



Catalogue no. 75F0002MIE — No. 005

ISSN: 1707-2840

ISBN: 0-662-43130-8

Research Paper

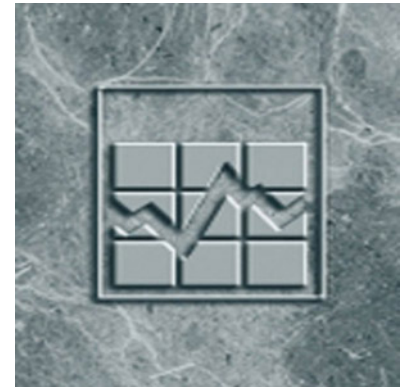
Income Research Paper Series

Data Quality in the 2004 Survey of Labour and Income Dynamics (SLID)

by Chris Duddek and Susan Storey

Income Statistics Division and Social Survey Methods Division
Jean Talon Building, Ottawa, ON K1A 0T6

Telephone: 613 951-7355



Statistics
Canada

Statistique
Canada

Canada

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Income Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: (613) 951-7355; (888) 297-7355; income@statcan.ca).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 75F0002MIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.



Statistics Canada
Income Statistics Division

Income Research Paper Series

Data Quality in the 2004 Survey of Labour and Income Dynamics (SLID)

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2006

All rights reserved. The content of this publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it is done solely for the purposes of private study, research, criticism, review, newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, name of product, catalogue, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopy, for any purposes, without the prior written permission of Licensing Services, Client Services Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

April 2006

Catalogue no. 75F0002MIE, Vol. 5

Frequency: Occasional

ISSN: 1707-2840

ISBN: 0-662-43130-8

Ottawa

La version française de cette publication est disponible sur demande (n° 75F0002MIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Abstract

The Survey of Labour and Income Dynamics (SLID) is a longitudinal survey initiated in 1993. The survey was designed to measure changes in the economic well-being of Canadians as well as the factors affecting these changes.

Sample surveys are subject to errors. As with all surveys conducted at Statistics Canada, considerable time and effort is taken to control such errors at every stage of the Survey of Labour and Income Dynamics. Nonetheless errors do occur. It is the policy at Statistics Canada to furnish users with measures of data quality so that the user is able to interpret the data properly. This report summarizes a set of quality measures that has been produced in an attempt to describe the overall quality of SLID data. Among the measures included in the report are sample composition and attrition rates, sampling errors, coverage errors in the form of slippage rates, response rates, tax permission and tax linkage rates, and imputation rates.

Table of contents

1. Introduction.....	6
2. Sample composition/Attrition.....	7
3. Sampling errors.....	9
4. Coverage errors.....	10
5. Response rates.....	11
6. Tax permission rates.....	15
7. Tax linkage rates.....	16
8. Imputation rates.....	17
9. Rounding of income data.....	21
Appendix 1: Sample composition in SLID by province, 1996 - 2003.....	23
Appendix 2: Slippage rates over time.....	26

1. Introduction

The Survey of Labour and Income Dynamics (SLID) is a longitudinal survey initiated in 1993. The survey was designed to measure changes in the economic well-being of Canadians as well as the factors affecting these changes. The target population consists of all persons living in Canada with the following exclusions: persons living in Yukon, the Northwest Territories, and Nunavut, persons living on Reserves, persons living in institutions, and military personnel living in barracks.

The SLID sample is comprised of 2 panels. Each panel remains in the survey for 6 consecutive years and a new panel is rotated in every 3 years. In January following the reference year, SLID sample households are contacted by telephone interviewers. Demographic information is collected for every person in the household. Complete survey data are collected for every person in the household 16 years or older. Questions are asked on labour (labour market activity, work experience, jobless spells and job information), educational attainment and income sources.

Reference year 2004 saw a major change in data collection for SLID. Previously, at the end of the January interview, respondents were informed that they would be contacted again in May when they would be asked to supply data on income as well as certain expense items. Starting in 2004, all of the questions were asked at the same interview, in January. The respondent may elect to grant permission to Statistics Canada to retrieve all the income data required from the T1 tax file, thereby avoiding the necessity of the second half of the interview. Collection of income data had been deferred until May so that the respondent would be more familiar with the required data (having just filed an income tax return).

Although originally designed as a longitudinal survey, SLID has always maintained the capability of producing cross-sectional estimates. This cross-sectional aspect took on new importance with the cancellation of the Survey of Consumer Finance after the 1997 reference year. At this time SLID became the primary source of cross-sectional household and family income data.

All persons who are members of selected SLID households in the beginning of the first year of a panel's existence are longitudinal sample persons for SLID. As such, it is these individuals that are followed longitudinally. Any (non-longitudinal) person living in a household with a longitudinal person is referred to as a cohabitant. Cohabitants living with cross-sectionally eligible longitudinal persons will also be cross-sectional sample persons.

For more information about survey concepts, definitions and design please refer to Statistics Canada publication: "*Survey of Labour and Income Dynamics - A survey overview*", <http://www.statcan.ca:8096/bsolc/english/bsolc?catno=75F0011X>

Sample surveys are subject to errors. As with all surveys conducted at Statistics Canada, considerable time and effort is taken to control such errors at every stage of the Survey of Labour and Income Dynamics. Nonetheless errors do occur. It is the policy at Statistics Canada to furnish users with measures of data quality so that the user is able to interpret the data properly. This report summarizes quality measures that have been produced in an attempt to describe the overall quality of SLID data.

2. Sample composition/Attrition

Although originally designed as a longitudinal survey, SLID also has the capability of producing cross-sectional estimates. Every non-longitudinal person living with a longitudinal respondent is also asked to participate in the survey. Such persons are called cohabitants. Table 2.1 and 2.2 show the composition of the SLID sample by province and by census metropolitan area (CMA) respectively, in terms of longitudinal sample persons who respond, longitudinal responding persons that are ineligible cross-sectionally (such as deceased, institutionalized and those who have moved outside the country), and responding cohabitants. Historical tables can be found in appendix 1.

The cross-sectional SLID sample coverage is maintained through the addition of cohabitants each year. The one exception is immigrants who arrive after the beginning of a panel and before the start of the next one and move into their own households, this introduces a small amount of under coverage. The longitudinal sample, however, is subject to attrition. Attrition is the gradual loss of respondents each year through the life of the panel. Table 2.3 shows the respondent status for persons originally selected as longitudinal respondents. In table 2.3 the responding longitudinal sample size is comprised of the in-scope respondents, the individuals who have moved to Yukon, North-West Territories or Nunavut, the individuals who have moved outside Canada, the institutionalized individuals and the deceased individuals.

