback from users would be welcome.

§§§§§§§§§§§§§§§§

**CONFIDENTIALITY  
PROTECTION FOR SLID  
MICRODATA**

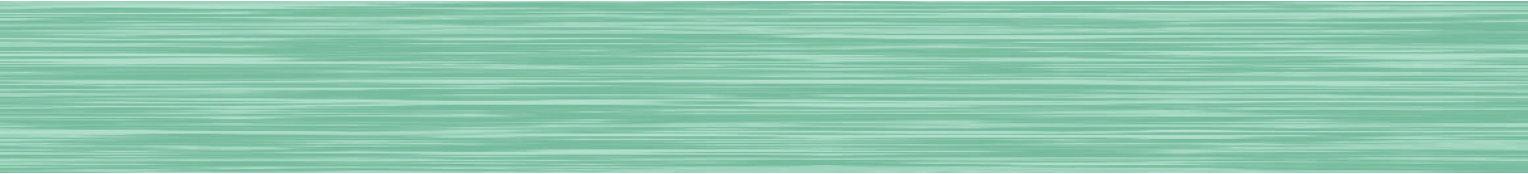
§ In preparation for data dissemination, some background research has been done on the principles of confidentiality for SLID data released in microdata files. We must have safeguards to prevent the identification of any person, either directly or indirectly. The issue of confidentiality is more complex for longitudinal surveys because the output data must satisfy confidentiality constraints over a number of years as well as for each instance of data collection.

Although SLID microdata files are viewed as an important survey output, we realize that some users will require access to more detailed data. These users will have two choices:

- A customized output service will be available. These customized outputs will be screened for confidentiality before transmission to the client.
- Exploratory work is under way to provide users with some type of remote access to unscreened SLID microdata files. This is different from the previous approach in that the user writes and submits the data extraction program. The output, however, will still be screened before release to the client.

In formulating a plan for confidentiality protection in microdata files, the goal is to avoid the unique identification of a person using SLID data variables, either by themselves or in combination. Two basic approaches have traditionally been suggested for protecting confidentiality: data modification (changing actual values to protect an individual's identity while ensuring minimal impact on any analyses) and data reduction (reducing the information content, such as grouping responses or suppressing a variable entirely). Because data modification tends to bias results on changes in individuals' characteristics over time, data reduction will be used to eliminate most disclosure risk for SLID.





Data reduction can take many forms. To optimize data availability, a mixture of approaches will be used. The principal approaches will be:

- Exclusion of variables from the file when the risk of disclosure is high or when the variable has limited analytical value. Examples are respondent's name and detailed geography of residence.
- Recoding to an anonymous coding structure. Variables related to the sample design, such as stratum, are required for accurate variance calculations. For analytical purposes, it is sufficient to identify which respondents are in the same stratum. However, the coding structure conveys some geographical information. Recoding stratum codes to another "meaningless" coding structure provides the analyst with no loss of information but reduces the risk of disclosure.
- Placing restrictions on population size. Minimum thresholds on the number of sample members are established for various subpopulations for which external data sources exist or which can be easily identified by an outsider.
- Down-scaling. Grouping numerical values so that the resultant variable is categorical (or qualitative).
- Collapsing categories. This is similar to the previous approach, except that it involves reducing the number of categories for qualitative (or categorical) variables by grouping.
- Top and bottom coding. Very high and very low values usually are rare in the population, thus increasing the risk of disclosure. All extreme values are recoded to a value which only indicates that the actual value is extreme.
- Rounding. Values, particularly monetary ones, can be rounded to a higher level. This allows minimal effect on analysis but a significant reduction in disclosure risk.

Although the first microdata file will contain data from only the Preliminary Interview and the first year of labour and income data, SLID is attempting to identify confidentiality procedures applicable to the entire range of six years of data. Although some work will be required prior to the release of subsequent cumulative data files, it will not be necessary to conduct the entire range of evaluations each year. For more information see Research Paper No. 94-14 or contact Pierre Lavallée at (613) 951-2892.

SSSSSSSSSSSSSSSS

## RESEARCH PAPERS

\$ The following are recently released Research Papers which can be ordered individually or received automatically. For more information, contact Anne Palmer by phone at (613) 951-2903, by fax at (613) 951-3253, or by mail at 11-D8 Jean Talon Building, Tunney's Pasture, Ottawa, K1A 0T6. Internet users can contact us at GILES@STATCAN.CA.

*94-08 Questionnaire and collection procedures for SLID income data collection - May 1994*

*Élaine Fournier, Susan Poulin*

In May 1994, SLID collected income data from our first panel of respondents using computer-assisted interviewing. Unlike other SLID interviews, a paper questionnaire is mailed to respondents prior to the interviewer's telephone call, to allow time for consultation of financial records. This document provides a description of the data collection procedures and income question wordings, as well as some rationale for the decisions that were made. It contains a copy of the questionnaire mailed to respondents and outlines how the interviewer collects the information.



*94-09 SLID Questionnaire for Demographics and Contact: 1994  
Ruth Dibbs, Bob Loverock, Alison Hale*

*94-10 1994 Preliminary Interview Questionnaire  
Alison Hale, Debbie Lutz, Mike Brule*

These two documents provide a paper "questionnaire" outlining the question wordings, response categories, and question flow pertaining to SLID data collection using computer-assisted interviewing. The questions in the Demographics and Contact modules are included in both the annual labour and annual income interviews. Information on household membership is collected in Contact. The Demographics module includes questions on: date of birth, sex, marital status, and household relationships. The Preliminary Interview is given to new household members and persons who turned 15 during the reference year.

*94-14 Confidentiality of SLID Microdata: General Approach  
Pierre Lavallée, Chantal Grondin*

A general overview of principles of confidentiality is provided along with a discussion of how these principles are applied in practice. The overall direction proposed for SLID is also described. Further Research Papers will outline detailed developments on this topic.