

Catalogue No. 97-04

**SELECTION OF A TOP UP SAMPLE
FOR CROSS-SECTIONAL INCOME ESTIMATES**

Product Registration Number 75F0002M

January 1997

Michel Latouche, Social Survey Methods Division

Sylvie Michaud, Social Survey Methods Division

Johanne Tremblay, Household Survey Methods Division

Ruth Dibbs, Household Surveys Division

The Income and Labour Dynamics Working Paper Series is intended to document detailed studies and important decisions for the Income and Labour Dynamics program. It is a continuation of the SLID Research Paper Series. These working papers are available in English and French. To obtain a summary description of available documents or to obtain a copy of any, please contact the Dissemination Unit, at 7-C6 Jean Talon Building, Statistics Canada, Ottawa, Ontario, CANADA K1A 0T6, by INTERNET (DYNAMICS@STATCAN.CA), by telephone (613) 951-4633, or by fax (613) 951-3012.

EXECUTIVE SUMMARY

The integration of the SLID and SCF income data collection is due to take place in fiscal 1997. We are currently planning to take most of the data from the SLID sample. In fact, two SLID panels will be active in the 1997 reference year. However, the first panel will be on its fifth wave of data collection and the second panel on its second. This paper sets out a number of summary analyses that support the recommendation to draw a top up sample. We first attempt to assess the impact of integration on the cross-sectional sample size, the variability of estimates and the representativeness of the sample. Next we study various options such as a dwelling or household sample, the appropriate time to draw the sample and earmark the LFS rotation group(s). *In the short term, in order to improve cross-sectional representativeness, we recommend adding a top up sample of 8,725 households which corresponds to an entire LFS rotation group. A dwelling sample would guarantee better sample cross-sectional representativeness but the associated costs would be much higher and would lead to a considerable decrease in the sample size. We therefore recommend that a household sample be drawn in August 1997. Rotation Group 4 in the LFS sample from April 1997 to September 1997 should be earmarked for the top up sample.*

The proposed solution is nonetheless a short-term one. In fact, lack of time and information have not enabled us to properly assess the full impact of the proposed strategy. A number of studies will be undertaken during the integration project in an attempt to quantify its impact as much as possible. There may have been a deterioration in the quality of the estimates if it turns out that there is some response bias caused by attrition and if the non-response adjustments do not compensate accordingly. The proposed strategy should be re-evaluated, therefore, within two years after integration.

TABLE OF CONTENTS		Page
1.	Introduction	1
2.	Integration and its impact on sample size and variability	2
3.	Integration and its impact on the cross-sectional representativeness of the sample	6
4.	Allocation of the top up sample	7
5.	Date of sample selection and dwelling vs household sampling units	9
6.	Selection of the sample (rotation group)	10
7.	Medium-term needs	12
Appendix 1.	Characteristics of persons by their interviewing experience: Full SIPP 1984 panel	13
Appendix 2:	Definition of the CMAs (Census Metropolitan Areas)	15
Appendix 3:	Preliminary longitudinal response rate of the SLID over three years	16
Appendix 4:	Drawing the top up sample assuming the same content	21
Appendix 5.	Options for a 1998 top-up sample	23
Appendix 6:	Sample selection	27

1. INTRODUCTION

The integration of the SLID and SCF income data collection is due to take place in fiscal 1997. We are currently planning to take most of the data from the SLID sample. In fact, two SLID panels will be active in the 1997 reference year.

However, the first panel will be on its fifth wave of data collection and the second panel on its second. This paper sets out a number of summary analyses that support the recommendation to draw a top up sample. We first attempt to assess the impact of integration on the cross-sectional sample size, the variability of estimates and the representativeness of the sample. Next we study various options such as a dwelling or household sample, the appropriate time to draw the sample and earmark the LFS rotation group(s). *In the short term, in order to improve cross-sectional representativeness, we recommend adding a top up sample of 8,725 households which corresponds to an entire LFS rotation group. A dwelling sample would guarantee better sample cross-sectional representativeness but the associated costs would be much higher and would lead to a considerable decrease in the sample size. We therefore recommend that a household sample be drawn in August 1997. Rotation Group 4 in the LFS sample from April 1997 to September 1997 should be earmarked for the top up sample.*

The proposed solution is nonetheless a short-term one. In fact, lack of time and information have not enabled us to properly assess the full impact of the proposed strategy. A number of studies will be undertaken during the integration project in an attempt to quantify its impact as much as possible. There may have been a deterioration in the quality of the estimates if it turns out that there is some response bias caused by attrition and if the non-response adjustments do not compensate accordingly. The proposed strategy should be re-evaluated, therefore, within two years after integration.

