Enhancing Research Opportunities on the Returns to Adult Learning with National Survey and Administrative Data Sources

Marc Frenette | Douwere Grekou | Ted Wannell*

April 2014

SRDC · SRSA

Social Research
and Demonstration
Corporation

Société
de recherche
sociale appliquée

Table of contents

Abstract

This paper explores how research on returns to adult learning could be informed by more complete use of the data resources of Statistics Canada and Human Resources and Skills Development Canada (HRSDC). Both survey and administrative sources are covered, as are current and potential linkages among administrative sources. Our review suggests that current data resources, even when linked, will impose some limitations on adult learning research.

Several Statistics Canada surveys have fairly detailed training information and several types of outcomes but policy analysis is limited by their small sample sizes and short (or non-existent) panel lengths. Administrative data resources overcome these limitations but have issues of their own. Statistics Canada administrative data sources are usually rich in outcomes and often have potential comparison groups. However, it is usually not possible to identify individuals who participated in specific adult training programs with Statistics Canada data. In contrast, HRSDC's data holdings have very detailed information on program participation but generally lack useful outcomes and comparison groups. Linking various Statistics Canada data files together would support analysis of adult learning conducted in postsecondary institutions, while Statistics Canada-HRSDC linkages would enable outcome studies of adult learning programs targeted towards EI recipients, youth and the disabled. However, training programs not conducted through HRSDC programs or postsecondary institutions would remain a data gap and constructing effective comparison groups for such linkages would be a challenge due to limited socio-economic information.

Executive Summary

The study of returns to adult learning is important in the context of a changing economy. Faced with industrial restructuring and geographically uneven job growth, governments must determine how to invest scarce public funds in effective training for displaced or low skill workers. A recent review of the literature on returns to adult learning concluded that studies are limited in their scope of outcomes and often do not focus on detailed types of adult learning (Myers et al., 2011). Specifically, most studies use survey data, focus on individual earnings, and usually consider postsecondary education as a single, broad category. The purpose of this paper is to discuss policy issues related to adult learning that could be informed by conducting research on current sources of Statistics Canada and HRSDC data, including surveys, administrative files and technically feasible, linked administrative files.

Our review suggests that current data resources, even when linked, will impose some limitations on adult learning research. Although survey data at Statistics Canada contain a wide variety of useful outcomes and some detailed information on adult training, their structure often does not lend themselves well to outcome analysis. In particular, small samples sizes and short (or non-existent) panel lengths impose the most important constraints.

In general, administrative data sources offer more potential than survey data sets. While both departments contain many pertinent administrative data files, each have their own set of limitations. At Statistics Canada, administrative data are usually rich in outcomes and often have potential comparison groups. These could answer questions about the returns to adult learning that could not be answered by survey data alone. However, the administrative files with long time series of outcome data at Statistics Canada have very limited information on learning activities: typically a single indicator of postsecondary participation.

In contrast, HRSDC's data holdings have very detailed information about adult learning programs and their participants. However, HRSDC's data holdings generally lack useful information on outcomes or comparison groups.

Linking administrative data within Statistics Canada would create the opportunity to examine outcomes related to adult learning taking place in postsecondary institutions. Opportunities also exist within HRSDC to link some limited outcome data to adult learning program data. Research possibilities would expand considerably by way of HRSDC-Statistics Canada linkages, since one could estimate the returns to very specific government initiatives in adult training programs, not just those delivered in postsecondary institutions, on many financial and non-financial outcomes. A case study is provided for one specific research question: understanding retraining behaviour after layoffs by linking the Longitudinal Worker File (LWF) and the Postsecondary Student Information System (PSIS).

Although the data linkages considered in the paper have the potential to enrich research on adult learning outcomes, some issues would remain unresolved. First, adult training that takes place outside of postsecondary institutions or HRSDC administered programs would remain *terra incognito*. For example, employer-provided training and training by private sector providers that is not sponsored by HRSDC is not covered in these databases. Furthermore, the construction of valid comparison groups to

program participants would be a non-trivial exercise due to the general lack of socio-demographic information (such as education and occupation) on tax-based administrative files.

Introduction

The study of returns to adult learning is important in the context of a changing economy. Faced with industrial restructuring and geographically uneven job growth, governments must determine how to invest scarce public funds in effective training for displaced or low skill workers. Such investments could help participants find jobs or be more productive at their work. Productively employed individuals may also benefit in other areas of their lives. For example, adult learners may feel less stress and enjoy better health as their career prospects improve. Firms may benefit from a larger pool of productive workers. And society may also benefit due to reduced demands on income support programs and public health expenditures.

In spite of large public investments in training programs, little is known about their impact on the full range of outcomes for individuals, firms and broader society. A recent review of the literature on returns to adult learning concluded that most studies in the field are limited in their scope of outcomes and often lack information on detailed types of adult learning provided (Myers et al., 2011). Is the relatively narrow scope of most such studies dictated by the data they use or the decisions of researchers?

The purpose of this paper is to explore how certain Canadian data sets could help researchers address the breadth of policy issues related to returns to adult learning. The data review focuses on Statistics Canada and HRSDC data sources.

The first section of the paper develops a typology of the research questions that can be applied to adult learning outcomes. It then outlines specific research questions that fit within the major types and gives examples of government programs to which such questions could be applied.

The second section starts with an overview of the adult training process. The overview is intended to illustrate the necessary characteristics of the data required to comprehensively evaluate the outcomes of adult learning programs. It then briefly discusses the types of data resources held by HRSDC and Statistics Canada, followed by a description of the relevant data holdings of each organization.

The third section describes the potential application of both organizations' data resources to the analysis of adult learning program outcomes. The utility of program, survey and general administrative sources is examined, along with the value added by the linking of data sources, whether already executed or technically feasible. A case study that illustrates the potential of such an approach is also described.

The fourth section identifies knowledge and data gaps that would persist in the presence of such file linkages. The conclusion summarizes the findings of the paper and makes some recommendations on how the identified data gaps might be addressed.

Adult Education and Training Research Questions

HRSDC's primary mission is to help Canadians get the skills they need to find and retain good jobs, thereby contributing to the quality of life for individuals and families. Economic outcomes of individuals and families are thus the primary metric in assessing labour and skills development policies.

HRSDC is also concerned with how well the labour market functions—particularly in the efficient matching of skilled workers to appropriate vacancies in firms. It makes sense, both in theory and practice, that the timely matching of job seekers with employers requiring their skills will contribute to positive outcomes for firms, as well as workers. Thus firm outcomes are another point of interest in policy research.

Economic and sociological research have also shown that having a skilled and engaged population can have positive effects beyond the standard economic impacts of good, stable jobs. Better health, lower rates of anti-social behaviour and higher civic engagement have all been attributed to higher levels of education, which is also related to better, more stable jobs. Other studies have suggested that these benefits extend beyond the individual to neighbourhoods, cities or society in general. So to fully measure the potential benefits of labour and skills development programs, individual and aggregate measures of non-financial outcomes would also be useful.

Overall then, the major lines of inquiry in labour and skills development policy can be characterized as:

- Studies focusing on the financial outcome of program participants;

- Studies that additionally examine non-financial outcomes of program participants;

- Studies that examine the outcomes of employers of program participants or firms that otherwise participate in skills development programs; and

- Studies that attempt to measure plausible societal outcomes of programs on non-participants, geographic areas or society in general.

Some research questions of each type, as well as examples of specific programs, are discussed below.

Financial outcomes of program participants

Many employment and training programs identify specific outcome goals for participants. For example, the Targeted Initiative for Older Workers is intended to help unemployed workers aged 55 to 64 with their return to work. So the primary outcome of interest is whether participants are re-employed within a reasonable timeframe after the completion of their program activities. Since the overall economic situation of participants is also of interest, the earnings in their new job and its relation to previous job earnings would be outcomes worth examining.

Other programs are more general in nature and aim to improve the general economic situation of participants. The Canada Student Loan Program provides financial assistance to encourage full-time enrolment in postsecondary studies. General studies on the labour market outcomes of graduates can provide broad estimates of the individual returns to study. Analyzing graduate data by field of study

and institution can provide useful information to prospective students, as well as to program administrators and planners.