Table 2.1 - Sample composition in SLID by province, 2004

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally*		Cohabitants		Cross-sectional sample size	
	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04
	Newfoundland	1,445	1,337	81	39	254	121	1,618
Prince Edward Island	920	931	59	28	151	103	1,012	1,006
Nova Scotia	2,148	2,180	144	74	433	274	2,437	2,380
New Brunswick	1,893	1,832	110	62	365	234	2,148	2,004
Quebec	6,031	6,337	436	221	1,343	856	6,938	6,972
Ontario	9,137	9,694	625	336	1,966	1,104	10,478	10,462
Manitoba	2,350	2,336	174	95	526	289	2,702	2,530
Saskatchewan	2,425	2,367	187	90	527	335	2,765	2,612
Alberta	2,733	2,727	139	71	717	449	3,311	3,105
British Columbia	2,691	3,048	187	111	540	414	3,044	3,351
Moved outside provinces	421	280	421	280	0	0	0	0
Total	32,194	33,069	2563	1,407	6,822	4,179	36,453	35,841

*This includes individuals who are deceased, institutionalized and those who have moved outside the country.

Table 2.2 - Sample composition in SLID by CMA, 2004

Census Metropolitan Area	Longitudinal sample size		Number of cohabitants		Cross-sectional sample size	
	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04
Halifax	493	461	126	79	619	540
Quebec City	510	432	142	86	652	518
Montréal	1,097	1,305	323	177	1,420	1,482
Ottawa - Gatineau	794	877	181	123	975	1,000
Toronto	1,719	1,695	406	234	2,125	1,929
Hamilton	420	408	80	52	500	460
St. Catharines - Niagara	422	477	101	44	523	521
Kitchener	445	478	142	49	587	527
London	433	433	84	51	517	484
Windsor	322	332	84	38	406	370
Winnipeg	1,031	1,095	273	164	1,304	1,259
Calgary	616	644	171	123	787	767
Edmonton	713	620	199	113	912	733
Vancouver	905	1,072	200	136	1,105	1,208
Victoria	238	293	38	37	276	330
Other CMA or CA	9,998	11,425	2,265	1,558	12,263	12,983
Do not live in a CMA	9,475	9,615	2,007	1,115	11,482	10,730
Not available**	2,563	1,407	0	0	0	0
Total	32,194	33,069	6,822	4,179	36,453	35,841

**This information is only available for those individuals who are cross-sectionally eligible.

Table 2.3 - Status of longitudinal persons, reference year 2004

Longitudinal status	Panel 03	Panel 04
In scope (respondents)	29,631	31,662
In scope (nonrespondents)	1,701	3,300
Moved to Yukon, NWT, Nunavut	10	6
Moved outside Canada	406	273
Institutionalized	617	384
Deceased	1,530	744
Removed from sample*	9,778	5,835
Duplicate person/error**	10	28
Total	43,683	42,232

* Respondents are removed from the sample for one of two reasons. If entire households have refused for 2 consecutive cycles they are said to be hard refusals and no further attempts are made to enumerate these households. As well after two years households that were not traced are not sent out for further attempts at collection.

**Respondents who were erroneously included in the household in the beginning of the first year of a panel's existence.

3. Sampling errors

Sampling errors occur because inferences about the survey population are based on data from a sample of that population rather than the entire population. The sample design, the variability of the characteristic being measured, and the sample size will all contribute to the magnitude of the sampling error.

The standard error is a common measure of sampling error. The standard error measures the degree of variation introduced in estimates by selecting one particular sample rather than another of the same size and design. Another widely used measure of the sampling error is the coefficient of variation (CV), which is the estimated standard error expressed as a percentage of the estimate.

In SLID, the bootstrap approach is used for the calculation of standard errors. This is a resampling method of variance estimation, often used when dealing with estimates from a complex sample design. Table 3.1 shows CV levels at the provincial and national level for a sample of key SLID estimates.

Table 3.1 - National and provincial coefficients of variation (%) for selected SLID variables

Variable	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Canada
Average total income	0.4	0.5	0.3	0.3	0.2	0.2	0.4	0.3	0.3	0.4	0.1
Average market income	1.6	1.3	0.9	1.0	0.5	0.4	0.7	0.7	0.5	0.8	0.2
Average wages and salaries	1.7	1.5	1.3	1.3	0.7	0.6	1.0	1.0	0.8	1.1	0.3
Average EI benefits	3.3	5.0	3.8	3.8	3.0	3.5	5.6	5.7	5.4	4.9	1.7
Average social assistance	6.9	13.9	10.9	9.6	5.2	4.9	12.2	10.7	6.1	9.9	2.6
Average other income	6.5	7.7	5.6	5.2	3.1	2.4	4.4	3.8	3.9	3.9	1.4
Prevalence of persons under LICO (after tax)	8.6	17.3	8.0	7.7	4.8	4.9	7.3	8.8	7.0	6.0	2.6
Counts of employed people	1.7	1.8	1.2	1.5	1.0	1.0	1.3	1.3	1.2	1.5	0.5

4. Coverage errors.

To produce good survey estimates, it is necessary that a survey sample adequately represent the survey population. To ensure proper coverage, SLID weights are adjusted using census population projections as control totals. The slippage rate is a measure of the percentage difference between these census projections and the survey estimate using weights prior to the application of this slippage related adjustment. More precisely, slippage is computed as

$$slippage_c = \frac{(CP_c - \sum_{k \in S_c} w_{kc})}{CP_c} * 100$$

where Class C is the group or class for which we want to calculate slippage rates. For example at a detailed level the groups are based on province, sex and age group.

CP_c is the census population projection for class C

w_{kc} is the survey weight for k_{th} responding unit in class C

S_c is the set of responding sample households in class C

Slippage rates for household surveys are generally positive because of frame under coverage.

Table 4.1 shows slippage rates at the person level by province and by age/sex groupings.