The paper refers throughout to reference years. The 1993 reference year deals with income for the 1993 calendar year. Collection of these income data is carried out in May of the following year. Section 2 discusses the impact of integration on the effective size of the cross-sectional sample and the variability of estimates. Section 3 covers the impact of integration on the representativeness of the cross-sectional sample. Section 4 puts forward a number of scenarios for allocating a top up sample. Section 5 compares a dwelling and a household sample in terms of representativeness and cost. Section 6 addresses the choice of a rotation group from the LFS sample to make up the top up sample.

2. INTEGRATION AND ITS IMPACT ON SAMPLE SIZE AND VARIABILITY

Integrating the two surveys will have an impact on estimates of income that will be different from the effect on estimate levels and/or the variables studied. The design of both surveys, one by nature purely cross-sectional and the other mainly for longitudinal use, implies there will indeed be impacts. A comparative study of the first wave of SLID data was carried out, but the lack of longitudinal data and the absence of the second panel limits comparisons. We have assumed that the integrated survey should maintain the current SCF levels of precision.

The first part of the study thus concentrates on the comparison of sample sizes in the two surveys. The first SLID panel was introduced in the 1993 reference year. In the first wave of interviews, around 15,000 households were interviewed. Since the SLID follows individuals for six years, a sample attrition (i.e., a cumulative non-response) was observed because it was impossible, for example, to trace a percentage of the sample of people who move. This is partly balanced by the fact that the sample is dynamic. Because the make-up of a household is dynamic over

time, a certain number of persons originally living together will leave to establish new households. New individuals also move into existing households. Every year, individuals who live with longitudinal respondents and who are part of these new households are interviewed. Table 1A describes the number of households interviewed for the first SLID panel. A noticeable decrease in the number of respondents can be seen at the beginning of the panel, followed by stagnation caused by the balancing of nonresponse and household creation. Other analyses (not presented in this paper) have shown that a similar phenomenon occurs at the provincial level.

Table 1A. Number of response households in the first panel

	Preliminary*	wave 1	wave 2	wave 3	wave 4**	wave 5**	wave 6**
Response households	15,006	14,694	15,224	15,207	15,434	15,445	15,465
% of the preliminary sample	88	86	91	92	91	91	91

* In fact, a sample of 20,000 households had been drawn, and more than 17,000 had responded but budget cutbacks limited the sample size to 15,000 households.

** These figures are projected.

A second SLID panel will be rotated in for the 1996 reference year. This new panel is made up of two rotation groups from the redesigned Labour Force Survey (LFS) sample. For the 1997 reference year, the second panel should be in the second year of data collection and the first panel in its fifth year. We can try to predict the number of households that will be eligible for the 1997 reference year by using the rates of increase/decrease in the first panel.

Table 1B. Projection of the number of households in the second panel

	Prel.+ wave 1	wave 2	wave 3	wave 4	wave 5	wave 6
Response households	14,229	15,056	15,221	15221	15056	15056
% of the preliminary sample*	86	91	92	91	91	91

* rates of first panel

For the first year of integration, the size of Panel 1 is estimated to be 15,445 respondents and Panel 2, 15,056 for a total of 30,501 households. These estimates are calculated on the basis of the response rate that includes households for which data on income were imputed. About 12% of the data on income were totally imputed. Thus, to make the response rates comparable to the SCF, the SLID rate must be decreased. We could thus expect to have a sample of about 27,000 response households.

The SCF draws its sample from among LFS respondents. Four rotation groups were drawn for the SCF which thus corresponds to a sample of 34,900 households. Traditionally, the SFC has a response rate of around 80% which would correspond to a sample of about 27,920 respondents.

The SLID sample seems to be lightly smaller, therefore, than that of the SCF. If we use only the sample size for purposes of comparison, we might infer that the SLID sample needs to be increased by about 1000 households.