Foundational learning programs provide participants with basic skills – such as literacy or second language proficiency – to which job-specific skills may later be added. Financial outcomes of such programs may take some time to develop as several sequenced foundational programs may be required to ready participants for occupation-specific training. Assessing the outcomes of such programs thus requires data covering long time horizons or a shifting of focus to their immediate outcomes, rather than final outcomes of interest.

Non-financial outcomes of program participants

In addition to the financial outcomes discussed above, other research has noted secondary benefits in health, well-being and reduction of anti-social behaviours. Much of this type of research is focused on youth employment programs, since a growing body of evidence is showing that early interventions can have positive long-term impacts along a number of dimensions. However, programs targeted to young adults or laid-off workers may have similar effects. Thus, long-term studies of learning program participants' health, civic engagement or contacts with the criminal justice system could yield more comprehensive measures of program benefits than studies that focus solely on economic indicators.

Other studies have examined the relationship between adult learning program participation and activities that may benefit others in participants' sphere of influence. For example, participating in a job-related training or educational program may raise participants' awareness of the general benefits of education for their children and the existence of programs that support their educational aspirations. Such reasoning provides the basis for studies that generally look at the relationship between adult training programs and the postsecondary participation rates of their children. Other lines of inquiry could examine more specific outcomes, such as whether increased use of the RESP-CESG program is associated with adult learning programs.

Firm outcomes

Governments fund a number of programs that are intended to help employers develop the skills required of their employees for the successful functioning of their businesses. Examples include apprenticeship programs, sector council development and foreign credential recognition programs. The existence of such programs raises three types of research questions. First, do such programs increase firms' training activities? Second, do the programs help to create a better balance between skill supply and demand – particularly in relation to skill shortages? Third, do training and other adult learning activities contribute to firm performance as measured by sales, profits or market share? The Canadian literature on firm outcomes is much smaller than for individual outcomes due to greater data demands and the difficulty of controlling for the many factors that can affect firm performance.

Societal outcomes

The possibility of beneficial secondary effects for adult training on participants and their families have been discussed above. There remain two additional types of positive societal outcomes that can also be

considered: effects feeding back to government programs and effects on those living in proximity to program participants.

The most obvious type of feedback effect on government programs relates to the primary outcomes of employment and training programs: if they are successful in getting participants into good, stable jobs then there should be less subsequent demand for such programs, as well as reduced demands on the unemployment insurance system. Such effects can be studied by examining the repeat use of the EI system by adult training participants or their use of social assistance.

The second type of feedback effect on government programs relates to the secondary benefits of adult training programs for individuals. If, for example, these programs do indeed have a positive effect on health then they should reduce the demand for health interventions. This sets the stage for research hypotheses that examine the relationship between adult training and the long-term use of the health-care system.

Finally, primary and secondary program benefits at the individual level may plausibly extend to those in proximity to adult training programs through several means of transmission. First, positive benefits of training may be observed or directly communicated to co-workers or neighbours, increasing their likelihood of participating in such training. Second, if participants' pro-social activities, such as volunteering, increase and their anti-social activities, such as crime or vandalism, decrease then they may contribute to virtuous cycles within their communities. In either case, the concentration of adult trainers in small geographic regions can be related to trends in EI use, social assistance, crime and other social indicators in the same regions.

This Section has considered the types of research questions one can pose in relation to the outcomes of adult education and training programs. The next section takes a closer look at the types of data necessary to support the analysis of such questions.

Data for Adult Education Research

This section ultimately describes the data resources available in HRSDC and Statistics Canada that could contribute to adult education research. But before delving into the data sources, it's useful to think about the process of adult education. Understanding this process provides a fuller picture of the demands that will be placed on data in the course of adult education research. Next, the types of data that are generally available for research will be described. Since data linkage provides the means to more fully enhance research opportunities, it will also be discussed. Although this section is primarily about data, rather than methodology, the presence of a comparison group or counterfactual is such an important element in credibly indentifying program effects that it is discussed in a sidebar.

The Adult Training Process

Adult training is a multi-stage process wherein each stage provides potentially useful information for outcome research. A brief description of each stage and how information on the stage can contribute to subsequent analyses follows.

1. Pre-training state / history

    Training is not a random event. Participants enter training programs as a consequence of their current situation and their past experiences. Thinking about a retraining program for laid-off workers, their current situation would include information about them (age, education, skills, duration of unemployment, etc.), while information on previous jobs and earnings would be key historical information of interest to researchers. Foundational training courses often aim to increase participants' language or literacy skills. In such cases, a base line measure of the skill is essential to the evaluation of the program.

2. Program

    The training program can be thought of as a production process. Both the provider and participants contribute inputs to the process required to complete the training event.

    a. Provider inputs

        Provider inputs can vary in both quantity and quality. Quantitative inputs are relatively easy to measure: the hours of training provided and the cost to provide it are the two most obvious examples. The quality of the inputs is more difficult to assess but can be discerned through measures such as the experience and qualifications of instructors or participants' satisfaction with the course.

    b. Participant inputs

        Participants are also contributing their time and money (including foregone earnings or leisure) to the course. Their knowledge and experience may also contribute to the outputs for themselves and other participants. As with providers, quality can also vary among participants along such dimensions as motivation and work effort.

c. Primary outputs

For many programs, the primary output is the completion of the course. For academic courses, a pass/fail barrier may be involved. For skills-specific courses, a test or certificate of competency may be involved. Note that the primary outputs are not the same as the final outcomes – most frequently stable, good jobs – but should contribute to the attainment of those outcomes.

3. Intermediate outcomes

Many programs involve multiple stages of training and development activities to reach the final outcomes intended by the program. For these types of program, enrolment in the next stage of training or accession to the next level of apprenticeship represents intermediate outcomes. Single stage courses can also have intermediate outputs. Consider, for example, a program that teaches job search skills. The implementation of a job search plan may not immediately result in a permanent job offer but it does increase the chances of meeting that long-term goal. Intermediate outcomes may not be particularly onerous to measure but measurement requires that some contact is kept with course participants for at least a short period following their program.

4. Final outcomes

For most employment-related adult learning programs, the desired long-term outcome is usually a stable, well-paying job. For those with significant barriers to employment, the goals may be more modest: assisted employment opportunities or enough work to increase their level of independence. In any case, the most important point to consider is that under some circumstances for some individuals it may take a significant amount of time for final outcomes to be achieved (or not) following a training event. Thus considerable data collection effort (perhaps repeated over time) is required to effectively measure final outcomes.

Types of Data Available for Adult Education Research

There are two primary types of data used in education research: administrative data and survey data. Administrative data is collected for the purpose of administering and monitoring programs while survey data is collected from households and businesses to monitor the broader progress of society and the economy. There are distinctions within each type of data that are relevant to adult education research.

1. Program administrative data

a. Single program data

Program administrative data is generated by the operation of the program under study. Program data is likely to include basic information on inputs and outputs: costs, hours of training, prerequisites, tombstone data on participants, completions, test scores, etc. Some information on intermediate outcomes, such as apprenticeship placements or enrolment in subsequent courses, may also be available.

There is a general movement among governments and training organizations to move their administrative systems towards the concept of performance management systems, which more fully document program processes from relevant client history to final outcomes (often termed strategic outcomes in this context). Performance management systems may ultimately be a boon to adult education research, but the costs and other hurdles to implementing such systems has thus far limited their use in adult learning research.

b. Multiple program data

Administrative data may also be aggregated across programs. Such databases allow ready comparisons among programs or analysis of the results of aggregations of programs. In order to ensure the consistency of data across programs, such databases may not include the breadth or depth of information found in the individual program administrative systems that feed into them.

2. General administrative data

General administrative data refers to data that is not directly related to the program(s) under study. General administrative databases are most frequently used to provide outcome and/or historical information for individuals. General administrative files are usually longitudinal: they contain a time series of records for each individual. For example, tax-based files contain an individual's yearly tax information for each year they file a return. Tax-system databases may also contain information about education or training programs that affect individual or corporate taxes.

3. Household surveys

Household survey is a generic term for information on individuals or families, usually collected through a sampling scheme based on household location or phone number. However, other sampling frames may be used. Of particular interest are graduate surveys, which sample from lists of graduates of education or training organizations.

Household surveys may be cross-sectional – covering a single period, or longitudinal – covering a number of successive periods.

4. Business surveys

Business surveys use comprehensive lists of companies, such as Statistics Canada's Business Register as sampling frames. The Business Register is hierarchical in nature so that different levels of organization within businesses, from individual locations to corporate conglomerates, can be sampled. Individuals with knowledge about the survey's main subject area are typically targeted within sampled units. In some cases, samples of employees within businesses are also surveyed.

