

Catalogue N° 98-15

**LABOUR MARKET DYNAMICS: INFORMATION
FROM NEW CANADIAN LONGITUDINAL SOURCES**

Product registration Number 75F0002M

September 1998

Maryanne Webber¹, Income Statistics Division

Mike Sheridan ¹, Income Statistics 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 Client Services Unit, at 7-B5 Jean-Talon Building, Statistics Canada, Ottawa, Ontario, CANADA K1A 0T6; by INTERNET: (DYNAMICS@STATCAN.CA); by telephone (613) 951-7355 or toll-free 1-888-297-7355; or by fax (613) 951-3012.

¹ Maryanne Webber, Statistics Canada, Ottawa K1A 0T6, Canada, webber@statcan.ca

¹ Mike Sheridan, Statistics Canada, Ottawa K1A 0T6, Canada, shermik@statcan.ca

EXECUTIVE SUMMARY

Since the early 1990s, a number of new longitudinal data sources have been developed and have begun to yield analytical results. Among these new data sources are the Survey of Labour and Income Dynamics (SLID), first conducted in 1993, and the Longitudinal Administrative Database (LAD), created retrospectively from tax records. These and other data sets are now being exploited. Research themes include transitions from school to work and from work to retirement, earnings mobility, labour market adjustment and the impacts of layoff on the family. This paper reviews some of the substantive findings that have emerged from recent studies. It then discusses the “growing pains” that can occur as these complex sources find their niche, and explores some of the lessons learned in the Canadian context.

KEY WORDS: Employment Dynamics; Labour market dynamics.

TABLE OF CONTENTS

	Page
1. Introduction	1
1.1 Background	1
1.2 Organisation of the paper	2
2. Recent longitudinal labour market data sources: Overview	2
2.1 Survey of Labour and Income Dynamics	2
2.2 Longitudinal Administrative Databank	3
2.3 Self-Sufficiency Project	3
2.4 Canadian Out-of-Employment Panel	4
2.5 Other recent longitudinal sources	5
3. Some highlights from recent studies	5
3.1 Family dynamics	6
3.2 Intergenerational Income Mobility	7
3.3 Earnings mobility	9
3.4 Consequence of Job Loss	10
3.5 Impact of work interruptions on women's earnings	11
3.6 Persistence of low income and dependence on Income Assistance	12
4. Lessons learned in launching panel surveys	14
4.1 Design issues	15
4.2 Development of processing system	17
4.3 Dissemination strategy	19
4.4 Protecting confidentiality	20
5. Conclusion	21
Bibliography	22

1. INTRODUCTION

1.1 Background

In the 1990s, several new longitudinal data sources emerged in Canada, designed to provide information on labour market and income dynamics. The research community is just beginning to exploit these sources, most of which currently have two or three years of available data. It is well known that longitudinal data sources can take several years to pay dividends in terms of analytical results that are directly useful to the development of social and labour market policy. One hurdle is inherent in the very nature of these datasets: enough time must pass for events of interest to occur and for outcomes to become evident. A second hurdle is that researchers must generally invest considerable time to understand any particular longitudinal dataset. This task is especially challenging for the pioneers who are mining a data source that few other researchers have used.

We can speed up the process of obtaining useful analytical results, by supporting analysis projects that can be done using two or three years of data. Research projects undertaken early in the life of a new survey provide valuable feedback to survey staff on the strengths and weaknesses of the dataset. The sooner this feedback begins the better.

This article reviews highlights from several recent longitudinal studies. The intention is to provide a flavour of the types of projects that have been initiated by researchers in Statistics Canada, by other departments of the federal government and in universities. Although it is still early days, we can anticipate how these results will influence social and labour market policy. The paper then discusses some of the pitfalls faced in the design and implementation of new longitudinal data sources.

1.2 Organisation of the paper

In Section 2, several new longitudinal data sources dealing primarily with labour market and income issues are described. This is followed by a selection of highlights from recent studies in Section 3. Section 4 describes some of the lessons learned in the development of one of these sources, the Survey of Labour and Income Dynamics.

2. RECENT LONGITUDINAL LABOUR MARKET DATA SOURCES: OVERVIEW

Four recent longitudinal sources are outlined below. The first two are intended for a variety of uses. In contrast, the last two were developed with very specific policy applications in mind.