Table 4.1 - Person level slippage rates by province/sex/age group

Sex	Age group	Province										Total
		N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B. C.	
Male	0-6	4	-1	21	11	9	18	10	11	20	26	16
	7-15	-7	4	3	0	-8	10	19	-14	10	13	5
	16-17	-18	7	0	-1	-4	4	18	-14	8	9	2
	18-24	8	9	34	19	20	22	22	25	27	30	23
	25-34	28	7	19	28	26	34	14	22	28	36	30
	35-44	11	7	18	11	17	21	19	6	16	32	20
	45-54	4	1	20	5	14	11	7	11	15	23	13
	55-59	17	-19	-2	-6	9	8	8	-5	12	21	9
	60-64	22	-3	-3	-4	12	17	6	14	21	26	16
	65-69	20	4	9	1	5	8	9	18	17	16	10
70+	13	0	5	5	8	13	6	18	14	15	12	
	Total	10	2	14	9	12	17	13	10	18	25	16
Female	0-6	8	-6	17	11	9	20	10	9	20	25	17
	7-15	-9	-1	1	-4	2	9	8	-9	14	13	7
	16-17	-4	0	1	-4	6	0	5	-7	17	8	4
	18-24	8	13	17	3	22	11	14	13	22	26	17
	25-34	12	7	15	22	15	29	10	9	23	25	23
	35-44	0	2	10	3	6	12	11	-8	14	27	12
	45-54	4	4	10	0	15	6	5	0	6	21	10
	55-59	12	-7	-4	-10	2	4	2	7	19	11	5
	60-64	22	10	4	-6	6	11	9	12	18	17	10
	65-69	6	5	-2	12	6	12	4	17	15	15	10
70+	2	11	-4	11	4	11	4	10	14	17	10	
	Total	5	4	7	5	9	13	8	4	16	21	12
Total		7	3	11	7	11	15	11	7	17	23	14

As a comparison we will look at the person level slippage rates for the Labour Force Survey (LFS) by province. We will look at the slippage rates from the LFS at the beginning of the panel for each panel in tables 4.2 and 4.3. These rates are the rates associated with the rotation groups used by SLID.

Slippage rates were also computed at the household level and are summarized in Table 4.4. For slippage rates for previous reference years, see Appendix 2.

Table 4.2 - Person level slippage rates by province of SLID sample coming from the LFS, Panel 3

	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Canada
Total	10	10	11	2	6	7	6	7	14	14	8

Table 4.3 - Person level slippage rates by province of SLID sample coming from the LFS, Panel 4

	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Canada
Total	17	10	16	14	10	12	12	16	16	19	13

Table 4.4 - Household level slippage rates for provinces by household size

Province	Households size 1	Households size 2	Households size 3+	All households
Newfoundland	4	9	4	6
Prince Edward Island	-4	1	2	0
Nova Scotia	5	12	10	9
New Brunswick	-4	1	11	4
Quebec	13	11	9	11
Ontario	12	16	12	13
Manitoba	3	4	13	7
Saskatchewan	-12	10	8	3
Alberta	-7	18	18	12
British Columbia	14	23	21	20
Canada	9	15	13	12

5. Response rates

Since SLID has taken on the role of both a longitudinal and a cross-sectional survey, respective response rates are calculated. Cross-sectional response rates are calculated both at the person level and at the household level. Since sample persons have the option of giving tax permission thereby avoiding the income questions, it is possible to have complete data for income with no actual contact made during the reference year. Because of this the definition of a nonrespondent is not straightforward.

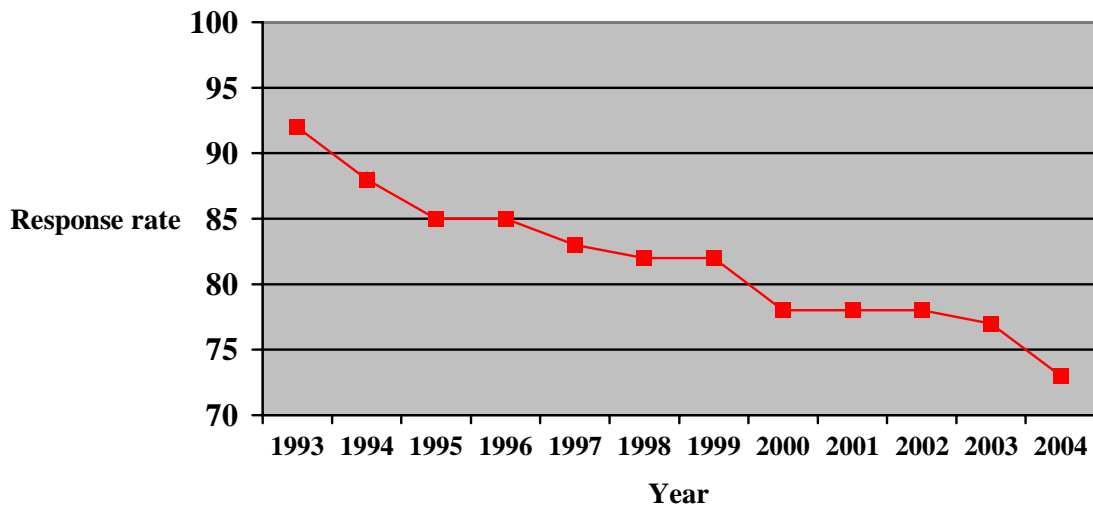
If all persons in a household are non-respondent to both labour and income questions, then these persons (and households) are nonrespondent.

With respect to those persons in households which are non-responsive to the labour questions but for whom we have tax data, it is determined whether the person is in the same household as the previous year (as of December 31). If the household is different this means the respondent has split from the original household. Since we have no information at all on the household composition of the new household, such persons are defined to be nonrespondent.

Persons in households which are non-responsive to the labour questions but for whom we have income data and for whom the household has not changed since the previous year, are considered nonrespondents if the household was a non-responding household to the labour questions the previous January. Since updates to household composition are collected with the labour questions, this means that the household composition has not been updated for 2 consecutive years. Persons in households that have been nonrespondent to labour questions in 2 consecutive January collections are therefore considered to be nonrespondents to SLID.

Figure 5.1 shows the cross-sectional person response rates to SLID throughout the years of the survey. The person level response rates are calculated by dividing the number of cross-sectionally eligible respondents to the labour and/or income questions by the total number of cross-sectionally eligible people. An assumption is made that nonrespondents are still in the target population unless there is evidence to the contrary. As a result this may somewhat underestimate response rates.

Figure 5.1 - Cross-sectional person response rates (%) (Age>15)



A household is considered a respondent household if at least one person in that household is considered a respondent. Household response rates are calculated by dividing the number of cross-sectionally eligible responding households by the total number of cross-sectionally eligible households. Once again an assumption is made; non-responding households are assumed to be still in the target population unless there is evidence to the contrary. As a result this may somewhat underestimate response rates.

Nonresponse can potentially introduce a bias in the data. A bias is created if characteristics of respondents differ from those of nonrespondents and this difference has an impact on the variable being studied. It is difficult to determine whether nonresponse is introducing bias, because there is a limited amount of information for nonrespondents. Figure 5.2 shows the household response rates by province.