Similar to household surveys, business surveys may be cross-sectional or longitudinal.

Data linking

Data linking or matching typically refers to the combining of individual level records from two different data sources to enable analyses that would not be possible with either data source in isolation. The linkage process is either exact – based on a unique matching key, such as Social Insurance Number – or

probabilistic, based on characteristics that are likely to provide a unique match, such as age, sex and address. The primary example would be the linking of training program data to tax data to examine the long-run earnings and employment outcomes of participants.

Not all data linkage is based on the matching of unique records. Many studies also link individual to aggregate level information. For example, program participant information could be linked to neighbourhood level crime or civic engagement indicators to examine secondary outcomes of program participation.

While tools and techniques for data linkage are well developed, the institutional barriers are often non-trivial. Access and use of information about individuals is subject to federal and provincial privacy legislation, as well as the policies and legal obligations of data collection and holding organizations. A full discussion of the confidentiality and legislative aspects is beyond the scope of this paper, but suffice to say that they can often dwarf the technical considerations.

Box 1   In search of the counterfactual – a further demand on data

In order to understand the need for a counterfactual or comparison group it is useful to consider a concrete example. Take four individuals laid off at the same time from the same company. Two, A and B, decide to take a retraining course to improve their chances of finding a new job. Their two former co-workers, X and Y, decide to search for new jobs without any training. A month after the training is complete, A finds a job and B remains unemployed. Coincidentally, X has found a job at about the same time and Y remains unemployed.

To this point, the outcomes are purely anecdotal – nothing can be inferred from so few cases. Start to increase the number of cases available and the instinct to draw inferences becomes stronger. Say that records are available for hundreds of people who took the same training as A and B. If the ratio of As (who found jobs) to Bs (who didn't find jobs) is very high, then the program starts to look quite successful. However, what if someone was also keeping tabs on the Xs and Ys – those who didn't take the training?

If the ratio of Xs to Ys is similar to the ratio of As to Bs, then those who didn't take the training were just as likely to find jobs as those who did. If the decision to train were entirely random, then the positive outcomes for the Xs represent a valid counterfactual and the impact of the program would be the ratio of As to Bs minus the ratio of Xs to Ys. However, the decision to train is typically non-random: people asses their skills with respect to the demands of the labour market and thus self-select into training on a non-random basis. Analysts must control for these selection effects to create valid counterfactuals. Such controls can be based on observable or unobservable characteristics. Methods that rely on observable characteristics place greater demands on the breadth of data available as controls, while those that rely on unobservables typically place greater demands on the longitudinal dimension of the data.

Alternatively the counterfactual for new programs can be constructed through trials that randomly assign individuals to treatment groups that are offered training or control groups that are not offered training. However, such trials remain relatively rare so other sources of counterfactual information are required. Finding appropriate counterfactuals is a preoccupation among policy researchers that may involve program, survey, administrative, aggregate or linked data, as well as literature searches. For a more detailed discussion of counterfactuals and other methodological issues, please see the *Practical Guide* – the 3rd report in the *Adult Learning and Returns to Training Project.*

HRSDC Data Holdings

HRSDC's data holdings consist primarily of program administrative data. As such, these data are typically rich information about adult learning programs and their participants but provide little or no information on outcomes.

Many of HRSDC's data holdings are related to the Employment Insurance program. These databases typically contain information on labour market-related, workplace-related or foundational training. The Benefit and Overpayment (BNOP) and EI Status vector have information on EI-sponsored training that can inform research on labour market-related learning. Training and skill development activities pertaining to employment could also be examined with other employment benefits files like the Employment Benefits and Support Measures (EBSM), the Aboriginal Human Resources Development Agreements (AHRDA), and the Targeted Initiative for Older Worker (TIOW). These files have information on activities that help in obtaining, maintaining or reintegrating into employment, such as labour-market related training, foundational or workplace-related training.

The Labour Market Agreements for Persons with Disabilities (LMAPD) contain data on activities aiming to develop employability skills of job seekers with disabilities. It may therefore inform the analysis of labour-market related and foundational learning.

The Youth Employment Strategy (YES) has activities aiming to develop basic employability skills and work experience that could serve to identify labour market-related, workplace-related and foundational learning.

Finally, the Needs Assessment Reporting System (NARS) has information on the program type and the field of postsecondary study of the population of student aid applicants. As a result, these data sources could be quite valuable in answering detailed questions about the returns to those specific programs. However, HRSDC's data holdings generally lack useful outcomes and data on non-program recipients who could form comparison groups.

Statistics Canada Data Holdings

Statistics Canada collects and maintains both survey and administrative data sources. The main strength of survey data at Statistics Canada is the sheer number of potential outcomes of interest. However, there are a number of surveys that combine outcome information with training and adult education information, although the later is often lacking in specificity.

Business surveys contain outcome information on many firm-level performance measures, such as operating revenues and expenditures and productivity-enhancing behaviour measures, such as innovation and research and development. The Workplace and Employee Survey combines firm-level information with pertinent information about their employees, including: classroom and on-the-job training episodes; information on literacy/numeracy training; leaves taken by employees, volunteer activities, parental involvement in the lives of children, and detailed skill use in jobs and away from work.

Household surveys also contain fairly detailed information on adult training, especially on formal learning and occasionally non-formal learning. However, the quality of the information on adult

learning activities in survey data varies significantly by source. For example, the Access and Support to Education and Training Survey (ASETS) has most types of learning; the Programme for International Assessment of Adult Competencies (PIAAC) has workplace-related training as does the Survey of Labour and Income Dynamics (SLID), which also has information on most types of learning including foundational learning and hours of training. Finally, the National Graduates Survey (NGS) and Follow-up of Graduates (FOG) have information on higher education. In addition, some data sets, including PIACC, SLID, LISA, NGS and FOG, have field of study information. The survey data at Statistics Canada may therefore be suitable to study features (e.g. delivery, fields of study, location) of different learning activities and may be the only source of information on informal or personal/social learning.

At Statistics Canada, administrative data are usually rich in outcomes and have the potential to create comparison groups for many types of studies. The types of outcomes include income, wages, permanent and temporary layoffs, marital status, geographic mobility, justice, and health outcomes.

When they are available, educational activities in Statistics Canada administrative data refer to postsecondary studies. However, the level of detail varies considerably. For example, the presence of tuition credits and education deductions in the Longitudinal Administrative Databank signal that an individual attended a postsecondary program in a particular year. However, we know nothing about the nature of the program, how long it lasted or whether it was successfully completed. Nor does the file contain information on other forms of adult learning activities that do not qualify for tuition credits or education deductions, such as workplace training or foundational learning delivered outside of postsecondary institutions. Similarly, other tax-based files contain only a very broad indicator of postsecondary education participation. It is therefore very difficult to identify the type of learning from these data sets.

In contrast, Statistics Canada also maintains several education and training administrative files that are rich in program input and output information, but lacking in the area of outcomes. The Registered Apprenticeship Information System (RAIS) contains information on registration and certification in trade/apprenticeship programs, and can therefore inform on the returns to workplace-related and labour market-related training. The Postsecondary Student Information System (PSIS) contains the most detailed information on postsecondary participation and completion among the entire list of administrative databases. It includes program and course information for all students registered at postsecondary institutions in Canada. The information is very detailed, including flags for co-op, articulated, brokered, and collaborative programs, a continuing education indicator identifying adult learners, the method of course delivery, the location of instruction and, importantly, the program level. The program level includes indicators for apprenticeship and upgrading programs, among others. Upgrading programs largely correspond to foundational learning. Information on foundational learning is not available in any other administrative data set at Statistics Canada. For those reasons, the PSIS file is potentially an invaluable source of information on adult learning.

It is clear from this discussion that most current data sources contain either detailed information on adult learning, like PSIS and RAIS, or relevant financial and non-financial outcomes – such as the health, social assistance, and justice files – but not both. In such cases, record linkages can substantially broaden the scope of research in the area.

Similarly, linking files across departments would considerably expand research possibilities since one could estimate the returns to very specific government initiatives in adult training programs (e.g., foundational and labour market-related learning) on many financial and non-financial outcomes. The research possibilities associated with data linkages within Statistics Canada and between Statistics Canada and HRSDC sources are considered in more detail later in the paper.

