

Catalogue No. 93-19

**SAMPLE REPRESENTATIVENESS
FOR THE SURVEY OF LABOUR AND INCOME
DYNAMICS**

December 1993

Pierre Lavallée, Social Survey Methods Division

The SLID Research Paper Series is intended to document detailed studies and important decisions for the Survey of Labour and Income Dynamics. These research papers are available in English and French. To obtain a summary description of available documents or to obtain a copy of any, please contact Philip Giles, Manager, SLID Research Paper Series, by mail at 11-D8 Jean Talon Building, Statistics Canada, Ottawa, Ontario, CANADA K1A 0T6, by INTERNET (GILES@STATCAN.CA), by telephone (613) 951-2891, or by fax (613) 951-3253.

EXECUTIVE SUMMARY

SLID is a longitudinal survey, collecting data for the same sample of respondents over several years. Among the many new issues which arise with a longitudinal survey are the procedures required to ensure that the survey results are representative of the population. Special procedures are necessary during data collection. As well, the production of sampling weights (used for estimation) is complicated due to the fact that the Canadian population constantly changes while the sample only changes periodically.

This document discusses the issues and the procedures developed to handle them. The mathematical treatment of the derivation of the sampling weights is left to another document, and not treated here.

1. Introduction

The Survey of Labour and Income Dynamics (SLID) will follow each panel of respondents for six years. While annual (cross-sectional) estimates will be produced, the survey's priority is clearly the longitudinal aspect, and the analytic capabilities that this offers.

This paper will discuss the survey plans relevant to ensuring that SLID estimates will be representative of the target population, and will describe the various sampling weights which will be produced. The mathematical treatment for the derivation of the sampling weights will follow in another research paper.

2. Sample Design

The first panel of respondents was selected from households rotating out of the Labour Force Survey (LFS). Although no decision has been taken on subsequent panels, in this document it is assumed that they will be selected in the same way. The first panel was introduced into the survey in January 1993, and the second will be in January 1996 (see Table 1). These two panels will cover reference years 1993 - 1998 and 1996 - 2001, respectively. With the introduction of the second panel, the SLID sample size will reach its target (around 40,000 households), and remain there. In January 1999, another panel will be introduced (to cover reference years 1999 - 2004), replacing the first panel. Following this pattern, a panel will be replaced in the sample every three years.

When a panel is introduced, all persons living in the selected households at the time of sample selection, regardless of age, become members of the SLID longitudinal sample. These persons remain so for the life of the panel, even if

TABLE 1

SAMPLE ROTATION PLAN														
Year														
	93	94	95	96	97	98	99	00	01	02	03	04	05	06
Panel														
1	P	I	I	I	I	I	I							
2				P	I	I	I	I	I	I				
3							P	I	I	I	I	I	I	
4										P	I	I	I	I
P = Preliminary Interview I = Labour and Income Interviews														
<i>Each panel has 15,000 households at the outset.</i>														

they move. No additional persons will become members of the longitudinal sample for that panel. Thus, for each panel, the longitudinal sample is fixed when the panel is introduced (i.e., January of the first reference year covered by the panel) and remains fixed for the life of the panel.

3. Data Collection

This section will discuss aspects of the data collection that are particularly relevant to the maintenance of sample representativeness. More details on SLID data collection are available in other SLID Research Papers.

When a panel is introduced into SLID, a Preliminary Interview is conducted with each member aged 15 and over of the selected households in January. For the next six years, starting with the year after the panel is introduced, data is collected on labour and income. Note that a typical longitudinal respondent will be contacted thirteen times over seven years: the first year for the Preliminary Interview and in each of the subsequent six years for the Labour and Income Interviews. In January, the Labour Interview is conducted while the Income Interview is conducted in May. In both cases, the data are collected using the previous calendar year as the reference period. Labour data are collected as soon as possible after the end of the reference year to minimize recall errors. The primary reason for collecting the income information in May is to enable respondents to refer to their completed income tax forms.

For the Labour and Income Interviews, data will be collected for all persons aged 16 and over in the longitudinal sample, as well as for all "cohabitants" (persons residing with a member of the longitudinal sample on January 1 of the year of interview), provided that they meet the conditions of eligibility outlined below in Section 5. Cohabitants will be interviewed for SLID only as long as they reside

with a member of the SLID longitudinal sample. The sole exception is when a cohabitant moves away between the January and May interviews. Since he/she was still a cohabitant on January 1, it is necessary, for both longitudinal and cross-sectional purposes, to trace this person and administer an income interview.