Figure 5.2 - Cross-sectional household response rates by province (%)

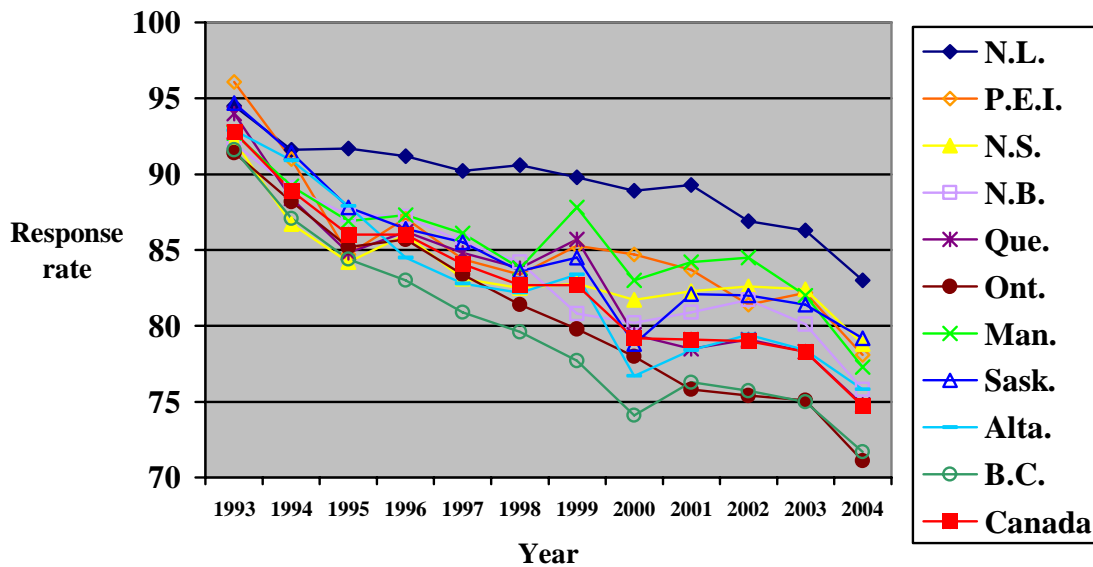


Table 5.1 shows the person response rates by phase. ‘Respondent to labour questions’ and ‘Respondent to income questions’ are the percentages of those who responded to only the labour or income sets of questions respectively whereas the ‘Respondent to both sets’ is the percentage of all those who responded in full or in part to both sets of questions.

Table 5.1 - Cross-sectional person response rates by phase (%)

Type	'93	'94	'95	'96	'97	'98	'99	2000	2001	2002	2003	2004
Respondent to labour questions	10	11	10	11	12	10	14	17	10	11	8	6
Respondent to income questions	6	3	3	3	2	3	3	5	4	5	5	5
Respondent to both sets	76	75	72	72	69	69	66	56	63	62	64	62
Non-response	8	11	15	14	17	18	18	22	22	22	23	27

Due to the conceptual difficulty in defining a longitudinal household, only person level longitudinal response rates are calculated. Table 5.2 shows person level longitudinal response rates by panel. These rates are calculated by dividing the number of longitudinal respondents by the original number of longitudinal persons selected in that panel. Figure 5.3 shows the longitudinal non-response by panel and wave.

Table 5.2 - Longitudinal person response rates (%) (All ages)

Panel	'93	'94	'95	'96	'97	'98	'99	2000	2001	2002	2003	2004
1	93	90	87	84	82	82
2	90	87	85	83	79	77
3	84	83	83	80	76	74
4	81	83	78

... not applicable

Figure 5.3: Longitudinal non-response by wave

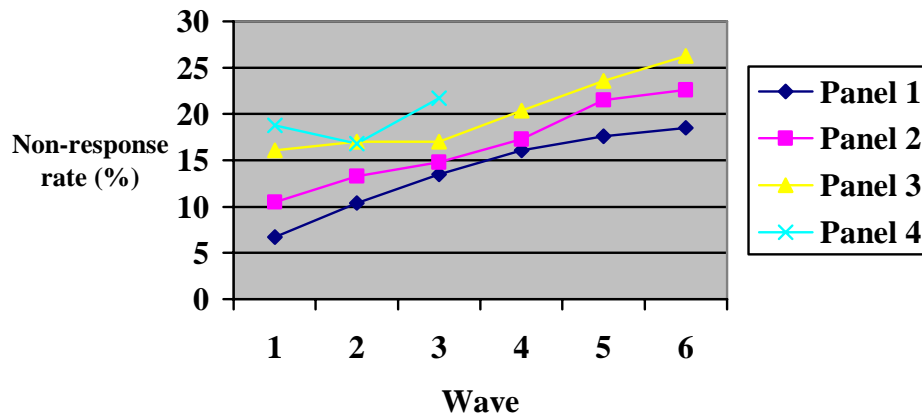
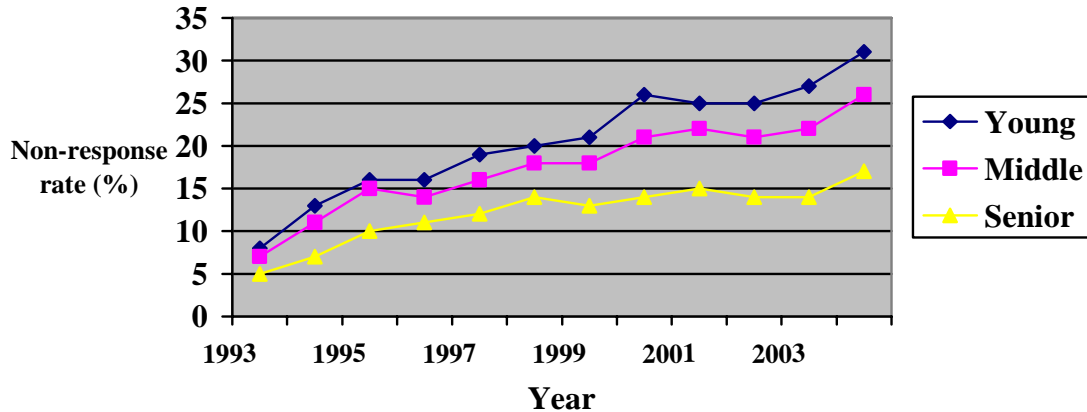


Figure 5.4 shows the longitudinal non-response rates each year by age group. ‘Young’ are people at least 16 years of age but less than thirty, ‘Middle’ are people thirty years of age or older but less than sixty years of age and ‘Senior’ are people at least sixty years of age.

Figure 5.4: Longitudinal non-response by age group



6. Tax permission rates

Prior to reference year 2004, there were two interviews every year: in January the interview was about activities such as working, going to school, looking for work or retirement. The second interview in May was about income, but wasn't necessary if the respondent gave Statistics Canada permission to obtain the required data from tax records. The tax source should provide consistent data of high quality and so a high permission rate should ensure good quality survey income estimates. The respondent was asked for this permission at the end of the January interview. If permission was not given, the respondent was contacted again in May. At this time the respondent was once again asked if he/she would prefer to give permission to access tax records. If permission was not provided, the interview proceeded. Starting in reference year 2004, permission was asked only once, in January. If it was not provided the interview continued immediately with the income questions.