Applying data to research questions

Statistics Canada survey data[1]

The main strength of survey data is the number of potential outcomes of interest and the extent of other socio-demographic variables that can be used as controls. Moreover, all of the surveys discussed below contain data on the occupation of respondents, a key item that is usually not available in administrative data sets.

In general, there are more cross-sectional than longitudinal files with focused content on adult learning outcomes. Of particular interest are the Access and Support to Education and Training Survey (ASETS) and the Programme for International Assessment of Adult Competencies, (PIAAC). In addition to standard economic outcomes, ASETS includes non-economic outcomes such as volunteer activities and parental involvement in the lives of children. PIAAC's greatest strength is the information on skills used both on the job and away from work. The main issue with these cross-sectional surveys is that outcome measures can only be observed after training has taken place, not before. This means that one cannot associate training activities with changes in outcomes, only post-training outcomes.

Longitudinal surveys provide the opportunity for improved measurement of outcomes, but often do not have the specificity of training information available on cross-sectional files. The recently cancelled Survey of Labour and Income Dynamics (SLID) consists of a six-year longitudinal panel. Changes in outcomes before and after program participation may be observed with this file, but the six-year window is often not long enough to precisely measure outcome trends before and after training episodes.

Researchers will often estimate difference-in-differences models with longitudinal data, but, without sufficient data prior to training, there is no choice but to assume common trends in outcomes among individuals who trained and those who did not. In reality, some individuals may have chosen to train because their earnings had been declining. Only with several years of pre-training data could one rule out such a source of bias.

The new Longitudinal International Study of Adults (LISA) offers some hope in this regard. In principle, the duration of this panel is only limited by funding, which is determined on a year-to-year basis. However, the LISA is limited in its details on training information. The field of study of the current educational program is collected at a very broad level.

More generally, the quality of the adult learning activities in survey data varies significantly by source. Most surveys have information on current or recent formal educational activities, while information on short-term workplace or foundational training is only available in some surveys. For example, ASETS has both types of information, SLID has foundational learning (through the detailed Classification of

---

[1]   Appendix Table 1 lists existing survey data sets at Statistics Canada that could inform the research questions mentioned in the preceding section. The table includes the population covered, the age variable (to help identify adults – however defined), adult learning activities, the outcomes, and the length of the longitudinal panel.

Instructional Program – CIP – code) and hours of training, and PIAAC has workplace training information.

The recently cancelled National Graduates Survey (NGS) focused more on formal training. It collected detailed information on all types of training leading to a certificate, diploma or degree in postsecondary institutions. Moreover, it contains information on graduates' economic outcomes up to five years after the completion of a program as well as retrospective information on their labour market experience prior to the program so it could support longitudinal analysis, albeit with some limitations.

The Workplace and Employee Survey (WES) is another unique source of training information. Conducted from 1999 to 2007, it was a business survey that also collected information from a sample of the businesses' employees. It contains many firm level performance measures, such as operating revenues and expenditures and productivity-enhancing behaviour measures, such as innovation, research and development. The WES has information on workplace training – classroom and on-the-job – and other educational activities related to career advancement, including – for example – trade certification and literacy/numeracy training. Although longitudinal analyses are somewhat limited with the WES due to the maximum seven-year panel for firms and two-year window for employees, its novelty is that it allows employer outcomes to account for employee characteristics activities and vice versa.

The main caveat with survey data in relation to adult learning studies is that, in most case, sample sizes are too small to allow researchers to study the returns to meaningful classifications of adult learning. While sample sizes may be quite large, the proportion of respondents who report training episodes is usually quite small.[2] Once pertinent respondent and program characteristics are considered, it is often difficult to estimate precise estimates of their impact on outcomes. The NGS is a notable exception to this rule but since it only includes graduates in its sample, it does not provide a meaningful counterfactual for the estimation of training effects (see the sidebar *In search of the counterfactual – a further demand on data* in the previous section).

To summarize, survey data are rich in outcomes and socio-demographic controls and often have fairly detailed information on training activities. However, small sample sizes and short (or non-existent) panel lengths impose important constraints when one is attempting to estimate the returns to adult learning.

---

[2]   According to the ASETS survey, only 11% of individuals between the ages of 25 and 54 participated in formal adult learning activities between 2007 and 2008 (Rubenson, 2012).

Statistics Canada administrative data sources

Administrative data sets at Statistics Canada are, in some respects, the antithesis of the survey data holdings.[3] Files that provide economic outcome information generally have long time series and very large samples or nearly complete coverage of the population. Files that cover learning activities also have very good coverage and very specific program information. However, to be informative as a 'stand-alone' administrative data set, a file requires some information on adult learning and some outcome of interest. Only a select number of data sets satisfy these conditions.

First, there are the files that contain personal tax information. The most informative in this group is the Longitudinal Worker File. From the T1 personal income tax file (a component of the LWF), we know if the individual attended a postsecondary institution in a given year from the presence of tuition credits and education deductions. However, we know nothing about the nature of the program, how long it lasted or whether it was successfully completed. The flag also misses all forms of adult learning activities that do not qualify for tuition credits or education deductions: workplace training or foundational learning delivered outside of postsecondary institutions, to name two examples. The LWF contains several outcome measures including employment, temporary and permanent layoffs, wages and salaries and net self-employment income, employer contributions to a pension plan, union membership, marital status, and geographic mobility.

Using the LWF, Frenette et al. (2011) made a first attempt to estimate the role of postsecondary participation following layoff on subsequent earnings. A challenge they faced was the aforementioned lack of detailed data on education – an issue that can be addressed through data linkage (discussed below).

With the exception of layoffs, all of the outcomes in the LWF could also be examined with the other personal income tax files. These files differ slightly in their structure. The T1 covers all individuals who filed their taxes, while the T1 Family File (T1FF) also covers their dependents and allows researchers to group individuals into census families. The Longitudinal Administrative Databank (LAD) is like the T1FF, except that it is a 20% random sample and records are already linked across years. The LAD also contains information from immigrant landing records. The T1 and the T1FF only contain annual cross-sectional information and have no information on immigrants.

Although the income tax files contain only a very broad indicator of postsecondary education, which limits analysis considerably, they are well suited for rigorous econometric evaluations given their longitudinal nature. In difference-in-differences analyses, a treatment group and a control group are ideally followed over a long time period before the treatment (the adult learning activities) to account for differences in trends in the absence of the treatment. A long follow-up period following the treatment is also desirable to examine final outcomes.

---

[3]   Administrative data sets at Statistics Canada are described in Appendix Table 2. Note that the items in the tables that summarize administrative data are slightly different than in those summarizing survey data. In particular, the length of the longitudinal panel is not binding with administrative data since they are collected over several years. Also, administrative record linkage possibilities will be discussed, which requires describing potential linking variables.

A linkage was recently approved between the T1FF, the Registered Apprenticeship Information System (RAIS) and the National Apprenticeship Survey (NAS). The resultant file contains the broad indicator of postsecondary participation from T1FF. In addition, RAIS includes information on registration and certification in a trade/apprenticeship program, and whether that trade is Red Seal certified.[4] It also contains outcomes from T1FF described above. Since the file contains information only on those who have registered in a trade, it can be used to estimate the returns to some types of trades training, but not necessarily the returns to taking a trade. For example, the returns to completing a Red Seal exam can be estimated. One could also estimate the returns to completing a trade program. However, when it comes to estimating the returns to taking a trade, the best evaluation method that can be adopted with these data is a pre-post analysis (i.e. comparing outcomes of individuals before and after taking a trade). Ideally, one would want to adopt a difference-in-differences approach, where the outcome pathways of those who never registered for a trade could be used as a control. However, such a comparison group is not available on the file since it is limited to individuals who registered in a trade.

Potential Statistics Canada record linkages

It is clear from the discussion of the recent T1FF-RAIS-NAS linkage exercise that research opportunities can be bolstered by linking administrative data sources at Statistics Canada. In many cases, current stand-alone data sources either contain detailed information on adult learning – such as PSIS and RAIS, or relevant outcomes – such as health, social assistance, and justice files – but not both. In such cases, record linkages can substantially broaden the scope of research.