2.1 Survey of Labour and Income Dynamics

SLID is a multi-purpose household panel survey, designed to address a range of issues including school-to-work and work-to-retirement transitions, family income stability, job quality, the determinants of income and earnings inequality, duration and concentration of unemployment, and the incidence and duration of spells of Income Assistance (welfare). The sample consists of overlapping panels, each one lasting six years. Each panel starts with about 15,000 households. All household members are followed through time and new people with whom they live during the six-year period are also covered. In addition to extensive “historical” information, covering marital history, fertility, work experience and educational attainment, persons 16 and over are interviewed every January about their labour market activities throughout the previous year. Detailed income information is obtained from their tax records, unless they do not file a tax return or would prefer to provide this information by interview. Income interviews are

conducted in May. SLID's first panel started in 1993 and the second one in 1996. The third panel will begin in 1999, when the first panel ends.

2.2 Longitudinal Administrative Databank

The LAD file is built up from personal tax records and currently covers the period 1982 to 1995. It is a 10% sample of the "family file", which groups tax filers and their dependants into census families (roughly, nuclear families) by linking filers' tax returns and making use of information on dependants contained on each filer's tax return. The LAD file currently has over two million records. Every year, a sample of "new filers" is added, so the LAD remains representative cross-sectionally.

The LAD includes information on each sampled person, on the person's spouse or parent, on his or her family and children. In addition to basic demographics, the file includes a detailed set of income sources and many additional variables, which can be derived from the income sources. Because of the sample size, it is ideal for studying such topics as the impact of family events on income and geographical mobility. Also, because the file goes back to 1982, it is a good source for studies that cannot be done with shorter panels. The trade-off is its relatively limited content.

2.3 Self-Sufficiency Project

Human Resources Development Canada funds the Self-Sufficiency Project (SSP) and Statistics Canada is one of many partners. SSP is a research demonstration project. It is designed to determine the effectiveness of an earnings supplement for single parents receiving Income Assistance. Many IA recipients are employable but could not command wages that are high enough to support their families, considering the costs of childcare, transportation and other employment-related expenses. Currently, earned income is deducted from IA benefits

(although specific practices vary somewhat from one jurisdiction to another). The purpose of the project is to test the effectiveness of an earnings supplement for single parents who obtain a job and agree to leave IA. The earnings supplement can last up to three years. During this time, the parent is accumulating work experience that could result in higher wages.

The sample consists of 9600 parents selected from the Income Assistance administrative files of two provinces from November 1992 to April 1995. Sampled individuals are randomly assigned to the program group or a control group. A baseline interview is conducted, with three follow-up interviews spaced out over four to five years. The interviews cover, among other things, human capital questions, labour market activities, marital history, childcare needs, hardship, income from various sources, attitudes about work and welfare and home life. The second follow-up survey also includes a battery of questions for the children.

2.4 Canadian Out-of-Employment Panel

In 1994, the Canadian government implemented a major reform of the Employment Insurance program. Chief among the changes were an increase in the number of weeks worked required to be eligible to Employment Insurance benefits and a reduction in the benefit rate. In addition, individuals who left their job voluntarily or were dismissed for cause became ineligible. To understand the impact of these changes on workers and their families, Human Resources Development Canada sponsored the Canadian Out-of-Employment Panel (COEP) Survey.

The sample is selected from an administrative file known as the Record of Employment file. A "Record of Employment" is generated when a work interruption occurs. Between July 1, 1995 and December 31, 1997, a sample of 4,000 persons was selected each quarter from this file. Each sample is a cohort.

Respondents are asked about their employment history over the past 12 months. Ten months later, a follow-up interview.

The survey collects information on labour market activities, background demographics, family situation, income, expenditures, assets and debts.

2.5 Other recent longitudinal sources

There are several other new panel surveys in different fields, which include summary labour market and income information. These sources are important because they will offer insights into the determinants of labour market success and socio-economic well-being. Examples include the National Population Health Survey and the National Longitudinal Survey of Children and Youth, both of which were launched in the early 1990s. In addition, the National Graduate Survey and Graduate Follow-up Survey has been a valuable source of information on the labour market integration of recent graduates.

Although designed as a cross-sectional survey, the General Social Survey includes many retrospective questions. This data source is often exploited using longitudinal techniques and for this reason, a GSS-based study is included below.