Table 2 sets out the coefficients of variation for estimates of total personal income before taxes and for the number of individuals 15 years of age and over with a family income before taxes under the LICO for the SLID and the SCF. The CVs

refer to the 1993 reference year. For Canada, SLID estimates remain valid for personal income considering that the SLID uses only two rotation groups. For the number of LICO individuals, the differences for Canada are more significant with the SLID CV being almost twice as high as the SCF. At the provincial level, the differences vary widely depending on the province and the interest variable. In some cases, the SLID CVs are very close to the SCF even with a smaller sample. In other cases, the differences are greater even if we consider that the SCF's sample size is double that of the SLID.

Table 2. Comparison of SLID and SCF CVs for personal income before taxes and for the number of individuals 15 years of age and over with a family income before taxes under the LICO.

PROVINCE	Personal income before taxes		Number of individuals under the LICO	
	SLID (2 "rotations")	SCF (4 "rotations")	SLID (2 "rotations")	SCF (4 "rotations")
NF	7.2	1.9	16.4	6.3
PE	1.6	2.3	15.6	12.2
NS	3.7	3.2	12.2	6.9
NB	7.2	1.4	10.0	5.6
QC	2.4	1.2	6.4	3.9
ON	1.4	1.1	8.4	3.5
MB	3.0	1.9	6.2	6.3
SK	2.3	1.5	6.8	5.2
AB	2.0	2.4	6.9	5.2
BC	1.3	2.1	7.6	6.0
Canada	0.9	0.7	3.5	1.9

3. INTEGRATION AND ITS IMPACT ON THE CROSS-SECTIONAL REPRESENTATIVENESS OF THE SAMPLE

Even if we could obtain a sample of similar size without selecting a top up sample, it is reasonable to ask whether such a sample would have cross-sectional representativeness. In fact, the experience of other longitudinal surveys suggests that there are differences between respondents and non-respondents. Although weighting procedures attempt to make up for these differences, it is always difficult to assess the ultimate impact. Moreover, since the SLID combines income and labour variables at the same time, it is not clear that adjustment will be equally effective for both components.

The assumption of differences between non-respondents because of attrition and respondents is supported by the literature. Appendix 1 compares the American “Survey of Income Program Participation” (SIPP) longitudinal sample, by breaking up the sample into various components: those who responded to all interviews, attrition, missing only one interview, and missing two or more interviews. When we compare the distribution, we can see that as in Canada, attrition is greater in some areas (for example, Los Angeles), in large cities and for people who lived in rented accommodation during the first wave. These are similar to the results the SLID observed in its first wave. Table 3 compares the estimates of size-1 families to size-2+ families during the first wave of the SLID and the SCF (some demographic characteristics were compared but this particular feature seemed significant).

Table 3. Distribution of the number of persons by family sizes 1 and 2+ for the SLID and the SCF (after post-stratification, integrated household weighting in both cases)

	SCF %	SLID %	Difference of total estimates * (SLID-SCF)/SCF
Size-1 families	13.4	12.9	- 4.0 %
Size-2+ families	86.6	87.1	+0.4 %

*This difference is obtained from estimates of the number of persons. It cannot be produced, therefore, directly from this table.

The SLID seems to understate size-1 families and overstate size-2+ families. This may probably be explained by the fact that the SLID sample must trace people who move. We thus lose some individuals who leave their original household and create single-person households, hence size-1 families. A good breakdown between these kinds of families is very important for SCF estimates. It may also be that non-traced people who move have different characteristics from those who did not move. A supplementary sample would partly offset the representativeness bias that could occur in the sample because of attrition. Thus, although the net loss in terms of sample size is small, it is nevertheless recommended that a top up sample be drawn to enable the possible bias caused by the longitudinal sample to be offset.

4. ALLOCATION OF THE TOP UP SAMPLE

The tight time frames leave little room to manoeuvre into deciding how to allocate a top up sample. Various hypotheses have been studied: selecting a whole LFS rotation group; selecting a CMA (Census Metropolitan Area) subgroup in one or

two LFS rotation groups; and various combinations of a complete group with a CMA subgroup. The various scenarios are set out in Table 4.