The files containing personal income tax information (LWF, T1, T1FF, and LAD) have limited information on PSE participation, but many individual outcomes. It would be technically possible to link the LWF to the T2-LEAP through the business number, which could, in principle, allow researchers to study the role of adult learning in determining firm performance (profits and sales).[5] This would be valuable information from the firm's and society's perspectives since training-related improvements in workers' productivity could also benefit the firm's performance, as well as productivity in the economy as a whole.

The LWF and the remaining personal tax files (T1, T1FF, and LAD) could be linked deterministically to health, social assistance, and justice data through the social insurance number or probabilistically through the date of birth, full name, sex, and postal code. By doing so, many questions of individual and social interest could be answered. For example, one could ask "Does adult learning lead to improved health outcomes?" The answer would be beneficial to individuals considering participation in adult learning, but also to firms who may reap the benefits of improved productivity, and the public, whose tax bill depends in part on the demands placed on the health care system. Many outcomes could be examined with the health files, including longevity and cause of death, experiencing a stillbirth, death

---

[4] A Red Seal endorsement acknowledges competence and ensures recognition of the certificate throughout Canada. See http://www.red-seal.ca/w.2lc.4m.2@-eng.jsp for more details.

[5] A firm-level analysis is preferred to a worker-level analysis since some workers change firms following a training episode and associating the difference in performance of those firms to the training may not be possible.

caused by a fatal injury, and being diagnosed with cancer. Adult learning may exert a positive influence on all of these outcomes by reducing stress or improving self-esteem (e.g. from the feeling of moving on in one's career, or from finding a job).

Similar arguments could be made for the receipt of social assistance benefits[6] and contact with adult criminal courts. If adult learning leads to improved job prospects, we may expect less use of social assistance benefits and less contact with the criminal court system. This would not only benefit the individuals involved, but also society as a whole through reduced public expenditures, reduced criminal victimization and an increased feeling of safety in one's community.

The PSIS contains the most detailed information on postsecondary participation and completion among all the administrative databases discussed: program and course information for all students registered at PSE institutions in Canada. The file includes flags for co-op, articulated, brokered, and collaborative programs, a continuing education indicator, the method of course delivery, the location of instruction, and, importantly, the program level. The program level identifies apprenticeship and upgrading programs, among others. Upgrading programs largely correspond to foundational learning. Very few studies have estimated the returns to foundational learning (Myers et al., 2011). Information on foundational learning is not available in any other administrative data set at Statistics Canada. For those reasons, the PSIS file is potentially an invaluable source of information on adult learning. However, it contains no outcomes, so its usefulness for this literature is contingent on the file being linked.

A PSIS-T1FF record linkage would allow researchers to estimate the returns to very specific types of training delivered in postsecondary institutions. The long timeframe of the T1FF data would allow analysts to carefully account for trends in earnings for years prior to and following the adult learning activity. Specifically, the linkage would allow researchers to analyze the returns to fields of study, which could help determine the relationships between fields with negative returns and an oversupply of students. Such information is useful both to prospective students and policy makers concerned with the efficiency of the postsecondary education market.

The poor economic performance of immigrants in recent decades is another policy issue to which linked data could be applied. Many possible factors have been cited, such as a lack of foreign credential recognition, declining returns to foreign experience, changing source countries, etc. However, it is not clear what can be done to improve the labour market outcomes of immigrants. One possibility is to invest in second language training for recent arrivals. By linking the PSIS to the LAD, it may be possible to estimate the returns to second language training. The LAD includes several pieces of information from immigrant landing records, including an immigrant identifier, the country of origin, the year of arrival to Canada, and the level of education at landing. The foundational learning programs in PSIS include basic literacy, second language training, skills upgrading, and college admission upgrading. Although these are all grouped together in one category, it may be reasonable to assume that the vast majority of well-educated, recent immigrants from non-English and non-French speaking countries

---

6   Social assistance benefits are also available on the personal income tax files, but it is not always
    required to fill out this information.

who enrol in upgrading programs are actually taking second language training since well-educated immigrants would have very little reason to work on their basic skills.

The PSIS and the RAIS could also be linked to the health, social assistance and justice files. Given the detailed program information in the PSIS and the RAIS, this would represent an improvement over a linkage between the personal tax files and health, social assistance, and justice files. Finally, the PSIS could also be linked to the LWF to analyze retraining behaviour after layoffs.

HRSDC administrative data sources

Only a few of HRSDC's administrative files satisfy the conditions of a stand-alone data set that contains information on both adult learning and outcomes of interest.[7] First, there are files that contain information on employment benefits. The most complete file in this group is the Benefit and Overpayment (BNOP) file, followed by the Employment Insurance (EI) status vector, whose parent file is the BNOP. Both files contain information on EI claimants. These files indicate whether an individual received EI-sponsored training. On the outcome side, the claim details and benefits received by the individual are included. For example, one could estimate the returns to EI-sponsored training on repeat EI use.

The Labour Market Agreements for Persons with Disabilities (LMAPD) program is designed to improve the employment situation of persons with disabilities by enhancing their employability and increasing their employment opportunities. The LMAPD file contains information about the training activities received. However, the use of the LMAPD as a stand-alone file is limited since the information on outcomes, such as employment participation and employment opportunities, is collected for participants only. Thus, there is no comparison group in the file.

The Youth Employment Strategy (YES) has three interventions: Skills Link, Career Focus and Summer Work. The strategy targets individuals aged 15-30,[8] particularly those facing barriers to employment. The program intends to help youth get the information and gain the skills, work experience and abilities they need to make a successful transition into the labour market. It includes information about the activities participants undertake to enhance their employability skills and acquire career-related work experience. Once again, even though information about outcomes of interests such as employment search, career choices or the employment status are collected, they are available for participants only.

Training and skill development activities pertaining to employment could also be examined with other employment benefits files like the Employment Benefits and Support Measures (EBSM), the Aboriginal Human Resources Development Agreements (AHRDA) and the Targeted Initiatives for Older Workers (TIOW). The EBSM and AHRDA have similar structures though they target different groups. Both target

---

[7]   Appendix Table 3 lists existing administrative data sets at HRSDC that could inform the literature on the returns to adult learning. As with Appendix Table 2, the table lists the coverage of the data, the age variable, adult learning activities, the outcomes, and potential linking variables.

[8]   We do not clearly define 'adults' in this paper. Some target group members in the YES may be considered adults.

EI- and non-EI claimants and offer multiple interventions. The files contain information about the labour market development activities participants engage in to find and retain jobs. The TIOW targets unemployed older individuals[9] from vulnerable communities. The TIOW file includes information about the activities participants undertake to reintegrate into the labour market. However, as with the LMAPD and YES, the EBSM, the AHRDA and the TIOW have limited value as stand-alone files since they cover only program participants and thus comparison groups are not available.

Some adult learning activities are not fully sponsored by the government. Indeed, adult learners may need to borrow to upgrade their human capital, which could be considered an outcome. The Needs Assessment Reporting System - Disbursement (NARS-DISB) file contains detailed information on student debt, although only among those who applied for government student aid.

Another potential outcome of adult learning is saving for the postsecondary education of one's children. When parents contribute to a Registered Education Savings Plan (RESP), the federal government will match a percentage of their contributions to the RESP with a savings grant. This government top-up is called the Canada Education Savings Grant (CESG). As a stand-alone file, however, the CESG is limited for two reasons: it only contains information on families who have received CESG grants, and it does not contain information on the parents' training activities.

Overall then, most of HRSDC's administrative files are somewhat limited in the application to adult learning outcomes. Most significantly, since most files cover program participants only, they do not have any comparison group. Furthermore, even though these files sometimes include self-assessed evaluation types of questions ("Has this program worked?"), participants may not have good grounds to assess what their outcome might have been in the absence of the program – their individual counterfactual. Therefore, a more rigorous evaluation would require finding valid comparison groups.

Potential HRSDC record linkages

In the case of Statistics Canada, we suggested record linkages to add outcomes or adult learning activities to existing files. This was possible because at least one of the data sets involved was not restricted to adult learners and, in principle, potential comparison groups could be found within that data set.

With HRSDC files, data are usually collected for participants only. Constructing comparison groups can therefore be challenging since individuals who are not included in a given program may not qualify for that program and are thus very different from program participants. Nevertheless, some possibilities exist.