Table 6.1 shows permission rates by panel for each phase of the survey. The option to give tax permission was given for the first time in the May collection for the 1994 reference year. Prior to this, all income data were collected through interview. Percentages in the table are based on the number of respondents over the age of 15 who are cross-sectionally eligible. Permission rates in reference year 2000 and 2004 are the same for both January and May because there were no May interviews in those years.

Table 6.1 - SLID permission rates by panel*

Panel (start date)	Wave											
	1		2		3		4		5		6	
	Interview Jan	Interview May	Interview Jan	Interview May	Interview Jan	Interview May	Interview Jan	Interview May	Interview Jan	Interview May	Interview Jan	Interview May
Panel 1 (1993)	59	70	75	79	82	84	85	86	88
Panel 2 (1996)	53	69	76	79	83	86	86	88	88	88	87	90
Panel 3 (1999)	55	72	76	76	78	85	87	90	90	92	93	93
Panel 4 (2002)	58	77	81	85	89	89

*Permission was asked for the first time in May 1994

... not applicable

7. Tax linkage rates

While respondents may grant Statistics Canada permission to use their tax data, they are not asked for their Social Insurance Number (SIN). Without a SIN to identify SLID respondents on the tax file, it is necessary to perform a linkage operation to find a respondent's SIN. The generalized record linkage system (GRLS) developed at Statistics Canada is used to perform this linkage.

After preprocessing of both the tax file and the SLID file to ensure compatible formatting of all match variables, a direct match is performed using 7 key matching variables. These matching variables are: Sex, province, soundex¹ code for surname, surname, date of birth, postal code and first initial. The SLID record can have no missing data for key matching variables. Output for the direct match is manually reviewed for errors where a SLID record matches to more than one tax record, where more than one tax record matches to a SLID record, and where the first given name is not the same on the 2 sources (only first initial is used in the tax match). The match rate on the direct match is approximately 55 percent.

The unmatched records are then run through a statistical match. Pockets² for matching are defined. The files are segmented into pockets with sex, province and surname soundex code defining a pocket. Every record within a pocket on the SLID file is compared with every record within the same pocket on the tax file. Factors of importance are assigned for full agreement, partial agreement, and disagreement. These factors are numeric values and are used to evaluate the likelihood that a pair of records (one from

1. Soundex is a name coding routine used in order to remove any common spelling errors from the surnames of respondents. This encoding is done based on the sound of the surname.

2. Pockets are groups of individuals on both the tax file and the SLID file with the same sex, province and soundex code.

SLID and one from tax) represent the same person. Factors are defined for each of the matching variables. Thresholds are defined whereby records are determined to be definite matches if their total factor is greater than the upper threshold or definite non-matches if their total factor is below the lower threshold. Manual verification is done to ensure the quality of the matches. Figure 7.1 gives the percentage of the SLID sample giving tax permission for which a SIN can be found. Since some respondents who give tax permission have not filed a tax return not all cases for which a SIN is found will result in successful tax linkages. Figure 7.2 gives tax linkage rates for those in the SLID sample for which we were successful in finding a SIN.

Figure 7.1: Percentage of SINs found

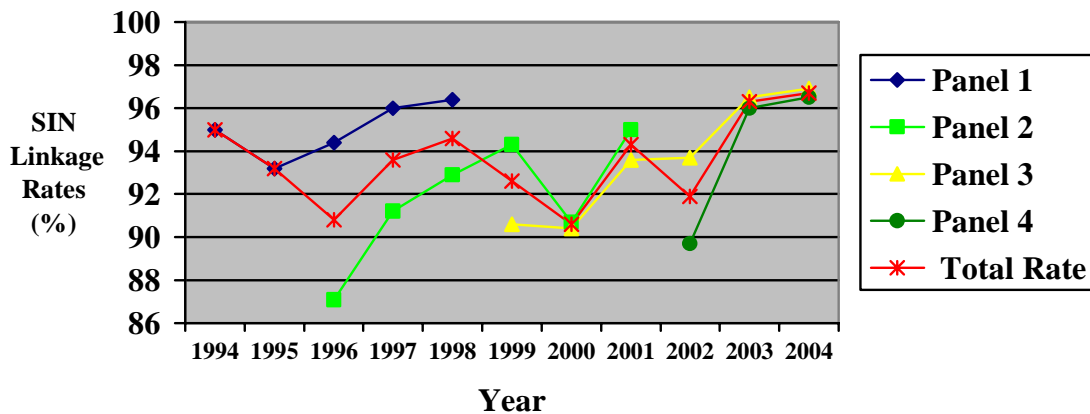
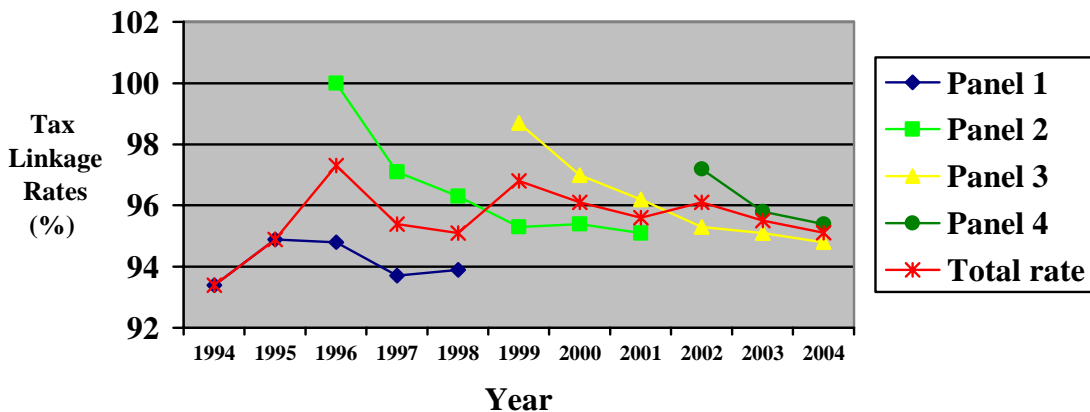


Figure 7.2: SLID/Tax linkage rates



8. Imputation rates

To compensate for non-responding households in the SLID sample, a non-response adjustment is applied to SLID weights. However, partially responding households are kept in the sample and any income data that is missing for individuals within responding households is imputed. These individuals may require complete imputation of all income

variables or they may require only certain fields to be imputed. Imputation rates in SLID may be thought of as a measure of partial non-response in the survey.