3. SOME HIGHLIGHTS FROM RECENT STUDIES

Our objective is to provide a bird's-eye view of recent work, including studies in progress. We begin with family dynamics. Though research in this field often entails challenging data handling problems, it offers important insights into the links between family events, labour market behaviour and income. With the focus still on the family, we then stretch the time dimension to look at intergenerational income mobility. Next, we outline some recent work on earnings mobility and on the impact of Employment Insurance changes on the consequences of job loss. Then, as a particular case of life-cycle labour market dynamics, we report a study

of the impact of work interruptions on the earnings of women. Finally, we review some work on the persistence of low income spells and dependence on Income Assistance.

3.1 Family dynamics

Cross-sectional data show that family dissolution and re-formation is an everyday reality of Canadian life. The longitudinal sources hold some promise for understanding the financial origins and outcomes of family change. In its first year, the SLID panel showed that one in seven people aged 15 and over experienced a major change in their family composition during 1993 (*Dynamics, 1996*). Thus, large segments of the population live through an important family event, such as a birth, death, marriage or divorce in their family in any one year. If the change involves the gain or loss of a breadwinner, it can have major repercussions on the family's financial picture.

Well-known studies based on U.S. panel surveys, including the Survey of Income and Program Participation and the Panel Study on Income Dynamics, show that when a major family event occurs, it often triggers a flow into or out of poverty (*Ruggles, 1987; Duncan, 1984*). In Canada, the longitudinal sources are being investigated to see what they can tell us about family instability and income inadequacy. Using two years of data from SLID, Picot and Zybblock (*1998*) found that family events, particularly marriage and separation, are very strongly associated with the probability that a child entered or left a state of low income during 1994. Parental separation increased the risk that a child would fall below the low-income line in 1994 from 6% to 66%, relative to a child whose parents did not separate that year.² Conversely, marriage or a new common-law union increases the probability that a child will rise above the low income line to “virtual certainty”.

² This study was based done using low income measures, defined as 50% of median income, adjusted for family size and composition using an equivalence scale.

A LAD-based study by Galarneau and Sturrock (1997) depicts the strikingly different situation of male and female spouses after separation. They looked at married people, with children at home, who separated between 1987 and 1993. One year after separation, 89% of the women were living with children, compared with 36% of the men. (Five years later the rates had converged somewhat.) For the women, family income, expressed on a per capita basis, dropped on average 23% the year after separation; for men it rose 10%. Five years after separation, women had recovered most but not all of the lost ground, with per capita income still 5% below its pre-separation level.

In a separate study, Galarneau and Sturrock (1998) looked at couples with no children at home who separate. As a general rule, these individuals are older. Financially, the male/female gap in income after separation is smaller than for couples with children. This study attests to the importance of family reconstitution: within five years of separation, over half were in a new relationship, despite the fact that over 40% of the men and 30% of the women were over age 50 at the time of separation.

The linkages that exist between changes in financial status and marital stability are complex. One SLID study of common-law unions found a relationship between increases in family income, hours worked, occupational skill level and the likelihood of remaining together (Wu, 1998). However, it is difficult to draw causal inferences.

3.2 Intergenerational Income Mobility

The studies referred to above illustrate the short-term impact of an important event on family income. Another area of research, sometimes based on

administrative data, sometimes on retrospective survey data, concerns intergenerational change.³

Corak and Heisz (1997) use data from the income tax system to look at the level and composition of income of young men and women according to their father's income. The authors demonstrate that "there is a strong relationship between the income levels of a father and the income level that his son or daughter will ultimately earn as an adult". Investment income and reliance on transfer income also seem to have some transmitted effects.

Parental loss through death or divorce during the teenage years has also been shown to be negatively related to measures of financial independence in young adulthood (Corak, 1998). This study used earnings as a young adult (age 25-32), reliance on government transfers and marital/fertility behaviour as outcome measures. These outcome measures were different according to whether or not the young adult's family had remained intact throughout his or her teenage years. However, there was little difference between young adults who had lost a parent through death versus divorce, after controlling for the labour market activities and income of parents prior to the bereavement or divorce. Thus the parental conflict preceding divorce did not appear to be that influential in determining the young adult's financial status.

Information collected through retrospective questions in household surveys has been used to look at intergenerational changes in education and occupation. Working with SLID data, Mata (1997) showed that parents, particularly fathers "transmit" their education level to their children, but not all population subgroups are able to convert their higher schooling levels into occupational and monetary rewards. In particular, significant blockages exist in the transmission process for

³ In 1997, a conference on *Intergenerational Equity in Canada* brought together a large number of studies on such long-term effects. Some of the studies mentioned in this section were presented at that conference and published in Corak (1998 ab).