One is to link the EI-status vector file with the files containing information on employment benefits (EBSM, AHRDA and TIOW) or the LMAPD. These files have information on training and skill development activities, but limited information on outcome – with the exception of the LMAPD. Because EI-recipients are eligible for these programs, it would be technically possible to link them to the EI-status vector through the Social Insurance Number. In principle, this could allow researchers to

---

[9] The program targets 55 to 64 year olds and, when there is no risk of displacing the target age group, 50 to 54 year olds and those 65 years old and above.

study the role of training on the dynamics of employment insurance claims and benefits. In the case of the LMAPD, this could provide valuable information about difficulties that are specific to persons with disabilities.

Another issue that could be analyzed through linkages of HRSDC administrative databases is the association between adult learning and savings for the child's postsecondary education. To this end, researchers could link the CESG files to the EBSM, AHRDA, or TIOW. The linkages could be done through the Social Insurance Number.

Potential HRSDC-Statistics Canada data linkages

Collectively, Statistics Canada data are rich in outcomes and potential comparison groups, but they lack information on government-sponsored adult training programs. HRSDC data have detailed information on program participation, but often lack outcomes or comparison groups. By linking Statistics Canada and HRSDC administrative data together, the research possibilities expand considerably. In the end, the linked data sets would have information on the inputs to government-sponsored training programs, many important outcomes, and the potential to construct credible comparison groups.

Specifically, HRSDC program participant data could be linked to the broader Statistics Canada data containing income, social assistance benefits, health, and justice outcomes. This would help researchers address several important questions. A non-exhaustive list of potential research questions together with the required linkage is included in Table 1 in the following section.

Box 2   Case study: Understanding the returns to retraining after layoff

An understanding of the returns to retraining after a layoff is fundamental to the evaluation of many policy interventions set in place or funded by HRSDC. Researchers have not consistently found large effects for government-sponsored job training.[10] A plausible reason for the inconsistency of the effects is the short duration of many such programs. Training that is delivered in post-secondary institutions, and thus captured by PSIS, will often be of longer duration than other government-sponsored training programs. This may be the case even when foundational learning is pursued at post-secondary institutions.[11] To facilitate the study of this issue, the LWF contains several labour market outcome – such as wages and salaries and net self-employment income, employment, permanent and temporary layoffs, employer contributions to pensions, and union membership – and covers a 27-year period. It thus enables researchers to observe labour market outcomes before and after training took place.

Frenette et al. (2011) made a first attempt to study this issue with the LWF alone. They identified post-displacement training through tuition credits and education deductions on the T1 files. These variables have been able to correctly identify the actual person in the household who is taking the training since the late 1990s, shortening the time period available for analysis. Moreover, there is no information on the nature of the training. The data simply infer that the individual was registered in a post-secondary institution for training that was eligible for the credits and deductions, but offers no indication of how long the training lasted or whether it was completed.

By linking the PSIS file to the LWF, the potential for analyzing the issue increases considerably. As mentioned earlier, the PSIS contains detailed course information for all registered postsecondary students, including: course, field of study, program duration and certification obtained. Of particular interest for policy is the fact that we can also identify foundational learning on the file. The file also identifies co-op, articulated, brokered or collaborative continuing education programs, which are likely geared towards adult learners. Further information on the method of course delivery, the location of instruction, and the program level is also available.

Although PSIS only contains SINs for some institutions, linkage would be feasible through exact names, date of birth, and full address.

Further enhancing the usefulness of a linked PSIS-LWF file is the fact that the LWF has recently been linked to the 1991 Census. The Census adds socio-demographic characteristics that could further inform policy on adult retraining programs, including disability, ethnicity, immigrant, and Aboriginal status. It also contains 1991 family level information, such as the number and age of children.

---

[10]   This US evidence generally finds little to no effects (Heckman et al., 1999), while the evidence is mixed in Canada, ranging from no effects for some programs to large effects for others (Park et al., 1996).

[11]   Our typology of foundational learning includes basic education, high school equivalency and language training (see Conte, Myers and Rubenson, 2010). Other types of learning consist of higher education, workplace related learning, labour-related learning and personal/social learning.

Data / Knowledge Gaps

Table 1 provides a summary of the adult learning issues discussed throughout this paper, the data sources that could be linked to address these issues, as well as a brief assessment of the value added by the linkage and the data or knowledge gaps that would persist assuming the linkage took place. Several clear patterns emerge.

As noted above, most of the linkages involved the connection of detailed adult learning program information (inputs and outputs) to long-term economic outcomes, primarily based on tax data. Although these data meet most of the requirements that are necessary to evaluate the final outcomes of adult learning, two distinct shortcomings arise as common themes in the table.

First, the program data maintained by HRSDC and Statistics Canada cover either adult learning activities offered in postsecondary institutions or covered through Employment Insurance. Thus the characteristics and outcomes of adult learning that falls outside of these two domains remain as knowledge gaps in the context of these linkages. Given that HRSDC funds many programs indirectly through transfers to the provinces, this knowledge gap limits the ability to compare the effectiveness of programs and their return on investment to other programs with similar objectives.

Second, the reliance on linked program-tax data impedes the construction of comparison groups or counterfactuals with which to assess outcomes. Although the program files may contain copious data on participants, they of course have no information on comparable non-participants. The linked tax-based files have data on just about everyone, but not in sufficient detail to readily construct groups that are comparable to program participants.

The implications of these themes for data development related to the returns to adult learning are discussed in the conclusion.

Table 1   Research questions and required linkages

Individual Economic Outcomes	Linkage Required	Assessment / Gaps
Returns to very specific types of training delivered in PSE institutions	PSIS with T1FF	Good longitudinal pre- and post-program outcomes, detailed inputs Gap: programs outside of PS institutions, lack of comparison groups
Returns to fields of study	PSIS with tax files (T1, T1FF and LAD)	Good longitudinal pre- and post-program outcomes, detailed inputs Gap: lack of comparison groups
Returns to second language training skills upgrading	PSIS with LAD	Good longitudinal pre- and post-program outcomes, detailed inputs Gap: programs outside of PS institutions
Returns to retraining following layoff	PSIS with LWF	Good longitudinal pre- and post-program outcomes, detailed inputs Gap: programs outside of PS institutions
Role of training on the dynamics of employment insurance claims and benefits	EI-status vector file with employment benefits files (EBSM, AHRDA and TIOW) or the LMAPD	Good longitudinal pre- and post-program outcomes, detailed inputs Gap: lack of comparison groups
Do government-sponsored training programs have long lasting effects on outcomes such as employment, earnings, and low-income?	BNOP or EI status vector with LAD	Good longitudinal pre- and post-program outcomes, detailed inputs Gap: lack of comparison groups
How much student debt is incurred by adult trainers in PSE?	NARS-DISB with PSIS or RAIS	Good input and outcome data Gap: borrowing from non-government sources
Individual / Family Non-economic Outcomes	**Linkage Required**	**Assessment / Gaps**
Relationships between adult learning and savings for the child's postsecondary	CESG files with EBSM, AHRDA, or TIOW	Good input and outcome data Gaps: non-EI training program effects - counterfactual
Do government-sponsored adult training programs foster PSE among the children of participants? If so, what is the role of parental savings for PSE in this relationship?	BNOP/EI status vector and CESG with PSIS	Good input and outcome data Gaps: non-EI training or postsecondary program effects - counterfactual

Individual Non-economic Outcome with Societal outcomes	Linkage Required	Assessment / Gaps
Does government-sponsored training of older unemployed workers in vulnerable communities improve the health of its participants, and thus, reduce the burden on health care costs?	BNOP or EI status vector with health files	Good input and outcome data Gaps: non-EI training program effects - counterfactual
Do government-sponsored training programs aimed at under-employed youth reduce crime?	BNOP or EI status vector with justice files	Good input and outcome data Gaps: potential for counterfactual is not clear
Education and health or justice outcomes	PSIS and/or RAIS with health, social assistance and justice data	Good input and outcome data, longitudinal outcomes Gap: lack of comparison groups
▪ Relationships between adult learning and health outcomes (e.g. longevity, cause of death, cancer, etc) ▪ Relationships between adult learning and justice outcomes (e.g. court appearances, charges, etc) ▪ Social assistance benefits and health or justice outcomes	LWF and/or Personal tax files (T1, T1FF, and LAD) with health, social assistance and justice data	Good input and outcome data, longitudinal outcomes Gap: lack of comparison groups
Employer Outcomes	**Linkage Required**	**Assessment / Gaps**
Role of adult learning in determining firm performance (profits and sales)	LWF with T2-LEAP	Good longitudinal outcome and firm characteristics data Gap: training inputs; non-government supported training

Conclusion

This paper has reviewed the data resources of Statistics Canada and HRSDC with respect to their utility in supporting research on the returns to adult learning. Overall, many opportunities to further research on learning outcomes were noted but some data gaps persist.