Imputation of income variables in SLID is done using a nearest neighbour approach. A set of basic consistency rules is defined and for a given record requiring imputation a set of consistent donors is identified. A set of matching variables, each of which are correlated with the variables to be imputed, is also defined. Through combined use of both a score function (for categorical matching variables) and a distance function (for numeric matching variables), the most similar consistent donor record is identified and used to impute data for the record.

The score function used in SLID income imputation is:

$$s(X, Y) = \sum_{k=1}^K p_k I(X_k, Y_k), \quad \text{where } I(X_k, Y_k) = \begin{cases} 1 & \text{if } X_k = Y_k \\ 0 & \text{if not.} \end{cases}$$

Note that p_k is a number allowing us to assign more or less importance to the matching variable k . X_k is the value of the receiver's variable k and Y_k is the value of the donor's variable k .

The distance function used in SLID income imputation is the same as the function used in the generalized edit and imputation system (GEIS). Suppose we have two records X and Y . The distance between the two is defined as:

$$d(X, Y) = \max_{j=1}^J |u_j(X) - u_j(Y)|$$

Where $u_j(X)$ is the function of the rank³ of X_j :

$$u_j(X) = \frac{\text{rank}(X_j)}{n_j + 1}.$$

J represents the number of quantitative variables used to calculate the distance, X_j represents the value of the quantitative variable j of the record X and n_j is the number of records with a valid value for this variable. When several records have the same value of the variable j , they are assigned a mean rank. Excluding these cases of equality the $u_j(\cdot)$ are uniformly distributed along the interval (0, 1).

The percentage of persons within responding SLID households that were subject to total or partial imputation is shown in Table 8.1. Recall that a responding SLID household is one in which at least one household member has responded partially or completely to either the labour or income questions of the survey.

Table 8.2 compares the proportion of records from tax to those collected in the telephone interview. In total eighteen income variables are imputed during SLID income

3. The rank is a method by which a numeric variable can be normalized. This way a numeric variable with a range from 0 to 9 and a numeric variable with a range from -999999 to 999999 have the same level of importance in the distance function.

imputation. Many individuals require only partial imputation. Partial imputation is when one or more income items is imputed with some information being supplied by the individual.

In table 8.3 we compare the percentage of tax data records requiring imputation to the percentage of records for which data is collected through the telephone interview. The need for partial imputation is determined after combining responses to both the labour and income questions. Inconsistencies are corrected through the imputation process. Table 8.3 also shows the percentage of individuals subject to partial imputation who require between one and seventeen variables to be imputed.

Table 8.1 - Persons requiring imputation of income variables, by province in 2004

Province	Total imputation (%)*	Partial imputation (%)**	No imputation (%)
Newfoundland	1	15	84
Prince Edward Island	1	15	84
Nova Scotia	1	17	82
New Brunswick	2	1	81
Quebec	2	15	83
Ontario	2	20	78
Manitoba	1	16	83
Saskatchewan	2	15	84
Alberta	2	19	80
British Columbia	2	19	80
Canada	2	18	81

*No information provided by the respondent. All data items imputed.

**One or more data items imputed with some information provided by the respondent.

Table 8.2 - Proportion of respondents coming from tax or interview in 2004

	Tax (%)	Interview (%)	Other* (%)
Proportion	83	5	12

*This comprises records that are not linked to Tax and without responses to Income questions.

Table 8.3 - Tax records and interview records requiring partial or total imputation in 2004

Record Type→	Tax (%)	Interview (%)	Other*(%)	All (%)
No imputation	93	73	n/a	81
Total imputation (18 variables)	n/a	n/a	13	2
Partial imputation	7	27	87	18
1 variable imputed	7	14	0.0	7
2 – 9 variables imputed	0	13	0.0	1
10 – 17 variables imputed	0	0	87	10

*Records that are not linked to Tax and without responses to the income questions. Some of these records are partially imputed based on the information collected from the labour questions.

In 2002, new housing content relevant for housing research and policy development was added to SLID in cooperation with the Canada Mortgage and Housing Corporation (CMHC). The survey now collects information for the following sub-populations beginning with the 2002 reference year: the need for repairs (as determined by the dwelling occupant); the principal heating fuel of the dwelling; and whether a farm or home business is operated from the property. Also from homeowners the amount of regular mortgage payments; the amount of annual property taxes; and whether the dwelling is part of a registered condominium is collected. From renters the following is collected: the amount of monthly rent, what amenities are included in the rent (*e.g.*, heat, water, electricity); and whether the rent is subsidised by government or an employer.

The above information is in addition to information about home ownership and type of dwelling (since 1994) and information on the presence of a mortgage and the number of bedrooms in dwellings (since 1999).

Table 8.4 - Households requiring imputation of housing variables, by province in 2004

Province	Total imputation (%)*	Partial imputation (%)**	No imputation (%)
Newfoundland	5	32	63
Prince Edward Island	6	32	63
Nova Scotia	6	28	67
New Brunswick	14	18	67
Quebec	4	25	72
Ontario	2	35	63
Manitoba	4	35	62
Saskatchewan	4	32	63
Alberta	3	36	61
British Columbia	2	37	61
Canada	4	31	65

*No information provided by the respondent. All data items imputed.

**One or more data items imputed with some information provided by the respondent.

Because of non-response to specific questions, imputation of housing related content was introduced in SLID in 2002. Two methods of imputation were used, longitudinal imputation and cross-sectional donor imputation. The cross-sectional donor imputation uses a similar method to that used in the income imputation, making use of the score function described above. Table 8.4 shows the percentage of responding SLID households that were subject to total or partial imputation.

In total twenty housing variables are imputed during SLID housing imputation. Many households require only partial imputation. Table 8.5 shows the break down of those requiring partial imputation.

Table 8.5 - Records requiring partial imputation in 2004

Partial imputation requirement	All (%)
Total	31
1 variable imputed	11
2 – 5 variables imputed	10
6 – 19 variables imputed	11

9. Rounding of income data

A small percentage of SLID income data comes from data collected in a telephone interview. While data obtained from the tax file is thought to be consistent for the most part, the quality of data coming from collection is not known. While some respondents may give precise amounts, it is possible that many of the responses given are estimates or approximations, which therefore are stated in hundreds or thousands of dollars rather than precise dollars and cents.

To test for the possible presence of rounding, distributions of each of the last 4 digits of reported variables were produced. One would normally expect the distribution to be approximately uniform with the digits 0 to 9 each comprising about 10 percent of the distribution. A prevalence of zeroes in the last digit would indicate rounding to the nearest 10, in the second last digit rounding to 100, etc. Table 9.1 shows the distribution of each of these digits for all reported values greater than ten thousand of the variable wages and salaries from both collected data (e.g. collected by interview) and tax data. Table 9.2 shows the prevalence of zeroes in each of the last 4 digits for all reported non-zero values for a selection of SLID variables.