Canadian-born and foreign-born visible minority daughters as well as Aboriginal sons and daughters.⁴

3.3 Earnings mobility

The increase in earnings inequality in Canada has manifested itself through a growing gap between older and younger workers, particularly among men. The availability of longitudinal data offers some prospect of understanding the individual trajectories underlying changes in inequality. Morissette (1998b) uses tax data to assess the long-term impact of the drop in real earnings of young men. Specifically, the earnings of men who were 18 in 1975 are traced over the subsequent ten years and compared with the same information for men who were 18 in 1984. He showed that “today’s young workers will spend a greater fraction of their career receiving low earnings than their counterparts in the mid-seventies...they start at lower wages and do not catch up even after ten years.” This increase in inequality is due to two factors: a divergence in hourly wage rates and polarisation of weekly hours worked.

This study points to a decrease in earnings mobility. Finnie and Gray (1998) have also examined the issue using the LAD file. Specifically, they looked at the probability that an individual would change income quintiles over time according to how long he or she had been in a given quintile. They showed that the probability of exiting a given quintile declines strongly as the duration of the spell in that quintile increases. This study also shows that upward mobility has declined in the 1990s.

Using three years of data from SLID, Morissette and Drolet (1998) examined the upward mobility of workers with low wages. Specifically, they defined “low wages” as equivalent to the low income cutoff for a family of two living in a city of 500,000+ population. Their study population consisted of persons employed as

⁴ SLID results for Aboriginal Peoples refer to persons living off-Reserve in the ten provinces.

paid workers at the end of both 1993 and 1995, who were not in school full-time, and whose weekly earnings were less than the low income cutoff. These individuals were defined as “moving up” if their 1995 earnings exceeded the cutoff (adjusted for inflation) by at least 10%.

Their study showed that roughly one quarter of all workers had low weekly earnings in 1993. This is partly because part-time workers are included. About 8 in 10 part-timers had weekly earnings below the cutoff. About one in five of the workers with low weekly earnings in 1993 moved up in 1995. The “escape” from low weekly earnings came easier to workers with high levels of education, in professional occupations, who changed jobs voluntarily and who moved from a non-unionized to a unionized job.

Morissette and Drolet also noted that family composition has a big impact. Only 8% of working lone mothers moved out of low weekly earnings, compared with 17% of unmarried women without dependent children. Workers who quit one job and began a new one were 1.5 times more likely to have “moved up” than workers who were laid off or fired.

3.4 Consequence of Job Loss

The COEP survey has been used to study several facets of the recent changes to the Employment Insurance program. One study examines the impact of eliminating benefits for individuals who quit their jobs voluntarily or are dismissed for cause. Crossley and Kuhn (1998) showed that the proportion of individuals quitting voluntarily who applied for EI dropped by nearly 10 percentage points after the changes to the program were enacted. Three potential reasons are that (1) “quitters” were not applying for EI, (2) they were applying but their claims were being rejected, or (3) the disentanglement was acting as an incentive to find another job more quickly. They showed that the first of these effects accounted for about three-quarters of the total decline in claims. However,

the study also showed an increase in Income Assistance take-up by these workers, demonstrating ‘measurable interactions’ between the EI and Income Assistance systems.

COEP’s design has also made it possible to assess the impact of a reduction in EI benefits (an action taken at the individual worker level) on the economic well-being of the family. In principle, the effect may be muted by income from other sources and by changes in the labour market activities of other family members, possibly triggered by the job loss. Browning (1998) shows that EI is the major source of income for those who receive it. Thus, household consumption does decline as a result of a drop in the amount of the benefit payment. Repeat users suffer a greater drop in consumption than do those who do not claim EI regularly. A repeat user is defined here as someone with at least three claims in the past five years.

3.5 Impact of work interruptions on women’s earnings

Sources like SLID and LAD offer some promise of understanding the long-term effects on women’s earnings of extended work interruptions to raise children. So do some retrospective surveys with information on fertility, child rearing patterns and work history.

Phipps used retrospective data from the General Social Survey to look at the impact of interruptions of six or more months. The average duration of such absences among women aged 25 to 54 and working at the time of the survey (1994) was 4.4 years. (Phipps reports that in the United States, it is 7.5 years, based on a study by Jacobsen and Levin using from 1984 SIPP. The difference is thought to be attributable to the availability of maternity benefits in Canada, which encourage women to take a brief absence rather than withdraw altogether from the labour market.)