Although survey data at Statistics Canada contain a wide variety of useful outcomes and fairly detailed information on adult training, their structure is seldom well-suited to outcome analysis. In particular, small sample sizes for sub-groups of interest, short (or non-existent) panel lengths and lack of detailed training program information are the most frequently encountered constraints.

In general, administrative data sources offer more potential to study final outcomes than existing survey data sets. However, Statistics Canada and HRSDC administrative data sources, as currently structured, have some limitations with respect to adult learning research. By linking files within departments and across departments, the range of policy-relevant questions that could potentially be answered would be substantially broadened.

By linking files at Statistics Canada, it would be possible to estimate the returns to long-term economic outcomes such as employment, earnings and repeat use of EI benefits. Other linkages would bring a broader range of outcomes of interest: firm performance, individual health outcomes, social assistance use, and contact with adult courts. In these analyses, adult learning could include detailed types of programs, such as many types of foundational training, including perhaps second language training for immigrants. A case study highlighted one particular linkage – the PSIS and LWF – that could be used to substantially enhance our knowledge of the returns to adult training following layoff.

The benefits of linking HRSDC files together are perhaps less obvious given both the absence of long-term outcomes and valid comparison groups in most cases. This is because many HRSDC files only contain information on program participants. Nevertheless, some questions could be answered through record linkages. For example, it is possible to estimate the returns to specific EI-sponsored training programs on EI claims and benefits. A related issue that could be examined is the relationship between adult learning and savings for the postsecondary education of one's children.

Greater linkage benefits would come from interdepartmental linkages. Such linkages would allow researchers to estimate the returns to participating in specific, government-sponsored training programs on a variety of economic and non-economic outcomes. HRSDC program participant data have previously been linked to broader administrative data files. For example, this approach was implemented for the summative evaluation of the AHRDA (HRSDC, 2009). However, such activities have largely been limited to income tax data. There are a large number of data files held at Statistics Canada that contain a wealth of non-financial outcomes that could, in principle, be linked to program participation data. Most of these files are longitudinal in nature, which would provide much credibility to the estimates.

The search for better and better data will always hit a wall at some point, mainly due to costs but also due to quality issues associated with the imperative for both broad ranging and in-depth data. The alternative is to use both program knowledge and other sources of data to create reasonable counterfactuals.

Many programs are not distributed evenly across the country and/or were introduced at different times in different regions. These situations offer opportunities for natural experiments. The combination of knowledge about such variations in program delivery and the detailed geographic information available in the tax files gives researchers the opportunity to employ econometric techniques to construct reasonable counterfactuals.

Finally, since considerable public, firm and private resources are being spent on adult learning, some further data development activities could help to ensure reasonable returns to those expenditures.

- Although the shortcomings of sample surveys as a primary source of program evaluation information were noted in this paper, these types of surveys can be a good source of comparisons between adult learners and non-learners due to their inclusion of a wide range of relevant socio-demographic information, such as occupation and educational attainment. Since LISA is the sole longitudinal household survey still being collected by Statistics Canada, HRSDC should ensure that adult learning information adequate to provide general information on the returns to adult learning is collected by the survey.

- Since programs vary across and within provinces and some provinces are already developing learning program performance management systems, data sharing agreements between HRSDC and the provinces could significantly increase the evaluation opportunities among programs with similar goals. Although comparable information on non-participants would remain a gap, this type of sharing can help to identify best practices in the targeting and delivery of programs.

- Further opportunities should be explored to add more socio-demographic information to the tax-based files. As noted in the paper, occasional linkages to the long-form census (now the National Household Survey) can add static socio-demographic information for large, cross-sectional samples at five-year intervals. Another possibility would be to assess the possibility of coding the occupation information collected on the T1 tax form. Having occupation in conjunction with program participation and longitudinal earnings would greatly improve the ability to create meaningful comparison groups of non-participants.

- Since good administrative data is maintained for employer outcomes, some means of linking information on their employees' learning activities is required to support estimates of the returns to training for employer outcomes. Since work on a replacement for the Workplace and Employee Survey was recently cancelled, other options might be considered. Links to program information through the tax files is conceivable, but would only cover a limited range of the adult learning activities of employees. Thus occasional employer surveys of training and related learning activities that could be linked to longitudinal firm outcome data comprise one of the only feasible options.

Note that all of these activities would require considerable resources to create new analytical opportunities. Given current constraints on data collection and linking activities, they should thus not be considered in isolation. Rather, the costs, potential benefits and likelihood of achieving those benefits should be considered across the range of opportunities in order to maximize the return on investment of public funds.

Bibliography

Conte, N., Myers, K., and K. Rubenson. 2010. *Adult Learning Typology.* Social Research and
Demonstration Corporation.

Frenette, M., Upward, R., and P. Wright. 2011. *The Long-term Earnings Impact of Post-secondary
Education Following Job Loss*. Statistics Canada, Catalogue No. 11F00192011334.

Heckman, J., Lalonde, R., and J. Smith. 1999. *The Economics and Econometrics of Active Labor Market
Programs*, in Ashenfelter, O. and Card, D. (eds.), Handbook of Labor Economics, Volume 3A.
Amsterdam: North-Holland, 1865-2097.

HRSDC. 2009. *Summative Evaluation of the Aboriginal Human Resources Development Agreements*.
Evaluation Directorate, Strategic Policy and Research Branch

Myers, K., Frenette, M., Sweetman, A., Dostie, B., and C. Langlois. 2011. *State of Knowledge Review of the
Wider Benefits of Adult Learning*. Social Research and Demonstration Corporation.

Park, N., Power, B., Riddell, C., and G. Wong. 1996. *An assessment of the impact of government-sponsored
training*. Canadian Journal of Economics XXIX, Special Issue.

Rubenson, K. 2012. *Review of the Social Research and Demonstration Corporation's Adult Learning
Typology*. Social Research and Demonstration Corporation.

Appendix

Appendix Table 1   Statistics Canada survey data

Database	Coverage	Age variable	Adult learning activities	Outcome(s)	Length of the longitudinal panel (if applicable)
Access and Support to Education and Training survey (ASETS)	A random sample of about 72,000 Canadians aged less than 65 years old, excluding individuals residing in the three territories, in the North, and in institutions	Age (years)	Recent school enrolment (and level/field of study, including non-formal training and foundational learning) and highest level of education (and field of study)	Income, employment, occupation/industry, volunteer activities, parental involvement and expectations of children, savings for children's education, awareness of government student aid and tax incentives	N/A
Census of Population	A 20% random sample of the Canadian population	Date of birth	School enrolment (and level) in previous 9 months and highest level completed/field of study	Income and components, employment, occupation/industry, unpaid work, geographic mobility, homeownership, mortgage information, dwelling conditions, disability, overall health	N/A
Labour force survey (LFS)	A random sample of about 30,000 dwellings in one of the ten Canadian provinces, excluding institutions and those on reserves	Age (years)	Current school enrolment (and level) and highest level of education	Detailed labour market information (including wages, occupation/industry, job search), geographic mobility	N/A
Longitudinal International Study of Adults (LISA)	A random sample of about 100,000 civilian Canadians 15 years of age and over, excluding individuals residing in institutions and on reserves	Age (years)	Current and recent school enrolment (and level/broad field of study) and highest level of education (and broad field of study)	Survey data will provide detailed labour market information (including wages, occupation/industry), self-rated reading skills, health and disabilities; full administrative income tax data (including at the family level) and pension plan data will also be included	The survey follow-up depends on continued funding (currently determined on a year-to-year basis). To date, one year of data has been collected. The administrative income tax and pension data will be added to