Table 9.1 - Distribution of the last 4 digits of wages and salaries for collected data versus Tax data (greater than 10,000) in 2004

Digit	Fourth last digit (%)		Third last digit (%)		Second last digit (%)		Last Digit (%)	
	Collected	Tax	Collected	Tax	Collected	Tax	Collected	Tax
0	30	12	86	12	94	13	95	14
1	5	11	1	10	1	10	0	10
2	10	11	2	10	1	10	0	10
3	5	10	1	10	0	9	0	10
4	7	10	1	10	1	10	1	10
5	16	10	4	10	1	10	1	10
6	7	10	1	10	1	10	0	10
7	6	10	2	10	1	10	0	9
8	8	9	2	10	1	10	0	10
9	5	10	1	10	1	10	0	10

Table 9.2 - Prevalence of zeroes in the last 4 digits of reported data for selected variables in 2004

Variable	Fourth last digit (%)	Third last digit (%)	Second last digit (%)	Last digit (%)
Wages and salaries	24	75	91	95
Investment income	9	32	64	86
Social assistance	9	28	60	77
UI Benefits	6	46	78	87
Non-farm self-employment income	36	78	93	94

Appendix 1: Sample composition in SLID by province, 1996 - 2003

1996 Sample composition

Province	Longitudinal sample size		Longitudinal Sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
	Newfoundland	2039	1692	74	18	290	103	2255
Prince Edward Island	751	1180	33	5	125	56	843	1231
Nova Scotia	2300	2620	73	26	375	148	2602	2742
New Brunswick	2118	2441	62	21	322	168	2378	2588
Quebec	6146	7537	238	59	923	360	6831	7838
Ontario	9046	11972	335	84	1557	682	10268	12570
Manitoba	2245	2754	87	19	387	181	2545	2916
Saskatchewan	2415	2468	124	25	373	222	2664	3112
Alberta	3156	2915	89	25	695	222	3751	3112
British Columbia	2998	3280	71	27	563	227	3490	3480
Moved outside provinces	149	126	149	126	0	0	0	0
Total	33,352	38,985	1,335	435	5,610	2,312	37,627	40,862

1997 Sample composition

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
	Newfoundland	1998	1624	87	35	312	148	2223
Prince Edward Island	734	1120	41	13	140	90	833	1197
Nova Scotia	2234	2500	98	38	410	264	2546	2726
New Brunswick	2068	2308	79	36	369	258	2358	2530
Quebec	6070	7325	270	102	1104	664	6904	7887
Ontario	8831	11550	395	181	1841	1196	10277	12565
Manitoba	2193	2687	105	48	434	288	2522	2927
Saskatchewan	2368	2406	147	47	436	247	2657	2606
Alberta	3137	2862	102	51	870	397	3905	3208
British Columbia	2929	3161	98	60	598	357	3429	3458
Moved outside provinces	196	337	196	337	0	0	0	0
Total	32,758	37,742	1,618	810	6,514	3,909	37,654	40,841

1998 Sample composition

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
	Newfoundland	1961	1588	89	45	362	153	2234
Prince Edward Island	708	1076	54	29	149	127	803	1174
Nova Scotia	2206	2456	118	67	484	339	2572	2728
New Brunswick	2026	2250	97	58	447	287	2376	2479
Quebec	6007	7198	310	143	1268	865	6965	7920
Ontario	8682	11253	442	268	2057	1427	10297	12412
Manitoba	2130	2603	127	72	461	333	2464	2864
Saskatchewan	2318	2332	155	75	470	314	2633	2571
Alberta	3123	2900	97	65	972	539	3998	3374
British Columbia	2895	3084	125	78	656	413	3426	3419
Moved outside provinces	472	346	472	346	0	0	0	0
Total	32,394	37,086	1,952	1,246	7,326	4,797	37,768	40,637

1999 Sample composition

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
	Newfoundland	1550	1578	55	15	179	83	1674
Prince Edward Island	1065	1005	36	8	165	31	1194	1028
Nova Scotia	2384	2282	102	19	375	136	2657	2399
New Brunswick	2159	2110	68	15	336	113	2427	2208
Quebec	7017	7309	216	86	1048	272	7849	7495
Ontario	10758	10510	347	110	1723	482	12134	10882
Manitoba	2573	2843	93	27	398	136	2878	2952
Saskatchewan	2265	2783	94	35	369	205	2540	2953
Alberta	2871	2995	85	22	612	208	3398	3181
British Columbia	2988	3114	108	34	468	203	3348	3283
Moved outside provinces	375	130	375	130	0	0	0	0
Total	36,005	36,659	1,579	501	5,673	1,869	40,099	38,027

2000 Sample composition

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02	Panel 01	Panel 02
	Newfoundland	1495	1591	66	22	200	129	1629
Prince Edward Island	1031	1024	46	17	162	71	1147	1078
Nova Scotia	2274	2351	130	36	441	200	2585	2515
New Brunswick	2060	2194	91	29	359	210	2328	2375
Quebec	6493	6970	270	158	1179	526	7402	7338
Ontario	10302	10671	418	191	1913	853	11797	11333
Manitoba	2402	2747	120	48	409	244	2691	2943
Saskatchewan	2121	2664	116	58	414	268	2419	2874
Alberta	2735	2815	105	40	620	292	3250	3067
British Columbia	2809	2977	136	65	513	279	3186	3191
Moved outside provinces	446	235	446	235	0	0	0	0
Total	34,168	36,239	1,944	899	6,210	3,072	38,434	38,412

2001 Sample composition

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 02	Panel 03	Panel 02	Panel 03	Panel 02	Panel 03	Panel 02	Panel 03
	Newfoundland	1,477	1,591	75	42	232	214	1,634
Prince Edward Island	1,005	1,014	57	27	195	99	1,143	1,086
Nova Scotia	2,263	2,378	148	63	543	302	2,658	2,617
New Brunswick	2,024	2,214	109	53	435	306	2,350	2,467
Quebec	6,341	6,825	324	217	1,348	871	7,365	7,479
Ontario	10,063	10,376	518	289	2,233	1,335	11,778	11,422
Manitoba	2,407	2,739	143	79	466	406	2,730	3,066
Saskatchewan	2,087	2,785	140	87	494	421	2,441	3,119
Alberta	2,764	2,910	122	63	759	490	3,401	3,337
British Columbia	2,813	3,075	157	95	601	386	3,257	3,366
Moved outside provinces	472	337	472	337	0	0	0	0
Total	33,716	36,244	2,265	1,352	7,306	4,830	38,757	39,722

