Catalogue No. 94-16

MEASURING NON-RESPONSE IN A LONGITUDINAL SURVEY: THE EXPERIENCE OF THE SURVEY OF LABOUR AND INCOME DYNAMICS

September 1994

Sylvie Michaud, Social Survey Methods Division

Maryanne Webber, Household Surveys 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

This paper was presented at the Fifth International Workshop
on Household Survey Non-response held in Ottawa in September 1994.

Response rates and estimates of variance are key elements in evaluating the quality of a survey. The Survey of Labour and Income Dynamics is a new longitudinal survey being implemented in 1994. Special data quality measures are required in view of the fact that the survey is longitudinal and that data collection is done using computer-assisted interviewing. Moreover, the design has its own peculiarities, such as the "deferred" income interview and the use of dependent interviewing. This paper presents various data quality measures proposed for the survey, including wave and panel response rates, measures of the characteristics of non-respondents. A range of such measures is required to understand the various dimensions of the response mechanism.

TABLE OF CONTENTS

1. INTRODUCTION

The measurement of non-response error is usually divided in two parts: a non-response rate, and measures of the difference between respondents and non-respondents. A survey's non-response rate is relatively easy to measure and so it is often a prominent data quality measure. Differences between respondents and non-respondents are often more difficult to measure, because there is rarely much information available on non-respondents.

Statistics Canada has developed standards for the reporting of response rates, and such standards are useful. However, these standards were really designed with cross-sectional surveys in mind. As several new longitudinal social surveys are under development, it can be argued that the time is right to ask what additional measures are needed, to provide an accurate and consistent data quality picture for longitudinal surveys. Longitudinal surveys often have considerably more information on the characteristics of non-respondents lost after the first contact. This information should feed into non-response measurement.

The Survey of Labour and Income Dynamics (SLID) is one of these new longitudinal surveys. The paper will present the issues SLID is currently facing in trying to define response rates and other measures needed to convey an accurate and understandable picture of data quality to the users.

2. DESCRIPTION OF SLID

The following is a brief overview of the Survey of Labour and Income Dynamics, to situate the discussion on non-response measurement.

Design Features SLID is a household survey. The sample for the first panel was drawn from the Labour Force Survey (LFS). The LFS is a multistage probability area sample, covering the provinces, with the exception of Indian Reserves, the military and inmates of institutions. These selection criteria are all replicated in SLID (at the point of initial sample selection) with one exception: Armed Forces personnel living out of barracks are covered.

Each panel is retained for six years. In that time, thirteen interviews are done: a preliminary interview at the point of sample selection to collect baseline information, plus six labour interviews (every January) and six income interviews (every May). The labour and income interviews both refer to the previous calendar years.

The panels in SLID are overlapping. The first one was introduced in January 1993, when the preliminary interview was conducted. (The first wave of labour and income interviews for this panel were thus completed in 1994.) The second panel will start up in 1996, and at that point the survey will be up to full sample. In 1999. the third panel will start up and the first panel will be "retired".

The size of the first panel at the outset was 15,000 households. This includes about 31,000 persons aged 16 and over who will be asked labour and income questions. The size of the second panel has not yet been determined.

Following rules All household members in the dwellings originally selected for the survey will be followed through the six years, even if they move and form two or more separate households. These originally selected individuals are called *longitudinal respondents*. In addition, persons who join the household of a longitudinal respondent during the six years (called *cohabitants*) are also interviewed, as long as they continue to live with a longitudinal respondent.

Respondents are followed regardless of age. The labour and income interviews are completed for longitudinal respondents and cohabitants aged 16 and over as of January 1 of the survey year. Respondents who turn 16 become eligible for the labour and income interviews. Indeed, when they turn 15, they complete a preliminary interview in preparation for the following year's labour and income interviews.

The May interview is not a distinct wave, but is rather "deferred" from January because the income data are of better quality when respondents can consult their tax records. The fact that the January and May interviews are conceptually part of the same wave has implications for the following rules. For example, cohabitants who cease living with a longitudinal respondent between January and May are nevertheless traced and interviewed in May, to obtain the income data (which refer to the previous year). Only then are they dropped from the sample.

Operational Features SLID data are collected using computer-assisted interviewing (CAI). Data collection is decentralized: interviewers generally work out of their own homes, conducting interviews by telephone and transmitting data to the nearest Regional Office by modem. Proxy reporting is accepted, as long as the proxy respondent is knowledgeable. Otherwise, call-backs are arranged.