Table 4. Different scenarios for allocating the top up sample (in terms of households)

	0 complete rotation group	1 complete rotation group
No extra CMA		8725
1 CMA-25 group	4150	12875
1 CMA-16 group	3158	11883
1 CMA-9 group	2258	10983
1 CMA-3 group	1112	9837
2 CMA-25 groups	8299	17024
2 CMA-16 groups	6317	15042
2 CMA-9 groups	4515	13240
2 CMA-3 groups	2224	10949

The CMA-16 group corresponds to the CMAs for which the SCF provides estimates at the CMA level, the CMA-9s represent the CMAs for which the LFS has incorporated stratification by income and the CMA-3 groups refer to the three largest cities. Appendix 2 lists the CMAs that belong to each group.

The study is based on the assumption that we want to ensure the quality of estimates of Canadian personal and family income as well as the level of the number of people below the poverty line. Even though we want national estimates, there will probably be a minimum need for provincial estimates. By concentrating the sample on CMAs, national estimates will be improved through a larger sample in the large CMAs. Moreover, American experience suggests that attrition is more significant in large cities. On the other hand, when the

preliminary 3-year SLID response rates are analysed (set out in Table 3), the CMAs' response rate is no lower than the CAs. Likewise, when we compare the CVs in Table 2, it seems necessary to cover all the provinces if we want to guarantee good provincial estimates.

The ideal scenario would be one in which the top up sample would enable us to draw a complete rotation group plus a second rotation group that would only cover the CMA-9s. This scenario, however, produces a larger sample size than the one originally contemplated (10,983 households instead of 9,400). To give due regard to sample-size limitations, we need to choose between a complete rotation group (8,725) or two rotation groups from the CMA-25s only (8,299). Since the 3-year SLID response rates are no lower than the CMAs, we recommend using a complete rotation group for the first year. In carrying out research on attrition over the next few years, we will assess whether it would be preferable to concentrate the sample in the CMAs.

5. DATE OF SAMPLE SELECTION AND DWELLING VS HOUSEHOLD SAMPLING UNITS

Comparisons were made to attempt to determine the optimum time for selecting the top up sample. Various scenarios were considered. First, the top up sample could be a household sample drawn during the year preceding the collection, January 1997 at the earliest, like rotating in a new SLID panel in August 1997 at the latest, to enable the application of the January 1998 collection to be finalized. The top up sample could also be a dwelling sample drawn at the time of collection, as the SCF currently does.

The table in Appendix 4 describes in detail the advantages and disadvantages of these various options. The main disadvantage of the household sample drawn in 1997 is the need for tracing as in the original sample. The number of individuals to trace is much lower for the sample drawn in August than for the one in January. On the other hand, the sample drawn in January enables us to feed back labour data so as to reduce follow-up error and improve the quality of labour data. The dwelling sample has the main advantage of ensuring better cross-sectional representativeness. This feature makes it the best approach from the methodological point of view since the top up sample has precisely the same aim. Nonetheless, a study of the costs of this approach (see Appendix 5) indicates that the sample should be reduced by about one-third to balance the additional costs of face-to-face interviews and the need to adjust the application of the current SLID collection. It is thus suggested that a household sample be drawn for the first year of integration with the possibility of changing this approach after studying the representativeness of the SLID panels more closely, taking into account the fact that the application of the SLID collection will have to be changed in the next few years in any case.

6. SELECTION OF THE SAMPLE (ROTATION GROUP)

Although the top up sample will be a household sample made up of a rotation group drawn between January and August 1997, it is still necessary to decide which LFS rotation group will match the SLID top up sample. In addition to considering the tracing problem already mentioned in the previous section, the response burden, the overlap between LFS and SLID interviews, the content overlap and the cluster effect between some rotation groups are the determining factors for the sampling. The table in Appendix 6 sets out the various possible

samples, numbers 1 to 6 representing the six LFS rotation groups and the colour changes indicating a sample rotation for this group.

Groups 2 and 3 have been discarded as possible top up sample choices because of the cluster effect. This problem is generated by the approach used by the LFS to replace the sample. When a rotation group finishes its six months in the LFS, it is replaced by another sample, generally from the same cluster as the outgoing group. We estimate that about 90% of all rotation groups replaced in a year are replaced by samples from the same cluster. There is no cluster-effect problem for the LFS since there is only one group from the cluster in the sample at a time. For the cross-sectional sample of the SLID that will be made up of two panels and the top up sample, it would be preferable if the samples did not come from the same clusters. Since the first SLID panel comes from the old LFS design, it is unlikely that the same clusters would be considered. Nonetheless, the second panel is comprised of a rotation group from the August 1995 to January 1996 LFS sample (Group 2) and the September 1995 to February 1996 sample (Group 3). It is thus preferable to exclude all samples from rotation groups 2 and 3 to avoid concentrating the top up sample in the same clusters as the second panel.