The study showed that women returning to a different job after an extended absence paid a penalty in terms of lost wages, although this was not so for women returning to the same job. This penalty for those entering a new job is not simply “delayed movement along the same path as anyone else”. In other words, it is greater than what would be expected by taking into account the reduction in lifetime work experience; the profile of earnings according to work experience shifts downwards. Thus the results would offer support to a program of maternity benefits that results in an early return to the same job.

A second study, using the retrospective work history questions in SLID examines current wage rates according to whether or not employment has been intermittent from the date the first full-time job was held after completion of full-time studies. Simpson (1997) defined labour market intermittence as not working full-time, that is, spells of not working or working part-time. According to this definition, men had on average spent 81% of their available years in full-time, full-year work, compared with 52% for women and 45% for married women with children. This study also showed large differences in wage rates between workers with intermittent work experience and those with continuous full-time, full-year experience.

3.6 Persistence of low income and dependence on Income Assistance

Low income is commonly measured in Canada using one of two definitions. The best known is the Low Income Cutoff or LICO. This measure begins with average household spending on food, shelter and clothing, which stood at about 35% of pre-tax income in 1992. Then 20 percentage points are added to this number. The LICO is the income level at which a family will spend, on average, 55% of its pre-tax income on food, shelter and clothing. LICOs are calculated by family size (ranging from one person to seven or more) and for 5 different community sizes (ranging from rural areas to cities of 500,000 or more). The result is a matrix of 35 LICOs. A second set of LICOs are calculated on the basis of after-tax income.

The base proportion of spending on food, shelter and clothing is periodically updated to reflect changes in average spending patterns.

The second measure is called the Low Income Measure or LIM. This is 50% of median after-tax family income, adjusted for family size and composition. The equivalence scale used for this adjustment assigns a weight of 1 to the first family member, or to a non-family person. The second person receives a weight of 0.4, regardless of age. The third and subsequent family members receive a weight of 0.4 (for adults) or 0.3 (for children).

Low income studies in Canada tend to be based on one of these two measures. When the results of SLID's second wave were released, flows into and out of low income were highlighted, using the LICO measure described above. Based on this measure, 4.4 million people were below the low income line in 1993 or 1994; of this total, 2.3 million were below in both years (*Noreau et al, 1997*). The income changes involved were often large. Persons who dropped below the line in 1994 had a median drop in family income of \$8200. For those who moved above the line, the median increase was \$6200. These are substantial amounts, suggesting that major events had triggered the change in status.

The study by Picot and Zyblock, cited earlier, examined the factors underlying these substantial flows. However, studies of low income which look at a short period, such as the flows between two years, can give an exaggerated impression of the amount of turnover that occurs in the low income population. Laroche (*1997*) studied the persistence of low income spells using the LAD file. She showed that the probability of leaving low income declines as the duration of the low income spell increases. Interestingly, she found little evidence of repeated movements into and out of low income.

As noted earlier, the Self-Sufficiency Project is testing the effectiveness of an earnings supplement for single parents on Income Assistance. In order to receive

the supplement, the person is required to find a full-time job, that is, at least 30 hours per week, within one year. The supplement, available for a maximum of three years, effectively doubles the earnings of individuals receiving minimum wage

The effectiveness of this program in reducing dependence on IA will require a long-term view, but some short-term effects have been reported. Fifteen months into the program, about 25% of the program group members were working full-time and receiving the supplement (*Card and Robins, 1996*). The employment increase was roughly 50% greater than it was in the control group over the same period.

To assess the true impact of the program, employment levels in the program group can be compared with those in the control group. It seems that three people out of five in the program group who obtained jobs did so because of the incentive while two out of five would have found jobs even without the incentive. Ultimately, this project may lead to the accurate targeting of social programs designed to reduce dependence on Income Assistance.

4. LESSONS LEARNED IN LAUNCHING PANEL SURVEYS

There is clearly a great deal of promising research underway, which is encouraging given the immense effort that has gone into the design and early development of the data sources in question. This section looks at a few of the issues faced in the design and development of longitudinal data sources. It is based in large measure on the experience gained in the development of the Survey of Labour and Income Dynamics, which has documented both its successful and less successful innovations.

4.1 Design issues

Omnibus or focused: The results presented above are derived from two different types of longitudinal sources. SLID and LAD are designed for a broad range of research purposes. SSP and COEP are focused on a specific policy concern. A narrow focus can help to control survey development costs and deliver meaningful results more quickly. However, once an omnibus survey is up and running, the analytical potential is broader.