SLID is designed to take advantages of CAI's potential for improving data quality. First interactive editing of dates reported by the respondent is done to produce extensive spell data. CAI also facilitates dependent interviewing -- the feeding back of information reported in an earlier interview to improve recall and reduce seam problems and false transitions.

3. DATA QUALITY MEASURES

As mentioned earlier, the non-response error can be decomposed in two parts, the non-response rate and the difference between characteristics of respondents and non-respondents. Groves [1] notes that non-response can be further sub-divided into various categories, since behaviour may vary by kind of non-response (e.g. refusals, unable to trace, other non-response).

For longitudinal surveys, it is especially important to differentiate between type of non-response since the factors contributing to non-response may changes over the life of the panel. There is usually a higher non-response due to refusals at the beginning of the panel, but tend to decrease as the panel continues. On the other hand, there is usually no non-response in the first interview due to the inability to trace people, but it can become an important source of non-response in subsequent waves.

SLID has developed measures to monitor data quality. However, some of its design features create complexities when it comes to defining response rates. At the same time, the design leads to the interesting possibilities of the measurement of differences between respondents and non-respondents.

4. RESPONSE AND NON-RESPONSE RATES

The non-response rate is often a key measure of data quality. Statistics Canada developed standards to measure response rates [2]. Definitions must be adopted, before response rates can be calculated. However, then choice of the definitions can influence a lot the obtained response rate.

4.1 DEFINITIONS UNDERLYING RESPONSE RATES

Based on Statistics Canada standards, the definition of various units for purposes of calculating as follows (a brief description is provided in Appendix 1).

Total units	(1)	= (2) + (3)
Resolved units	(2)	= (4) + (12)
Unresolved units	(3)	
In-scope units (4)	= (5) + (8)	
Responding units	(5)	= (6) + (7)
Refusal conversion	(6)	
Other responding	(7)	
Non-responding units (8)	= (9) + (10) + (11)	
Refusals	(9)	
No contact	(10)	
Residual non-resp. (11)	= (16) + (17) + (18) (new)	
Out-of-scope units	(12)	= (13) + (14) + (15)
non-existent units	(13)	
Temp. out-of-scope	(14)	
Perm. out-of-scope (15)		

Residual non-response (11) should be further divided in three categories;

Unable to trace	(16)
Non received (transmission problems)	(17)
Other non-response	(18)

Two response rates are proposed: an operational response rate and a response rate for weighting purposes.

The operational response rate (OP RR) is defined as the number of respondents (that is completed interviews) over the number of people who were either resolved and eligible for the survey or unresolved. In the response rate for weighting (WGT RR), the number of unresolved units is divided into an expected number of eligible units and an expected number of non-eligible units. The expected number of non-eligible units is then removed from the calculations.

The response unit can be the dwelling or the person in the calculation of the rates, even if typically dwelling is used for a lot of social surveys.

4.2 SLID'S DESIGN AND ITS IMPACT ON DEFINING RESPONSE RATES

The first difficulty for SLID is due to the deferred interviews. The January and May interviews together form one "wave". One data file will be produced per year combining information from the two collection periods. When trying to classify eligible units into respondents and non-respondents, non-response can be defined legitimately in two ways. First, one can restrict respondents to cases where a response was obtained in both interviews. This would result in a substantial loss of

information. Alternatively, one can maximize the number of cases retained (and therefore the response rate) by including as respondent all persons who completed either the January or the May interview. This implies greater item non-response.

The second difficulty is that SLID covers all household members. The survey's cross-sectional weighting scheme is an integrated weighting procedure where the weight adjustment takes into account family composition, and ensures that everybody in the household gets the same weight. This raises the issue of what to do with households that are partially respondent. One can exclude them altogether, or include them by imputing for the missing household member. (Note that this imputation could simply be a matter of recording "unknown" to all items for the person in question).

An added complexity in the case of SLID is again the deferred interview: households are dynamic and the composition can change from January to May. Also, because the survey follows the longitudinal individuals, i.e. all the persons living in the original households, after the first year, the exact number of households is not known before collection.

The third difficulty is brought on by the implementation of computer assisted interviewing (CAI). CAI offers a lot of benefits, but the technology is still new at Statistics Canada and there is much to learn. As the following example illustrates, CAI has an impact on response and introduces new complexities in response measurement.

With the use of a decentralized CAI system, cases may not be successfully transmitted. These transmission failures appear to be divided in two categories and should be treated differently in trying to explain the response mechanism. The

first category consists of cases the interviewer never received. These cases labelled "not started" cases can probably be treated as missing at random.