Some groups have also been set aside because the response burden was too great and might have affected the response rate. These were the rotation groups already called upon by the SFC for 1997 income and some groups taking part in a number of LFS top up surveys (see Appendix 6). Since the top up sample will be a household sample that must be drawn by August 1997 at the latest, the recommended choice is rotation group 4 that will be in the LFS from April 1997 to September 1997.

7. MEDIUM-TERM NEEDS

In the medium term, an in-depth evaluation of SLID attrition will be carried out as the subsequent panels are dealt with. A priori, it is unlikely that we would ever want to drop the option of a top up sample to maintain the quality of cross-sectional observations, if we are trying to use these data as sources of official cross-sectional statistics. Cross-sectionally, the panel will always suffer from a representativeness problem. First of all, there are the understated “immigrant” households; but we may doubt the ability of a top up sample to properly target such a small group. More significant, however, is the problem of attrition. We note that attrition is often greater in some groups. Youth, a group not examined in this paper, is understated in taxation data matching and has been observed to have greater attrition. The best way to ensure cross-sectional representativeness may eventually be to slightly reduce panel sizes (for example, by sub-sampling of salaried workers in certain age groups) to increase the top up sample size and capture certain understated population segments. Time and the lack of longitudinal data did not allow us to make appropriate inferences about the problem for the time being. In particular, we were not able to quantify the impact of over-sampling in some groups. The recommendations are thus to draw a single short-term top up sample and re-evaluate after the first year of integration. Better comparative research on the representativeness of the panels as opposed to “updates” could be carried out. From this, we could quantify the impacts when the data from the first top up sample become available.

**APPENDIX 1. CHARACTERISTICS OF PERSONS BY THEIR
INTERVIEWING EXPERIENCE:
FULL SIPP 1984 PANEL**

	Restricted sample (1)	Completed all Interviews (2)	Attrition (3)	Missing 1 Interview (4)	Missing 2+ Interviews (5)
Total Number	32985	22814	6957	2330	884

Wave 1 Variables

Regional Office:

Boston	7.0	7.2	7.4	5.1	3.7
New York	7.1	5.8	9.4	10.5	12.9
Philadelphia	10.3	11.0	7.6	9.8	13.6
Detroit	8.5	8.6	8.3	7.1	8.9
Chicago	7.6	8.7	5.4	5.3	4.1
Kansas City	8.4	9.5	6.4	5.1	3.4
Seattle	8.6	9.1	6.9	8.6	8.9
Charlotte	9.0	9.4	8.7	8.0	4.9
Atlanta	11.4	10.6	12.5	15.5	11.8
Dalla	9.5	8.5	11.4	12.2	12.2
Denver	5.8	5.7	6.3	5.2	4.6
Los Angeles	6.9	5.9	9.7	7.6	11.0

	Restricted sample (1)	Completed all Interviews (2)	Attrition (3)	Missing 1 Interview (4)	Missing 2+ Interviews (5)
Residential char:					
Not an SMSA	25.2	27.4	20.1	22.4	17.2
SMSA: LT 100000	1.2	1.2	1.3	1.0	0.6
SMSA: 100-249 thou	9.1	9.6	8.1	8.4	6.7
SMSA: 250-499 thou	9.3	9.7	8.9	7.6	8.5
SMSA: 500-999 thou	13.4	13.2	14.6	13.6	11.1
SMSA: 1-2.9 mill.	24.2	23.2	26.5	25.7	29.4
SMSA: 3-14.9 mill.	17.4	15.7	20.5	21.4	26.6
Living quarters:					
House apt.flat	94.2	94.1	94.4	94.2	94.2
All other	5.8	5.9	5.6	5.8	5.8
Living quarters:					
Owned/Being bought	69.8	72.7	62.9	65.6	59.8
Rented for cash	27.8	24.8	35.1	32.1	37.3
Occ'd w/o cash	2.3	2.4	2.0	2.3	2.8
Race:					
White	86.8	88.5	83.8	82.9	78.2
Black	10.5	9.2	12.9	13.9	17.6
AmInd/Esk/alNative	0.4	0.3	0.7	0.5	0.9
Asian/Pac.Isl	2.2	2.0	2.6	2.7	3.3