Continuous measures versus sequence of snapshots: There are basically two ways of capturing the time dimension in a panel survey. The first is to link up the data from one interview to the next to create continuous measures or “vectors” of activity. The second is to construct a series of discrete snapshots of the individual at successive points in time. Researchers in the labour and income dynamics are often very interested in spell data -- unemployment spells, spells of welfare receipt, vectors of job attachment, and so on. These are essentially continuous measures. SLID collects the start and end dates of various activities (for example, job holding), and derives measures of spell duration in the most appropriate time unit, usually weeks or months. Spells may of course cross from one reference year into the next, so that right-censored spells in one year need to be “attached” to left-censored spells in the following year. This is a time-consuming task.

However, many variables in SLID are annual snapshots – either a snapshot at the end of the reference year (such as household composition on December 31, or main job held on December 31), or a summary measure for the year as a whole (such as total hours worked during the year).

At this point in the survey’s life, snapshot variables look more attractive than spells, for two reasons. First, as a general rule, they are easier to derive. Secondly, they are easier to handle analytically. It is quite possible that the spell data will become more popular as time passes.

Following rules: All household longitudinal surveys collecting household or family information have a set of rules to determine who should be covered and interviewed in the event of a change in the composition of the household. SLID was designed to follow all members of the originally selected household and interview any new people they live with during the life of the panel. Other surveys select one person in the household as “the” longitudinal person, and construct a family or household picture around that one individual through time. The SLID approach has the benefit of being able to trace (for example) what happens to different branches of a family after family dissolution. But nothing comes without its price. Researchers who want to exploit this information will need to apply a great deal of expertise.

Panel length: Surveys like COEP, with a focus on relatively short-term dynamics, can meet their objectives without getting into the complexities associated with extended panels. For an omnibus survey, however, there is a strong case for defining panels of ten years or more. At the moment, SLID’s panels last six years, which is one year longer than planned when the survey was on the drawing board. Many users have asked us to extend the life of the panel, in order to get a more complete picture of work to retirement transitions, low income spells, among other things.

The main arguments in favour of keeping panels short are attrition and response burden. As time goes on, there is a risk that the sample increasingly under-represents individuals who are highly mobile and over-represents intact, stable households. The ability of the survey to keep attrition down to acceptable levels is a function of many things, not the least of which is whether or not the survey’s content is interesting to respondents. In SLID’s case, the labour and income content is not that stimulating, although there are indications that the questions associated with family events can be quite interesting.

Response burden and respondent relations: Each respondent in a panel survey represents a major investment. Each respondent lost part way through the panel is a correspondingly major loss. The most obvious way of reducing these losses is by keeping the interview short and relatively painless. This can be a challenge with some types of questions, such as expenditures and wealth, where the respondent may be asked to check records, recall amounts or estimate amounts. For its income questions, SLID has offered respondents the option of having us access their tax records or complete an interview. Currently, about three quarters chose the “tax route”. Although it is difficult to quantify, this has surely helped us to reduce attrition.

Some respondents are not very interested in survey results, but many are. Panel surveys often feed back results from the survey to respondents and this can help maintain good relations with respondents. Even if respondents do not read the information sent to them, this material is also available to the interviewer who is better able to brief respondents at the time of the interview.

To be successful, a respondent relations program needs to recognize the heterogeneity of the responding population. No single vehicle or incentive will work for all. What seems to work for many panel surveys is to have a unit responsible for respondent relations. The staff get to “know” the respondents and can help to build up a rapport.

4.2 Development of processing system

Working with the time dimension: Longitudinal databases are more complex than cross-sectional ones because they must account for the time dimension. This includes dating and sequencing events, building spells accounting for geographical mobility and for the movements of people into and out of the sample. SLID uses three different types of variables to represent the time dimension. *Spells* are a state (marital state, or an association with an employer, for

example) with a start date, and end date and a duration. *Vectors* are a series of discrete observations for a pre-specified time interval, such as a week or a month. Labour force status is a *weekly vector*, that is, there is an observation recorded for each week an eligible person is in sample and responding. The third way time is represented is by a *transition*, for example a transition from employment into unemployment, from below to above the low income line, or from being married to being separated.