2002 Sample composition

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04
	Newfoundland	1,552	1,368	49	15	244	54	1,747
Prince Edward Island	982	972	40	6	136	68	1,078	1,034
Nova Scotia	2,307	2,239	91	28	386	116	2,602	2,327
New Brunswick	2,095	1,923	71	18	345	119	2,369	2,024
Quebec	6,544	6,557	320	89	1,084	371	7,308	6,839
Ontario	9,890	10,222	400	112	1,552	492	11,042	10,602
Manitoba	2,627	2,542	105	35	488	176	3,010	2,683
Saskatchewan	2,626	2,410	136	28	435	168	2,925	2,550
Alberta	2,846	2,829	97	31	607	262	3,356	3,060
British Columbia	2,897	3,126	135	34	476	186	3,238	3,278
Moved outside provinces	403	108	403	108	0	0	0	0
Total	34,769	34,296	1,847	504	5,753	2,012	38,675	35,804

2003 Sample composition

Province	Longitudinal sample size		Longitudinal sample ineligible cross-sectionally		Cohabitants		Cross-sectional sample size	
	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04	Panel 03	Panel 04
	Newfoundland	1,504	1,387	66	25	242	97	1,680
Prince Edward Island	949	1,003	52	15	138	90	1,035	1,078
Nova Scotia	2,232	2,268	117	54	425	197	2,540	2,411
New Brunswick	1,993	1,957	91	49	369	182	2,271	2,090
Quebec	6,285	6,748	373	161	1,241	692	7,153	7,279
Ontario	9,504	10,498	520	243	1,835	1,019	10,819	11,274
Manitoba	2,499	2,501	139	59	501	262	2,861	2,704
Saskatchewan	2,489	2,481	159	54	496	278	2,826	2,705
Alberta	2,768	2,866	124	47	637	363	3,281	3,182
British Columbia	2,785	3,186	153	73	522	291	3,154	3,404
Moved outside provinces	380	241	380	241	0	0	0	0
Total	33,388	35,136	2,174	1,021	6,406	3,471	37,620	37,586

Appendix 2: Slippage rates over time

Slippage rates by province and year, 1996 to 2003

Year	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Canada
1996	2	6	3	3	4	9	5	1	11	15	8
1997	-1	5	2	3	3	10	4	-1	13	18	8
1998	1	9	3	4	4	10	6	-1	13	18	9
1999	-1	4	5	2	2	11	-3	-2	13	17	8
2000	-7	3	3	-0	4	12	-0	-1	16	20	10
2001	-5	3	1.	-1	6	14	-2	-4	16	20	11
2002	7	8	10	7	9	12	8	8	17	20	12
2003	8	5	11	7	11	13	10	8	17	22	14

Slippage Rates for province by household size, 1996

Province	Households size 1	Households size 2	Households size 3+	All households
Newfoundland	0	6	-0	2
Prince Edward Island	10	2	5	5
Nova Scotia	-8	9	1	2
New Brunswick	6	1	1	2
Quebec	-3	5	4	2
Ontario	6	11	5	7
Manitoba	-5	13	2	4
Saskatchewan	-23	11	-2	-4
Alberta	1	11	-2	8
British Columbia	7	14	15	13
Canada	1	10	5	6

Slippage Rates for provinces by household size, 1997

Province	Households size 1	Households size 2	Households size 3+	All households
Newfoundland	-6	-2	2	-1
Prince Edward Island	8	2	5	5
Nova Scotia	-17	5	4	-1
New Brunswick	-2	-4	4	1
Quebec	-6	5	4	2
Ontario	6	13	8	9
Manitoba	-8	13	2	3
Saskatchewan	-21	9	-3	-4
Alberta	-5	10	12	7
British Columbia	4	18	18	14
Canada	-1	10	8	6

Slippage Rates for province by household size, 1998

Province	Households size 1	Households size 2	Households size 3+	All households
Newfoundland	-12	3	3	1
Prince Edward Island	3	6	11	8
Nova Scotia	-23	7	6	-1
New Brunswick	0	1	4	2
Quebec	-3	6	5	3
Ontario	4	13	9	9
Manitoba	-6	10	6	4
Saskatchewan	-25	15	-4	-3
Alberta	-14	15	12	7
British Columbia	-3	17	20	13
Canada	-3	11	9	7

Slippage rates for province by household size, 1999

Province	Households size 1	Households size 2	Households size 3+	All households
Newfoundland	-24	4	2	-2
Prince Edward Island	-9	-1	10	2
Nova Scotia	-13	7	7	2
New Brunswick	1	1	4	2
Quebec	-2	3	3	2
Ontario	8	16	10	12
Manitoba	-120	-7	-1	-6
Saskatchewan	-23	5	1	-5
Alberta	-10	8	13	6
British Columbia	3	14	21	14
Canada	-1	10	9	7

Slippage rates for province by household size, 2000

Province	Households size 1	Households size 2	Households size 3+	All households
Newfoundland	-32	-7	0	-7
Prince Edward Island	-10	-2	8	1
Nova Scotia	-13	2	6	0
New Brunswick	-6	-3	5	-0
Quebec	5	4	4	4
Ontario	5	15	11	11
Manitoba	-11	-9	3	-5
Saskatchewan	-16	5	1	-2
Alberta	-3	12	17	11
British Columbia	-1	16	25	15
Canada	1	10	10	8

Slippage rates for province by household size, 2001

Province	Households size 1	Households size 2	Households size 3+	All households
Newfoundland	-22	-6	2	-5
Prince Edward Island	-5	-1	8	2
Nova Scotia	-17	6	4	-1
New Brunswick	-13	-1	4	-2
Quebec	7	7	7	7
Ontario	7	18	12	13
Manitoba	-17	7	-0	-7
Saskatchewan	-17	1	-2	-5
Alberta	-7	9	18	9
British Columbia	1	16	25	16
Canada	2	11	12	9

Slippage rates for province by household size, 2002

Province	Households size 1	Households size 2	Households size 3+	All households
Newfoundland	-5	4	13	7
Prince Edward Island	8	6	10	8
Nova Scotia	-4	7	15	7
New Brunswick	-7	-1	15	5
Quebec	11	7	10	9
Ontario	12	12	12	12
Manitoba	-3	-3	11	2
Saskatchewan	-9	13	12	7
Alberta	8	13	17	13
British Columbia	11	15	24	17
Canada	9	10	14	11

Slippage rates for province by household size, 2003

Province	Households size 1	Households size 2	Households size 3+	All households
Newfoundland	5	8	5	6
Prince Edward Island	-1	3	4	3
Nova Scotia	1	11	11	9
New Brunswick	1	3	9	5
Quebec	14	11	8	11
Ontario	13	15	10	12
Manitoba	7	2	12	7
Saskatchewan	-5	11	8	5
Alberta	1	17	16	13
British Columbia	16	20	20	19
Canada	11	14	11	12