APPENDIX 2: DEFINITION OF THE CMAs (CENSUS METROPOLITAN AREAS)

	3 CMA	9 CMA	16 CMA	25 CMA
St John's			X	X
Halifax			X	X
Saint John			X	X
Quebec			X	X
Trois-Rivières				X
Sherbrooke				X
Montreal	X	X	X	X
Hull-Ottawa		X	X	X
Chicoutimi				X
Oshawa				X
Toronto	X	X	X	X
Hamilton		X	X	X
St. Cath/Niagara			X	X
London		X	X	X
Windsor				X
Kitchener			X	X
Sudbury				X
Thunder Bay				X
Winnipeg		X	X	X
Regina				X
Saskatoon				X
Calgary		X	X	X
Edmonton		X	X	X
Vancouver	X	X	X	
Victoria			X	X

**APPENDIX 3: PRELIMINARY LONGITUDINAL RESPONSE RATE
OF THE SLID OVER THREE YEARS**

CMA/CA	Number of respondents	Response rate
<i>Missing</i>	494	0.83
ST JOHNS	241	0.84
010	46	0.88
011	8	0.53
015	93	0.86
025	20	0.87
105	244	0.89
110	64	0.72
HALIFAX	282	0.80
210	69	0.85
215	147	0.77
220	88	0.78
225	140	0.77
305	267	0.80
SAINT JOHN	196	0.78
320	208	0.87
328	37	0.65
330	20	0.95
335	67	0.92
403	20	0.91
404	66	0.77
406	65	0.73
CHICOUTIMI- JONQUIÈRE	282	0.87
410	9	0.82
411	20	0.77
412	48	0.76
QUEBEC	228	0.75
SHERBROOKE	391	0.80
435	28	0.82

CMA/CA	Number of respondents	Response rate
440	24	0.96
TROIS-RIVIÈRES	163	0.87
444	78	0.88
446	32	0.91
447	39	0.57
450	82	0.68
452	35	0.76
454	35	0.76
456	14	0.74
459	94	0.85
MONTREAL	625	0.77
465	45	0.94
468	11	0.85
475	61	0.77
480	48	0.77
485	81	0.80
501	106	0.74
502	2	0.67
HULL-OTTAWA	526	0.72
515	25	0.81
521	91	0.73
522	109	0.83
529	107	0.66
530	7	0.78
OSHAWA	269	0.79
TORONTO	699	0.76
HAMILTON	301	0.78
ST CATHARINES/ NIAGARA	426	0.80
KITCHENER	417	0.83
543	114	0.83
544	55	0.86
546	8	1.00
547	12	0.75

CMA/CA	Number of respondents	Response rate
550	189	0.87
553	62	0.61
LONDON	322	0.78
556	13	0.76
557	51	0.94
WINDSOR	273	0.78
562	42	0.71
566	17	0.74
567	14	0.82
568	50	0.79
569	39	0.81
571	28	0.90
575	122	0.74
SUDBURY	252	0.84
582	52	0.84
584	14	0.93
585	7	0.70
586	17	0.71
590	58	0.81
THUNDER BAY	265	0.83
598	7	1.00
WINNIPEG	468	0.76
604	6	0.35
607	53	0.82
610	61	0.77
640	76	0.97
REGINA	336	0.86
710	41	0.87
715	74	0.76
720	63	0.86
SASKATOON	309	0.82
730	22	0.73
735	30	0.86
745	97	0.71

CMA/CA	Number of respondents	Response rate
750	34	0.67
805	56	0.86
810	74	0.83
CALGARY	473	0.84
830	72	0.81
833	14	0.78
EDMONTON	550	0.79
840	36	0.77
845	1	1.00
850	44	0.83
860	37	0.80
865	19	0.86
905	42	0.71
913	53	0.83
915	64	0.76
918	25	0.81
925	86	0.77
930	124	0.79
932	137	0.76
VANCOUVER	407	0.74
VICTORIA	202	0.81
937	2	1.00
938	71	0.68
940	22	0.63
943	29	0.93
944	18	0.78
945	15	0.83
950	26	0.72
952	6	1.00
955	21	0.72
960	33	0.94
965	26	0.72
970	75	0.89
975	24	0.83