It was a very time-consuming process to translate SLID's content into a set of variables and processes capable of capturing the time dimension and dealing with temporal inconsistencies. This was done through a data modeling exercise. The SLID data model has 44 entities (groups of variables). One important function of the data model was to define four basic types of entities. *Fixed entities* contain variables that do not change over time, for example, date of birth and mother tongue. *Annual entities* have one observation per year, for example investment income and total annual hours worked. Each annual observation is independent. *Cumulative entities* also are annual, but the value of the variables can increase over time. Examples include years of schooling and years of work experience. Finally, *spell entities* include variables where one record corresponds to a spell. This data-modeling exercise occurred after the survey's content had been defined. In hindsight, it would have been useful to have the two proceed in tandem because the structured thinking need to build the database would have been useful in defining the variables in the first place.

Take long-term view: In their review of the SLID experience, Giles and Lafrance (1995) recommended taking a long term view in developing the processing system. This means designing the file structure to hold the maximum data (in SLID's case six years), rather than rebuilding it from year to year. In essence, this means trying to anticipate what the file for the full panel will look like once it is finished. This can be challenging as the experience gained with each wave of the survey is bound to influence views on what the database should

contain. Nonetheless, it is well worth the effort, and will pay dividends by speeding up processing in later waves. In SLID's case, the first wave took about 24 months to process, the second wave took 12 months and the third wave took 9 months.

4.3 Dissemination strategy

Packaging microdata: The most important output of a longitudinal survey is its microdata, that is, its unit records. The types of analyses typically undertaken with panel data imply access to microdata. The full dataset may all be offered in one single microdata file, or there may be several files designed with particular applications in mind. This approach is bound to be inconvenient for some users, who want content from various files. However, a series of smaller files tailored to specific research topics is less overwhelming than one massive file.

Rectangular or relational file: A decision must also be made between a rectangular file and a relational database. In SLID's case, the main database is relational but the public use file has so far been rectangular. This is beginning to pose a problem because of the length of the file. With the second wave, it was necessary to break the content into a "person file" with demographic characteristics, family variables and income and a "job file" with characteristics on each job held over the two years. The file has one record per job; up to six jobs per year are captured. The user can link the person file and the job file through record identification numbers, or the files can be used independently. When we move to a relational public use file, there will be a software interface to allow the user to build rectangular files using subsets of the full dataset.

"Value added" variables: One important consideration is whether to derive variables from the basic survey content that are likely to be useful to analysts, or expect them to derive their own. These derived variables may speed up the work of the analyst but, of course, they add to the processing time and have an impact

on timeliness. SLID added many derived variables. For example, some labour market measures have been derived at the family level, such as the number of workers in the family and the total hours worked by all family members during the year. There are also many “flags” identifying subpopulations of interest, such as people who moved during the year or had a change in their family composition. Many of these derived variables are turning out to be quite useful, others are less so. Initially, there is a certain amount of guesswork involved but, now that the database is established, it is possible to add derived variables on the basis of demonstrated need. For example, there is clearly a requirement for a variable “main job held during the year”. This can be defined in various ways and by adding this variable, the survey staff will in essence be recommending a certain approach, unless the user has his or her own definition in mind.

Although the microdata is the most important product of a longitudinal survey, there is a role for tables as well. Not everyone can invest the time needed to master a longitudinal database, or the analytical techniques. Tables showing transitions from one state to another and other dynamic measures are a way of broadening the base of users that work directly with the data.

4.4 Protecting confidentiality

There are major challenges ahead if we are to maximise the availability of longitudinal data to the research community and continue to protect the confidentiality of respondents.

Panel surveys often have a massive amount of information in a single wave. In many cases, the events that are captured through time add to the risk of disclosure. SLID applies the traditional methods of protecting confidentiality, including suppression, collapsing, top-coding and addition of random noise. Because the database contains so many variables, SLID uses a systematic approach (first

developed for the Census) for detecting unusual combinations of responses that could entail a higher risk of disclosure.

The public use microdata file is thus a subset of the full survey content. To supplement this, some analysis is also done by academics under contract. In these cases, the researcher is sworn in under the Statistics Act and bound by the same provisions as employees. They may then have access to more lightly screened microdata at headquarters or in the Agency's regional offices.

SLID has experimented with "remote access". In these cases, a special file was created to meet the specific needs of a researcher, but the file remains at Statistics Canada. The researcher writes and submits programs via Internet, which we execute the same day. Assuming there is only limited demand, this approach seems to work reasonably well. Some researchers do not like it because the turnaround is too slow. Even for those researchers who can deal with the pace, this does not currently seem viable as a cost-recovery service. Other surveys have experimented with the same approach and the results are broadly similar.

