HOW ADMINISTRATIVE DATA ARE INFORMING RESEARCH ON CANADIANS' EDUCATION

KATHERINE WALL

Katherine Wall is an analyst in the Tourism and Centre for Education Statistics Division, Statistics Canada. She authored the May 2019 StatCan study "Persistence and representation of women in STEM programs" and discussed her research in the panel titled "Disseminating social research at Statistics Canada" at the conference, Statistics Canada – 100 Years and Counting, Ottawa, December 6-7, 2018.

Education in Canada has seen many changes over the years. In 1976, only one-third of young Canadians aged 25 to 34 had completed education above the high school level.[1] Today that proportion has doubled, with over two-thirds of young Canadians having a postsecondary credential (trades, college or university), including more than one-third who have a bachelor's degree or higher.[2]

There have also been major changes in gender representation in postsecondary education. Among the working-age population (age 25 to 64), women have been the majority of bachelor's degree holders in Canada since 1996, and the majority of master's degree holders since 2011. As of 2016, women were the majority of doctorate holders among young people aged 25 to 34 (excluding non-permanent residents).[3]

Likewise, the ways in which Statistics Canada collects information on education have changed. In 2002, the first release of the Youth in Transition Survey (YITS) results provided the first national Canadian data on young people's educational and labour market pathways over time. Such data can provide answers to a wide range of questions. How long does it take students to finish their degrees? How many students drop out, and what kinds of students are most likely to do so? How often do students change their field of study, and what fields

1 Census of Population 1976.

2 Highlight Table 1, Census of Population 2016.

3 Census of Population, 1991, 1996, 2001 and 2016, and 2011 National Household Survey.

of study do they switch to and from? How many students work while they are in school? And how do all these decisions affect their ability to find a job after leaving school, and their earnings? Over the next decade, as YITS participants provided information about their educational and labour market experiences every two years, the survey informed a flurry of papers on topics ranging from postsecondary dropouts to gap years to earnings after graduation.

As the Youth in Transition Survey came to an end, Statistics Canada was in the process of modernizing its approach to use large-scale administrative data to collect much of the information that was formerly gathered through smaller surveys. A new tool for extending the use of administrative data on education is the Education and Labour Market Longitudinal Platform (ELMLP). This platform allows researchers to connect anonymized information provided by apprenticeship programs, public colleges and universities, and tax files, enabling them to analyze students' pathways through the postsecondary education system and how their education and training affects their career prospects in terms of earnings and employment.

Administrative data enables Statistics Canada to understand the year-to-year progress of students' education, and of their jobs and income after graduation, without the need to re-contact survey participants every few years. It also reduces costs and avoids the challenges associated with longitudinal surveys, such as difficulties in contacting respondents who have moved within the last few years, and declining response rates over time.

While the YITS provided very detailed information on a representative sample of students, longitudinal administrative data provide information on all students at public colleges and universities in Canada: millions of students every year. This enables the investigation of topics that require a large population to analyze, such as what kinds of fields of study students switch between. And whereas the YITS provided information on two samples of students — one group who were age 15 in 1999, and one group who were age 18 to 20 in 1999 — administrative data provides yearly information on students of all ages. It can provide information on the number of mature students, and on their earnings before and after completing their new credential. It also allows researchers to look at how educational pathways change over time — for example, whether the length of time it takes students to complete their degrees is increasing.

The last year has seen the publication of the first Statistics Canada findings to use nationwide data integrated through the ELMLP. In particular, a new article released in May 2019, entitled "Persistence and representation of women in STEM programs",[4] endeavours to answer a long-standing question: are women who start a degree in STEM (science, technology, engineering, and mathematics and computer science) less likely to complete it than men are?

The question is pertinent because women have long been underrepresented in STEM fields of study

4 Statistics Canada, "Persistence and representation of women in STEM programs" (Katherine Wall, author), May 2019. www150.statcan.gc.ca/n1/en/pub/75-006-x/2019001/article/00006-eng.pdf?st=5QurJBoQ

and occupations. In 2016, women made up 34% of STEM bachelor's degree holders and 23% of science and technology workers among Canadians aged 25 to 64. Addressing the issue of women's representation in STEM requires identification of the points in women's educational and career pathways when they are lost from STEM. Results from the United States, many looking at a subset of universities or a single institution, have provided conflicting findings on the question of whether women are less likely than men to finish their STEM degrees, and no analysis of the sort has previously been done for Canada.

This article answers the question by using administrative data from all Canadian public colleges and universities, provided through the Postsecondary Student Information System (PSIS) and connected longitudinally by ELMLP, to look at the educational pathways of students who started an undergraduate degree in 2010 at the age of 19 or under.

The article finds that, while women who start a STEM undergraduate degree are less likely to continue in it (66%) than men are (72%), there are also more women than men who move from non-STEM programs (also called BHASE programs, including business, health, humanities, arts, social sciences, and education, among other fields) into STEM ones. As a result, women's representation in STEM remained stable over the course of six years. Within the target population, women made up 44% of first-year STEM students, and 43% of people who had either graduated with a STEM degree after five years or were continuing their sixth year of STEM studies. Thus, women's representation in STEM changes little over the course of a degree. Underrepresentation in STEM is primarily related to women being less likely than men to major in a STEM field when they start their degree, and less likely to work in science and technology occupations after completing a STEM degree.[5]

The population size and level of detail provided by the administrative data enable the research to look not only at STEM overall, but also at specific STEM fields of study. As a result, it is able to show that in engineering and computer and information sciences, fields where women make up less than a quarter of students, women are equally or more likely to remain in STEM than their male peers. Moreover, they tend to graduate more quickly than men in the same fields, completing their degrees in four or five years, while men are more likely than women to take six years or more.

And the longitudinal detail made possible by the ELMLP can go even further than that. Since women who leave STEM degrees mainly switch to non-STEM (BHASE) degrees, while men who leave STEM mainly drop out of undergraduate degree studies, it is valuable to know which fields of study women are switching to. The data shows that women who leave STEM mainly pursue non-STEM (BHASE) fields that are related to their initial STEM field. For example, most women who leave biological sciences switch to health care fields such as registered nursing or pharmacy, which require much of the same knowledge — and which

5 Statistics Canada, "Are young bachelor's degree holders finding jobs that match their studies?", *Census in Brief* (Ottawa: November 2017), Statistics Canada catalogue no. 98-200-X2016025.

may even lead to higher-paying careers than a biology degree. Women who leave mathematics and related studies often switch to fields like business administration, accounting, or finance, which have a strong mathematical component and would allow them to continue applying their mathematical skills and interests.

Moreover, since the ELMLP permits the integration of data on students' education with data on income (using tax variables), it lets researchers see how switching fields of study relates to students' income. Future research could look at whether women and men who switched, for example, from biological sciences to health care, or from math to business, have higher or lower incomes than the students who persisted in those fields.

This article is only an early example of the kind of analysis that is possible with ELMLP-integrated data sources. New projects, such as linking administrative data on students to the census, are already underway. This would enable researchers to carry out new analyses of educational pathways using census variables — for example, parental education, occupation or Aboriginal identity — that are not available from administrative sources. In a context where response rates to many surveys are in decline, administrative data are a key part of the present and future of Statistics Canada research.