4. Target Population

The target population for SLID is all persons, regardless of age, residing in Canada, **excluding** residents of:

- ! Yukon and Northwest Territories;
- ! institutions (length of residence exceeding six months);
- ! Indian Reserves;
- ! military barracks.

5. Eligible Population

The following criteria define the group of respondents for whom labour and income data are collected in 1994 and later years:

- ! Preliminary Interview: Cohabitants aged 16 years or older as of January 1 who joined the household of a longitudinal respondent since the last interview; all longitudinal respondents (and previously-identified cohabitants) who were 15 years old on January 1 of the year of interview.
- ! Labour and Income Interviews: All persons in the SLID longitudinal sample and all cohabitants aged 16 years or older as of January 1 of the year of interview.

A cohabitant who moves into the household between January and May will not receive an Income interview that year, since the person was not

residing with a longitudinal respondent during the reference year. He/she would receive a Preliminary Interview in May, and then Labour and Income interviews the following year (provided that he/she was still residing with a longitudinal respondent in January).

Although they are not eligible for the labour and income questions, basic demographic information will be collected for all persons under 15 years of age.

There are two main reasons for doing so:

- ! From an analytic point of view, certain family and household characteristics will be important as the number and characteristics of children influence the labour-market behaviour of all household members and also influence the household wealth;
- ! Some of these individuals will become "eligible" during the life of the panel, and therefore will be interviewed for SLID.

6. Target Population: Entrants and Leavers

Although the longitudinal sample is fixed at the point of sample selection, the composition of the target population is continuously changing (i.e., not the definition, but the actual people). For this reason, discussions of representativeness must refer to a particular point in time.

Entrants to the SLID target population will belong to one of three groups:

- a) Persons who have moved from Yukon or Northwest Territories, or immigrated from another country since the panel was introduced;
- b) Persons who were residents of an institution, Indian Reserve or military barrack when the panel was introduced, and who are now living in a dwelling that is in the sample frame;
- c) Persons born after the panel is introduced.

Conversely, leavers from the target population will also belong to one of three groups:

- a) Persons who move to Yukon or Northwest Territories, or emigrate to another country after the panel is introduced;
- b) Persons who move to an institution, Indian Reserve or military barrack after the panel is introduced;
- c) Persons dying after the panel is introduced.

7. Longitudinal Representativeness

For SLID, longitudinal representativeness is defined in relation to the composition of the target population on January 1 of the year in which the panel is introduced (i.e., 1993, 1996, 1999,...). Therefore, provided that each panel is representative of the target population when it is introduced, longitudinal representativeness is achieved. Changes in the composition of the population are not relevant to longitudinal representativeness.

There are two issues related to this:

! Effect of two panels in survey design

At any time (after the second panel is introduced), the SLID sample will consist of two panels. Each panel will be representative of the target population, but at two different points in time. In addition to a six-year panel weight, each panel (after the first panel) will have two three-year panel weights. To derive the latter, the overlapping portion of two successive panels will be collapsed and treated as if it were one three-year panel. The choice of weights to use for estimation will depend on the analyst's requirements (in general, whether a three year or a six year time period is needed). For example, a longer time frame allows one to study such

things as: durations of time spent in a particular state (eg., time a person spends unemployed); critical events and important transitions (eg., marital separation with information leading up to, and following from this event); cyclical events where a cycle may be several years in length. On the other hand, the sample size of one panel may not be adequate for many population subgroups of interest. For these analyses, users may opt for the three-year time frame, which affords a larger sample by amalgamating results for two adjacent panels. Among the many population subgroups of interest are: single parents with preschool-aged children, and youth making the transition from full-time education to full-time labour force. Also, some analyses directed at a particular level of geography may only be possible using both panels of respondents.