5. CONCLUSION

There has been a recent surge in longitudinal labour and income analysis, exploiting new Canadian data sources. Many of these studies represent early findings; a more complete picture will emerge as the length of the panels increases. Because the data sources are new and complex, these analysts have made a significant contribution towards establishing these surveys in the research community. At the same time, the survey staff are learning about any barriers that might exist to the full exploitation of the data, so they can address them early.

Bibliography

Browning, Martin. **Income and Living Standards During Unemployment**,

Human Resources Development Canada website

(http://www.hrdc-drhc.gc.ca/hrdc/co...rogeval/macro/ui/i222/brf14_e.html)

Card, David and Philip Robins. **Do Financial Incentives Encourage Welfare**

Recipients to Work?, Social Research and Demonstration Corporation, 1996.

Corak, Miles and Andrew Heisz. **How to Get Ahead in Life: Some Correlates of Intergenerational Income Mobility in Canada**, in Corak (1998 b).

Corak, Miles (ed). **Government Finances and Generational Equity**, Statistics Canada No. 68-513, 1998a.

Corak Miles (ed). **Labour Markets, Social Institutions and the Future of Canada's Children**. Statistics Canada No. 89-553-XPB.

Corak, Miles. **Death and Divorce: The Long-term Consequences of Parental Loss**, presented at the Canadian Employment Research Forum conference, Ottawa, May 1998.

Crossley, Tom and Peter Kuhn, **Effects of Bill-113 Disentitlement on UI and Welfare Take-up Rates**, Human Resources Development Canada website

(http://hrdc-drhc.gc.ca/hrdc/co...rogeval/macro/ui/i215/brf17_e.html)

Duncan, Greg (ed). **Years of Poverty, Years of Plenty**, Institute for Social Research, Ann Arbor, Michigan, 1984.

Finnie, Ross and David Gray. **The Dynamics of the Earnings Distribution in Canada: An Econometric Analysis**, presented at the Canadian Employment Research Forum conference, Ottawa, May 1998.

Galarneau, Diane and Jim Sturrock. *Family Income After Separation, Perspectives on Labour and Income*, Statistics Canada No. 75-001-XPE, Summer 1997.

Galarneau, Diane and Jim Sturrock. *Income After Separation: People Without Children, Perspectives on Labour and Income*, Statistics Canada No. 75-001-XPE, Summer 1998.

Giles, Philip and Pierre Lafrance. **Modeling Data Files for Longitudinal Surveys**, SLID Research Paper Series, Statistics Canada, No. 75F0002M, August 1995.

Laroche, Mireille. **The Persistence of Low Income Spells in Canada, 1982-1993**, Department of Finance, Working Paper No. 98-02, September 1997.

Mata, Fernando. **Transmission of Education and Socio-Economic Status: A Look at Immigrants, Visible Minorities and Aboriginals**, SLID Research Paper, Statistics Canada No. M 75F002M, March 1997.

Morrisette, René. *The Declining Labour Market Status of Young Men*, in Corak (1998 b)

Morrisette, René and Marie Drolet. **The Upward Mobility of Low-paid Canadians, 1993-1995**, SLID Research Paper, Statistics Canada No. 75F0002M, June 1998.

Noreau, Nathalie, Philip Giles, Maryanne Webber and Alison Hale. **Crossing the Low Income Line**, SLID Research Paper, Statistics Canada, July 1997.

Phipps, Shelley. **Long-term Consequences of Interruptions to Paid Work**, presented at the Intergenerational Equity in Canada conference, Ottawa, February 1997.

Picot, Garnett and Miles Zyblook. **Why do Children Move into and out of Low Income: Changing Labour Market Conditions or Marriage and Divorce?** Analytical Studies Branch Research Paper Series, Statistics Canada (forthcoming)

Ruggles, Patricia. **Transitions In and Out of Poverty: New Data from the Survey of Income and Program Participation**, Bureau of the Census, Report No. 8716, U.S. Department of Commerce, December 1987.

Simpson, Wayne. **Labour Market Intermittency and Earnings in Canada**. SLID Research Paper, Statistics Canada No. 75F0002M, August 1997.

Statistics Canada. *Life Events: How families Change*, **Labour and Income Dynamics**, Statistics Canada, No. 75-002-XPB, Vol. 5 No. 1, April 1996.

Wu, Zheng and Michael Pollard. **Economic Circumstances and the Stability of Nonmarital Cohabitation**, presented at the Canadian Employment Research Forum conference, Ottawa, May 1998.