The second category covers cases that were started by an interviewer but never made it back to Head Office. In contrast to the "not started" cases, these may generally be non-responding households lost in one of the transfer process. Some transfer processes are complex, involving various levels in the hierarchy of the data collection staff. For example, there is a refusal conversion process that involves a case transfer to a senior interviewer. Similarly the tracing procedure is done in two stage and involves the transfer of cases the interviewer is unable to trace to a Regional Office tracer. But since the cases were not received it is inherently difficult to know exactly what happened.

Another difficulty is the accurate assignment of response codes in CAI. To help the interviewer, certain response codes are assigned automatically, specifically where a complete interview is obtained and where the person in not eligible for an interview (for example, a person who has moved into an institution for more than six months). All the other response codes -- partial response, unable to trace and so on --have to be assigned manually by the interviewer. Because it is assigned automatically, the "complete interview" code is not on the menu of response codes that the interviewer can choose. Two problems arise. The interviewer may actually work right through an application, entering "refuse" or "don't know" whenever an entry is required. This is treated by the system as a complete. Also, the interviewer must apply judgement in selecting a response code. For example, the definition of "partial interview" is not sufficiently clear and a number of cases coded as "partial" have to be converted back to non-response, because hardly and information is present.

4.3 SLID'S PROPOSED RESPONSE RATES

Based on the design features for SLID, it is felt that more response rates are needed beyond the two proposed in the framework. So SLID is proposing to calculate three types of response rates: a phase response rate, a wave response rate and a longitudinal response rate.

4.3.1 Phase response rate

The phase response rate is essentially a cross-sectional rate. It will be used to indicate what happened at a particular collection instance. So in SLID's case, it will be calculated twice a year, once for January and once for May. This rate will be calculated at the household level. It may also be calculated at the person level, but for the subset of longitudinal persons only. This is because the number of non-responding cohabitants is underestimated, since the number of cohabitants may not be recorded for non-responding households.

Table 1 shows a distribution of households in the sample by their response code, for the January interview. From Table 1, it can be seen that the most important cause of non-response is cases not received through transmission. Using data from May, it was estimated that around 5% of cohabitants (mainly non-respondents) were missed in January because of this problem.

As can be seen from the definitions, partial responses have to be categorised into respondents or non-respondents. In SLID we plan to classify them as respondents, and impute key items of the missing person in the household. This definition means that response rate will be of 87%.

Table 1. Breakdown of the sample of households for the January interview

	# of households
Original households	15006
Created households	1024
Total expected # of households	16031
Non eligible	138
Responding (complete)	13113
Responding (partial)	708
Non response- refusals	276
Non-response- unable trace	334
Non-response- not received	1156
Non-response other	306

4.3.2 The wave response rate

The wave response rate gives a global picture of a particular year's of collection effort. The contents of the labour and the income interviews are merged to create one data file that will be released yearly. The wave response rate is thus a relevant measure for the data user community. This rate is calculated at both the household level and the person level. Each serves a different purpose.

Because of the deferred interview, the wave response rate can be calculated two ways. The numerator can be restricted to persons responding to both interviews (that is, labour and income). Alternatively, it can be defined as a response to *either* the labour interview or the income interview.

Table 2 shows the sample distribution of the eligible longitudinal respondents (only people aged 16 and more are eligible for the interview) and the different outcomes. Column 1 and 2 show the response codes based on what was received through the computer application. Column 3 shows the impact of recoding (done mainly when not enough data was present). Finally Column 4 shows what happens when the two interviews are combined. For persons not received through transmission in January but received in May, the January non-response code was imputed from the May non-response code.

Based on Column 4 of Table 2, the wave response rate could be 71%, if the first definition is used, 91% if the second definition is used. The second definition clearly maximizes the response rate but it also maximizes item non-response rate. The causes of non-response are also quite different between the labour and the income interview. The recoding of a non-response code increases the rates of unable to trace and of refusals. Finally the recoding of response code also increases substantially the number of refusals to the income interview.

For SLID, the second response rate will be used. Imputation procedures are being developed to impute key items for the missing phase. However, respondents will be categorised in two components, the respondents who completed both interviews, and the ones responding to only one. The wave response rate at the person level will show users the number of people who will be weighted on the file, for cross-sectional purposes.