Database	Coverage	Age variable	Adult learning activities	Outcome(s)	Length of the longitudinal panel (if applicable)
					the survey data as long as respondents continue to be informed.
National Graduates Survey (NGS) and Follow-up of Graduates (FOG)	A random sample of about 38,000 recent graduates of Canadian public postsecondary institutions	Age (years)	Level and field of study of the recently completed postsecondary education and further postsecondary enrolment (and level/field of study)	Detailed labour market information (including wages, benefits, occupation/industry, job requirements, job satisfaction) - some of which are available pre- and post-program, geographic mobility (pre- and post-program, including brain drain to the US), student loan repayment	5
Programme for International Assessment of Adult Competencies (PIAAC)	A random sample of about 49,000 adults between the ages of 16 and 65, excluding individuals residing on reserves	Age and month of birth	Current and recent school enrolment (and level/broad field of study, including workplace training) and highest level of education (and field of study)	Detailed labour market information (including wages, occupation/industry), detailed job activities (including skill use), use of skills outside work, self-perception (e.g. work habits), volunteer activities, health and disabilities	N/A
Survey of Labour and Income Dynamics (SLID)	A random sample of about 17,000 Canadian households and all of their members, excluding households residing in the three territories in the North, in institutions, and on reserves	Date of birth	Current school enrolment (and level/field of study, including foundational learning and job-related training, and hours of attendance) and highest level of education (and field of study)	Detailed labour market information (including wages, income and its components, benefits, occupation/industry, reason for job separation), material deprivation, geographic mobility, homeownership, mortgage information, dwelling type, remittances, demographic events, health, stress, disabilities	6
Workplace and Employee Survey	A random sample of about 20,000 paid workers employed in	Year of birth	Detailed workplace training (classroom and on-the-job)	Firm level outcomes (firm size, job vacancies, operating revenues and	2

Database	Coverage	Age variable	Adult learning activities	Outcome(s)	Length of the longitudinal panel (if applicable)
(WES)	businesses operating in one of the ten Canadian provinces, with the exception of employers operating in crop and animal production or in fishing, hunting, and trapping, as well as private households, religious organizations, and public administration		and outside training taken by employees (including literacy and numeracy and hours in class), as well as their highest level of educational attainment	expenditures, unit production costs, productivity, sales growth, product quality, customer satisfaction, profitability, research and development, product development, developing new production/operating techniques, innovation, technology use) and employee outcomes (wages, benefits, hours of work, job duties, job requirements, work arrangements, paid and unpaid leaves taken, technology use, job satisfaction, family income)	

Appendix Table 2   Statistics Canada administrative data

Database	Coverage	Age variable	Adult learning activities	Outcome(s)	Linking variable(s)
Longitudinal Worker File (LWF): T4-ROE-LEAP-T1	A 10% random sample of paid workers who are followed-up indefinitely	Date of birth	Postsecondary participation	Income and components, permanent and temporary layoffs, marital status, geographic mobility, employer contributions to a pension plan, employment, union membership	Social Insurance Number, Business Number, date of birth, full name, sex, postal code
T2-Longitudinal Employment Analysis Program (LEAP)	Population of corporations			Profits, sales	Business Number
T1	Population of tax filers	Date of birth	Postsecondary participation	Income and components, marital status, fertility, geographic mobility, pension savings	Social Insurance Number, date of birth, full name, sex, postal code
T1 Family File (T1FF)	Population of tax filers and their dependents	Date of birth	Postsecondary participation	Income and components, marital status, geographic mobility, pension savings; spousal income and components, and pension savings	Social Insurance Number, date of birth, full name, sex, postal code
Longitudinal Administrative Databank (LAD)	A 20% random sample of tax filers and their dependents	Date of birth	Postsecondary participation	Income and components, marital status, geographic mobility, pension savings	Social Insurance Number, date of birth, full name, sex, postal code
T4	Population of paid workers	Date of birth		Wages and salaries	Social Insurance Number,

Database	Coverage	Age variable	Adult learning activities	Outcome(s)	Linking variable(s)
					date of birth
Record of Employment (ROE)	Population of paid workers experiencing job separation			Layoff	Social Insurance Number, Business Number
Postsecondary Student Information System (PSIS)	Population of postsecondary students	Date of birth	Very detailed program enrolment and graduation information		Social Insurance Number,[a] date of birth, full name, sex, postal code
Registered Apprenticeship Information System (RAIS)	Population of registered apprentices taking in-class and on-the-job training	Date of birth	Trade registration / certification, Red Seal		Social Insurance Number, date of birth, full name, sex, postal code
RAIS-National apprenticeship Survey (NAS)-T1FF	Population of registered apprentices taking in-class and on-the-job training who are tax filers or dependents of tax filers	Date of birth	Trade registration / certification, Red Seal, postsecondary participation, NAS details about the trade program	Income and components, marital status, geographic mobility, pension savings; spousal income and components, and pension savings	Social Insurance Number, date of birth, sex, full name, postal code
Vital Statistics/Death Database	Population of individuals experiencing death	Date of birth		Date and cause of death	Date of birth, full name, sex, postal code

Database	Coverage	Age variable	Adult learning activities	Outcome(s)	Linking variable(s)
Vital Statistics/Birth Database	Population of individuals experiencing birth	Date of birth		Date of birth and birth weight	Date of birth, full name, sex, postal code
Vital Statistics/Stillbirth Database	Population of individuals experiencing stillbirth	Date of birth		Date and birth weight of stillbirth	Date of birth, full name, sex, postal code
Canadian Coroner and Medical Examiner Database (CCMED)	Population of individuals experiencing death due to fatal injury	Date of birth		Date and cause of death due to fatal injury	Date of birth, full name, sex, postal code
Canadian Cancer Registry	Population of individuals diagnosed with cancer	Date of birth		Date and type of cancer diagnosis	Date of birth, full name, sex, postal code
T5007	Population of social assistance recipients	Date of birth		Social assistance benefits	Social Insurance Number, date of birth, full name, sex, postal code
Adult Criminal Courts Survey (ACCS)	Population of adults appearing in a criminal court	Date of birth		Adult criminal court appearances, charges and outcomes	Date of birth, full name, sex, postal code

Note: [a] In the PSIS, the Social Insurance Number is available for some institutions only.

Appendix Table 3 HRSDC administrative data

Database	Coverage	Age Variable	Adult learning activities	Outcome(s)	Linking variable(s)
Benefit and Overpayment (BNOP)	Population of EI claimants	Date of birth	Referred to EI-sponsored training	Employment Insurance claims and benefits	Social Insurance Number, date of birth, economic region
Employment Insurance (EI) Status Vector	Population of EI claimants	Age (years)	EI-sponsored training	Employment Insurance claims and benefits	Social Insurance Number, age (years), sex
Employment Benefits and Support Measures (EBSM)	Population of EI and Non-EI[a] claimants	Age (years)	Activities aimed at obtaining and maintaining employment		Social Insurance Number, age (years,) age category (10 categories), sex
Aboriginal Human Resources Development Agreements (AHRDA)	EI and Non-EI[a] claimants among the Aboriginal people	Age (years)	Activities aimed at obtaining and maintaining employment		Social Insurance Number, age (years,) age category (10 categories), sex
Targeted Initiative for Older Workers (TIOW)	Population of unemployed older workers[b] in vulnerable communities who received TIOW training	Date of birth	Activities aimed at reintegrating older workers into employment		Social Insurance Number, date of birth, full name, sex, street address
Labour Market Agreements for Persons with	Population of job seekers with disabilities	Date of birth	Activities to develop employability skills		Social Insurance Number, date of birth, full name,

Database	Coverage	Age Variable	Adult learning activities	Outcome(s)	Linking variable(s)
Disabilities (LMAPD)c					sex, street address
Youth Employment Strategy (YES)	Population of young people facing barriers to employment, employers and organizations.	Date of birth	Activities to develop basic employability skills and work experience		Social Insurance Number, date of birth, full name, sex, street address
Needs Assessment Reporting System - Disbursement (NARS-DISB) File	Population of student aid applicants	Date of birth	Postsecondary participation	Student debt	Social Insurance Number, date of birth, full name, sex, postal code
Canada Education Savings Grant (CESG)	Population of CESG beneficiaries	Date of birth of primary caregiver		Savings for children's post-secondary	Social Insurance Number, date of birth, full name of primary caregiver

Notes: a Non-EI claimants are those who have no substantive or recent labour force attachment, and include new labour force participants and individuals who were formerly self-employed.

b The target groups are those 55 to 64 years old. Individuals 50 to 54 and 65 plus may be eligible to participate, as long as it does not displace those in the target age range.

c The information for the LMAPD is based on the Canada-Nova Scotia agreement only. Information on other agreements could not be located and may be different.