CMA/CA	Number of respondents	Response rate
977	18	0.78
<i>Unknown</i>	232	0.56
OUTSIDE CMA/CA	9659	0.84
TOTAL	24986	

APPENDIX 4: DRAWING THE TOP UP SAMPLE ASSUMING THE SAME CONTENT

	HOUSEHOLDS SAMPLED IN JANUARY 1997	HOUSEHOLDS SAMPLED IN AUGUST 1997	DWELLING SAMPLED AT THE TIME OF THE COLLECTION
REPRESENTATIVENESS OF THE SAMPLE	<p>(-) Individuals who move could be under-stated if measures taken to trace them are not effective.</p> <p>(-) Immigrants for 1997 cannot be included in the sample. The top up sample therefore does not enable us to compensate for the immigrant under-coverage.</p>	<p>(+/-) The under-statement of movers is less significant since almost 2/3 of yearly moves have already taken place by August according to the 1994 SLID results.</p> <p>(+/-) Immigrants who enter the country before August 1997 could be included in the sample.</p>	<p>(+) Good representativeness of movers by identifying the new arrivals in the selected dwelling.</p> <p>(+) The top up sample will cover 1997 immigrants.</p>
TRACING	<p>(-) Requires more tracing and hence a higher collection cost and a lower response rate.</p>	<p>(+/-) Requires less tracing since a good proportion of moves have already taken place.</p>	<p>(+) No tracing is necessary because the household in the dwelling is selected when the sample is drawn.</p>
FOLLOW-UP ERROR	<p>(+) Reduces follow-up error for January 1997 because January LFS information is available for both groups.</p>	<p>(-) No follow-up is possible because using the SLID does not allow us to take advantage of the April LFS information available for both groups.</p>	<p>(-) No follow-up is possible.</p>

RESPONSE BURDEN	<p>(+) Period of 12 (or 11) months between the SLID contact and the last LFS contact.</p>	<p>(-) Period of only 4 (or 3) months between the last LFS contact and the SLID contact. An incentive letter would be received not long after the LFS.</p> <p>(-) One of the two rotation groups would have already taken part in the 1997 SCF if the two outgoing rotation groups are taken.</p>	<p>(-) Respondents were called on for the LFS and the top up SLID in January.</p>
COMPOSITION OF THE FAMILY	<p>(+) The composition of the family is defined at the start of the reference period. Change in the composition of the family may thus be identified during the reference year.</p>	<p>(+/-) The composition of the family is unknown for the first months of the reference year. Some changes in the composition of the family thus cannot be identified.</p>	<p>(-) Changes in the composition of the family cannot be identified.</p>
ELIGIBILITY OF MEMBERS	<p>(+) Eligibility could be defined the same way for both panels.</p>	<p>(-) More difficult to define eligible members.</p>	<p>(-) More difficult to define eligible members.</p>
OTHERS	<p>(+) The approach is more consistent with the other SLID panels since the samples are drawn in January.</p>		<p>(-) The SLID contact procedure would have to be modified.</p>

APPENDIX 5. OPTIONS FOR A 1998 TOP-UP SAMPLE

This note summarizes the collection implications and cost estimates of several options for a top-up sample starting in 1998. The basic comparison is between a dwelling sample targeted to CMAs compared, to a sample of 9,000 households using SLID's current collection method. The dwelling sample implies personal visits and a rewriting of the SLID contact and demographic module by January 1998.

All household surveys must be rewritten to function under Windows, and to deal with year 2000 issues. The target date is to be in the field in April 1999. All surveys will be expected to follow standards interfaces with case management and have a Windows "questionnaire". Development of standards will take place in 1997/1998 and would not be ready for us if we rewrite for January 1998 collection (starting April 1997 at latest). We will have to redesign the system again for January 2000 collection.

Thus implementing a dwelling sample implies programming and testing costs for a collection application that can be used only for two years.

Figure 1 shows assumptions and operational scenarios for the two approaches. Table 1 presents cost and sample size estimates for three dwelling-based samples, 9, 16 and 25 CMAs, and two household-based approaches, one complete rotation or two rotations targeted to CMAs or CAs. Finally, advantages and disadvantages of the approaches are compared in Figure 2.