Table 2. Response codes of the longitudinal sample of eligible persons, after the first year of interview

	Labour	Income	Income Recoded	Combined SLID

Complete	25792 (84.7 %)	24101 (79.1%)	23003 (75.6%)	21328 (70.0%)
Labour only	-	-	-	4464 (14.7%)
Income only	-	-	-	1675 (5.5%)
Refusal	598 (1.9%)	2246 (7.4%)	3694 (12.1%)	1194 (3.9%)
Unable to trace	312 (1.0%)	550 (1.8%)	546 (1.8%)	530 (1.7%)
Other	1071 (3.5%)	2354 (7.7%)	2153 (7.1%)	683 (2.2%)
Not received/ transmission	2435 (8.0%)	931 (3.1%)	786 (2.6%)	308 (1.0%)
Became not eligible	253 (0.8%)	279 (0.9%)	279 (0.9%)	279 (0.9%)
Total	30461 (100%)	30461 (100%)	30461 (100%)	30461 (100%)

The wave response rate at the household level is more complicated. Households are a dynamic structure, but to calculate a rate, they have to be defined in one point in time (in this case January).

Table 2 is a simplification of the table and the outcome codes. For example, new people become ineligible between January and May. These cases have to be recoded to non-response for the calculation of response rates, since the eligibility criteria are based on January. SLID also interviews in January people who are 15

(they complete a preliminary interview). At the person level, there should be a separate rate for the 15 years old.

4.3.3 Longitudinal response rates

A longitudinal response rate is calculated at the person level only. This is because of the difficulty in defining longitudinal households. This is the cumulative response rate, and serves as an indicator of loss through attrition. Essentially, it shows response in relation to the original sample.

For the longitudinal sample, the first contact for SLID was done one year before the first wave of labour and income interviews. This first contact is referred to as the preliminary interview. SLID had a 88% response rate to the preliminary interview for its first panel. The longitudinal response rate will show longitudinal respondents in each successive wave as a proportion of those who participated in the preliminary interview.

The longitudinal response rate could be calculated using either the phase response rate or the wave response rate. For SLID, the plan is to use the wave response rate, because it will represent the actual number of weighted records on the file. However, the longitudinal response rate will be divided into complete and partial, to distinguish between the records where a phase is imputed from the ones where no imputation is done.

5. DIFFERENCES BETWEEN RESPONDENTS AND NON-RESPONDENTS

The second aspect of response error concerns the differences that exist between respondents and non respondents. Longitudinal surveys are fortunate in that they often have considerable information about non-respondents who participated in some earlier wave. SLID is a supplement to the Labour Force Survey so even for those who did not respond to the preliminary interview, there are data available on demographic and labour market characteristics, and it is possible to discern differences between these individuals and those who did respond.

Three types of data quality studies will be undertaken using these data. These studies, described below, may be very useful in promoting our understanding of the sample composition and perhaps also of the attrition process. They could further help in selecting the most appropriate non-response adjustment procedure.

5.1 ANALYSIS OF CHARACTERISTICS OF NON-RESPONDENTS BASED ON PREVIOUS WAVES

A number of tables have been produced, summarizing characteristics of non-respondents to the preliminary interview and drawing comparisons to respondents. There were significant differences between respondents and non-respondents by province, urban/rural region, age groups, education, industry, occupation, job tenure, full-time vs part-time work, type of dwelling, number of adults in the dwelling and type of family. A second report of a similar nature is underway to study non-response to the first wave. The findings will be presented in a special data quality report, and will be used to verify the non-response adjustment proposed for the longitudinal sample.

5.2 STATISTICAL LINKING OF THE SAMPLE TO TAX DATA

A micromatch of the first panel sample to tax file data is currently underway, in order to measure data quality. The linkage will help us to evaluate some income characteristics of "attritors", particularly in terms of income level and sources in income received.

This linkage initiative is building on experience gained with data from a field test conducted for SLID in 1993. The test sample was also linked to the tax file. Tables 3 and 4 show selected test results. The statistical linkage was done using last name, first name, postal code, sex and date of birth. That information was available for most of the sample (a few non-respondents who could not be linked and they are excluded from the tables).

Table 3. Comparisons of wages and salaries for respondents and non-respondents in the Test 3 sample in Ontario.

Wages and Salaries	SLID non-respondents	Match SLID - tax		Unmatched SLID
Data source	TAX	TAX	SLID	SLID
# persons	650	848	848	188
% reported amount	71.4 %	72.5 %	70 %	57 %
average reported	$28k	$27k	$28k	$21k

Table 3 shows the reporting of wages and salaries. It can be seen that the unmatched sample is definitely different from the matched sample. However, differences between respondents and non-respondents are not that great either in terms of the number of people reporting wages and salaries or in the amounts reported. Table 4 gives a different picture.