In fact, longitudinal weights could be produced for every year. For example, a person moving in with a longitudinal respondent during the first year of the panel is a cohabitant, and therefore will have a six-year longitudinal weight of zero. Any six-year longitudinal analysis will not incorporate any of this cohabitant's data, except possibly as household or family data for a longitudinal respondent. Six years of data may be available for this cohabitant if he/she remains in the household for the life of the panel. If a five-year longitudinal weight was produced, the data for this person could be included in the analysis (albeit with a five-year time frame rather than a six-year one). One problem with producing longitudinal weights for every year is that SLID will not trace cohabitants if they cease to live with a longitudinal respondent. Therefore, following the same example, if the cohabitant moves out in year three, the five year longitudinal weight remains but no data will be available for years four, five and six. No decision has yet been made regarding longitudinal weights for every year.

! Effect of non-response

As with any survey, non-response by sample members affects the representativeness of the data. For longitudinal surveys, sample attrition (caused, for example, by persons stating that they do not wish to be contacted again) is an added factor. Discussion of these problems and their treatment for SLID will be left to a future paper.

8. Cross-sectional Representativeness

Although changes to the composition of the population do not affect the longitudinal representativeness, cross-sectional estimates (i.e., estimates referring to a one-year time period) are a SLID requirement. Therefore, sample weights must be defined for each year, with representativeness defined as of January 1 of the year of data collection (i.e., at the end of the reference year).

Cross-sectional weights must account for cohabitants, who can be one of two types:

- !** They can be "initially-absent individuals": persons who could not have been selected in the SLID longitudinal sample because they were not part of the target population when the panel was introduced. These are entrants to the target population.

- !** They can be "initially-present individuals" who were part of the target population when the panel was introduced but who were not selected for the SLID longitudinal sample. They "joined" a household with a member of the SLID longitudinal sample, for example, through marriage.

Cohabitants will be interviewed as long as they reside with a member of the longitudinal sample.

Treatment for each of the three categories of entrants and three categories of leavers, as outlined in Section 6, is discussed below. In general, handling entrants to the target population is a more difficult task than handling leavers. The decision to interview cohabitants helps in this, and the requirement for cross-sectional estimates was a factor which led to the decision to interview cohabitants for SLID. (Interviewing cohabitants is necessary also to allow the calculation of family and household variables (for example, family income) required for longitudinal analysis.)

Leavers:

- a) Persons moving to Yukon or Northwest Territories, or to the United States: To the extent possible, these persons will be traced and interviewed. (There are some issues to resolve yet regarding the collection of income data from respondents who move to the U.S., as some income sources would not be appropriate.) Regarding persons moving to other countries, tracing will be done to identify those who subsequently return. Those who do will be treated the same as other members of the SLID longitudinal sample.
- b) Movers to an institution, Indian Reserve or military barrack after the panel is introduced: Persons entering an institution will be traced, but not interviewed. Those who return will be treated as other members of the SLID longitudinal sample. Persons moving to an Indian Reserve or into a military barrack will be traced and interviewed if feasible.

- c) Persons dying after the panel is introduced: Obviously, no further interviews will be conducted. Data collected for these persons before their death will be kept on the data file.

Entrants:

- a) Movers from Yukon, Northwest Territories or immigrants from another country: This group is perhaps the most difficult one to deal with for SLID. Two sub-groups (mutually-exclusive and exhaustive) are of interest:

- ! entrants joining a household where at least one of the household members was in the target population when the panel was introduced;

- ! entrants in new households or joining households where no members were in the target population when the panel was introduced.

By interviewing cohabitants, SLID will have a representative sample of entrants joining a household where at least one member was in the target population when the panel was introduced. Since information on previous residence and immigration is collected for SLID, it can be used in the weighting of such persons interviewed for SLID.

Accounting for those households where no member was in the target population when the panel was introduced would require a separate sample. Three options have been suggested:

- ! Since the original SLID sample was drawn from a sample of dwellings, one could return to all original sampled dwellings and

determine whether the residents are all immigrants. Those who were would be interviewed for SLID. One obvious drawback is that new dwellings are not considered.

- ! A question could be added to the January interview of LFS to identify recent arrivals from the Canadian territories or from other countries. This would provide a sample of households which could be added to the panel. Households with a blend of people who were and some who were not in the target population when the longitudinal sample was selected are already represented in SLID; only households composed solely of new entrants would be of interest;

- ! A dual frame approach could be adopted using immigration records.