Figure 1 Assumptions for Collection

Dwelling sample	Household Sample
Vacant dwellings from month of selection need personal visit	Not in sample/not applicable
Vacant dwellings in January 1998 need personal visit	(Probably SLID “unable to trace” cases)
Movers in January 1998 not traced; dwellings need personal visit	Movers sent to telephone tracing
Movers from January to May need tracing—implies collection approach is different for each phase	Movers from January to May need tracing—collection approach is consistent for each phase
Major rewrite of contact module for January 1998 interview	Not needed
Because of the complexity of SLID application, probably requires separate applications for top-up and longitudinal cases in January	Not needed
Major rewrite of contact module for January 2000 interview (all household surveys move to Windows)	Major rewrite of contact module for May 1999 interview (all household surveys move to Windows)

Table 1 Sample size and costs

	Dwelling sample			Household sample	
	9 CMAs	16 CMAs	25 CMAs	1 rotation	targeted
Adjusted sample size ¹	4 924	6 889	9 050	9 000	9 000
Personal interviews ²	1 192	1 653	2 172	0	0
Jan. collection cost	\$102 676	\$139 664	\$180 347	\$126 368	\$126 368
May collection cost	16 063	22 475	29 526	30 537	30 537
Programming (1 year only) ³	19 000	19 000	19 000	n/a	n/a
Contact question as LFS note ⁴	n/a	n/a	n/a	1 500	3 000
Total Costs	\$137 739	\$ 181 139	\$228 873	\$158 405	\$159 905

1. Dwelling household counts given by Methodology, adjusted by vacant dwellings in selection month.

2. Estimates are conservative and based on 8% vacants both in 1997 and 1998 (Methodology) and 8% whole Hhld moves (SLID Wave 1). Cost is about \$30 per case compared to \$14 per telephone interview.

3. Assumption of 5 months CS2 work.

4. Recommend asking two rotation groups for the name of a contact person in 1997 LFS. Methodology could recommend either the targeted or 1-rotation household sample option with more information on attrition. By March, we should know raw response rates for Panel 2, Wave 1 (the great unknown). Impact of collection preparation work should be approximately the same for both options.

Figure 2 Advantages and disadvantages of the options

	Dwelling sample	Household Sample
Advantages	Best design to reduce attrition bias	Cheaper for equivalent sample size
	Possibly higher rate of permission to use tax files	Simpler procedures—maintains SLID's basic collection design
		No wasted effort rewriting contact program for use for only two years
		Minimal changes to interviewer training and manual
Disadvantages		
	Adds major complexity to already complex system	
	Sends contradictory signals to interviewers- tracing important for some but not others	
	Requires changes to Interviewers' manual; also either more training time or risk of less knowledge of SLID longitudinal procedures	
	Requires trade-off between cost and sample size— impact on CVs	Not as good for offsetting attrition bias
	Need for SLID staff to do specs for 1997/98 rewrite of Contact diverts their ability to analyse problems and optimize changes for mandatory rewrite in 1998/99	
	Requires more additions and changes to processing system.	Requires some, but fewer, changes to processing system
Unknown	Impact on response rates	Impact on response rates

APPENDIX 6: SAMPLE SELECTION

		J97	F	M	A	M	J	J	A	S	O	N	D	J98	Main reason to exclude by sample		
		1	1	1	1	1	1	1	1	1	1	1	1	1	SCF April 97	end LFS Dec 97	New LFS
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	Cluster effect	Cluster effect
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	Cluster effect	Cluster effect
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	Cluster effect	Cluster effect
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	Cluster effect	Cluster effect
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	Cluster effect	Cluster effect
																overlap w. LFS	overlap w. LFS
																overlap w. LFS	overlap w. LFS
																overlap w. LFS	overlap w. LFS

Option 1: Groups 4 and/or 5 drawn in March 1997

8 to 9 months between the end of the LFS and the start of the SLID
 3 to 4 months labour content overlap with the LFS
 drawn 10 months before the contact (tracing movers)

Extra burden:

- Travel Survey (November or December 1997, by group 4 ou 5)

one person per household

- Survey on the Importance of Wildlife (February 1997)

short interview of less than 5 minutes

- Longitudinal Survey of Children (February or March 1997)

for households with children between 0 and 23 months of age (about 6% of households)

Option 2: Groups 4 and/or 5 drawn in August 1997

3 to 4 months between the end of the LFS and the start of the SLID

6 months labour content overlap with the LFS

drawn 5 months before contact (tracing moves)

Extra burden: still not known