For Table 4, one can see that for people who do respond, there is an underestimate both in the reporting and in the amount. Non-respondents are also different from respondents, with more non-respondents reporting Unemployment Insurance benefits on the tax file.

These results can be used to distinguish the component of underestimation that is due to total non-response from that attributable to item non-response. Micro-

comparisons will also be used to estimate response error, where definitions of the income sources on the survey side and on the tax side are compatible.

Table 4. Comparisons of Unemployment Insurance income.

Unemp. insurance benefits	SLID non-respondents	Match SLID - tax		Unmatched SLID
Data Source	TAX	TAX	SLID	SLID
# persons	650	848	848	188
% reported amount	18.6 %	15.7 %	12.6 %	13.3 %
average reported	$523	$593	$498	$132

5.3 FOLLOW-UP OF A SAMPLE OF NON-RESPONDENTS TO THE PRELIMINARY INTERVIEW

A small subsample of the 12% who did not respond to the preliminary interview in January 1993 was included in the first wave for purposes of data quality evaluation. The 1994 wave response rate for this small subsample was 66% (if response is defined as response to either the labour or the income interview). The data will be examined for potential differences with respondents.

6. CONCLUSIONS

A single response rate is not enough to give a good image of data quality in complex surveys. Other measures are needed. However there are disadvantages to calculating and presenting too many numbers, in that it makes it harder for data users, who usually want only one (or at most a few) response indicators. SLID will no doubt produce a lot of documentation on response, but will also have a few standard measures, namely a wave response rate and a longitudinal response rate.

SLID is electing for an broad definition of "responding household", and it follows that item non-response will be higher than it would be under a more restrictive definition. Thus, imputation to handle item non-response takes on more importance. SLID will try to adjust the variance calculations to reflect the use of imputed values in its estimation of variance; otherwise variance will be underestimated. Flags will also be put to indicate to users what has been imputed.

More work is needed on evaluation of the differences between respondents and non-respondents. No response indicators have yet been developed that could reflect these measurement errors.

CAI does improve data quality, but it can also enhance data quality "artificially". For example, in CAI, an incorrect flow (in terms of skip patterns) is not possible assuming the instrument has been thoroughly tested. However, the flow recorded may not reflect the reality. The correction of a path error may be more difficult in computer assisted interviewing than in paper and pencil and such errors may be harder to detect. We need to think about methods to evaluate response error. There is also a need to conduct more research aimed at understanding the general response mechanism and its different components.

Some of the measures discussed her relate to the design peculiarities of SLID. However, all longitudinal surveys seem to have their peculiarities and considerable

effort may be needed to develop measures that are appropriate. This makes it all the more important to devise standards for measurement that can accommodate the individuality of the various longitudinal surveys.

REFERENCES

[1] Groves,R., (1989), Survey errors and Survey Costs, Wiley Series

[2] Methods and Standards Committee, (1993), Standards and Guidelines for Reporting of Non-Response Rates.

APPENDIX 1. DEFINITION OF UNITS FOR PURPOSES OF CALCULATING RESPONSE RATES

Based on Statistics Canada standards on reporting non-response rates, the following definitions have been adopted:

Total units	(1)	all units in the sample
Resolved units	(2)	units where the status has been resolved at the end of the collection period
Unresolved units	(3)	units where the status has not been resolved at the end of the collection period
In-scope units	(4)	units in the target population of the survey
Responding units	(5)	units that provided usable information
Refusal conversion	(6)	refusals (from this collection or previous collection) that have been converted to respondents by a special conversion effort
Other responding	(7)	respondents other than refusal conversion
Non-responding units	(8)	in-scope that are non-responding or are not providing usable information
Refusals	(9)	units who refuse to participate to the survey
No contact	(10)	in-scope units that can not be contacted
Residual non-resp.	(11)	non-responding units because of special conditions
Out-of-scope units	(12)	units that do not belong to the target population.
Non-existent units	(13)	out-of-scope units that were determined to be non-existent.In the CAI world, it could include things such as duplicates due to transmission problems

Temp. out-of-scope	(14)	units out-of-scope at the time of one collection that subsequently may become eligible again (ex. people who move abroad for one year)
Perm. out-of-scope	(15)	units that became permanently out-of-scope since the sample was selected (deaths)