Tables 2 and 3 below show the importance of immigrants in terms of households and individuals, based on 1991 Survey of Consumer Finances data. It can be seen that only 0.30% of households contain only new immigrants. For SLID Panel 1 with a sample size of 15,000 households, an estimated 45 new households would contain new immigrants only, which makes the first option clearly not a cost-efficient one. The other two options offer more promise. However, as the evaluation and implementation of these approaches would draw funds from other areas of the survey, it was decided not to have a separate sample of immigrants, at least for the first panel.

Table 2: Households with New Immigrants, 1991

	Unweighted		Weighted	
	Count	%	Count	%
No new immigrants in household	43,244	99.62	9,842,633	99.23
At least one, but not all, new immigrants	98	0.23	47,298	0.48
Everyone in household new immigrants	65	0.15	29,484	0.30
TOTAL	43,407	100.0	9,919,415	100.0

Data from April 1991 Survey of Consumer Finances, excluding persons under age 16; new immigrants are defined here as people who immigrated to Canada at some time between January 1990 and April 1991.

Table 3: 1991 Persons by 1990 Immigrant Status

	Unweighted		Weighted	
	Count	%	Count	%
Did not immigrate in 1990	89,476	99.71	20,591,309	99.40
Immigrated in 1990	258	0.29	124,758	0.60
TOTAL	89,734	100.0	20,716,067	100.0

Data from April 1991 Survey of Consumer Finances, excluding persons under age 16.

- b) Movers who were residents of an institution, Indian Reserve or military barrack when the panel was introduced:
- As with immigrants, the collection of data from cohabitants allows some accounting for part of this group. Also similarly, a separate sample of households could in principle provide representation for households where no member was in the target population when the panel was introduced. However it would be problematic to ask about whether a person had been in an institution at a specific date in the past, because the question is sensitive and the definition of "institution" is not clear cut. SLID will not

identify joiners who moved from an institution. Some loss of representativeness will occur, although it is believed that it is negligible.

c) Persons born after the panel is introduced:

Newborns are easily identified in a household, since date of birth is collected for SLID. Therefore, newborns who "join" households where at least one person was in the target population when the panel was introduced can be easily handled in the weighting (in an analogous manner to immigrants). An entire household of newborns is virtually impossible, so this situation need not be addressed. Newborns who "join" households where no member was in the target population when the panel was introduced can be handled in the same way that the entire household will be handled.

9. Survey Weights

This section briefly describes the weights which will be available to SLID data users.

Longitudinal Weights

Longitudinal weights will be calculated when a new panel of respondents is introduced. A longitudinal weight will be assigned to all longitudinal respondents in the SLID sample. These weights will be assigned based on the population estimates on January 1 of the first reference year.

For weighting, it is important to understand that data for any given reference year will be collected from two panels of respondents (after the introduction of the second panel). This is true even in a year when a panel is introduced (after the

second panel), when data will be collected from three panels of respondents. Table 1 above illustrates this important concept.

As mentioned in Section 7, three longitudinal weights will be assigned to any given longitudinal respondent, a six-year longitudinal weight and two three-year longitudinal weights. For a given longitudinal respondent, these weights will be calculated according to the compositions of the target population on January 1 of the first and fourth reference periods covered by the panel to which the respondent belongs. Cohabitants will have a zero value for their longitudinal weights.

An example, referring to Table 1, should help to explain these three longitudinal weights. To avoid the differences in weighting for Panel 1 due to the phased introduction of the sample, the example uses Panel 2.

- ! Panel 2 will be selected in January 1996. Data will be collected from respondents in this panel to cover reference years 1996 - 2001.

- ! A six-year longitudinal weight will be calculated (based on the January 1 1996 population). This weight will be used for longitudinal analysis using data collected for Panel 2 respondents only.

- ! A three-year longitudinal weight will be calculated (based on the January 1 1996 population). This weight will be used for longitudinal analysis using data collected for both Panel 1 and Panel 2 respondents.

- ! A three-year longitudinal weight will be calculated (based on the January 1 1999 population). This weight will be used for longitudinal analysis using data collected for both Panel 2 and Panel 3 respondents.

Cross-sectional (Annual) Weights

Cross-sectional weights will be calculated for every year based on the composition of the target population on January 1 of the year after the reference year (for all intents and purposes, at the end of the reference year). All SLID respondents (longitudinal and cohabitants at any time during the year) will be assigned a non-zero cross-sectional weight for any given year.