CITC

**CANADIAN ISSUES
THÈMES CANADIENS**

100

MADE TO MEASURE:
STATISTICS CANADA @100

RANDY BOSWELL

MICHAEL HAAN

SEN. DONNA DASKO

JAN KESTLE

MARTHA PATTERSON

MYRIAM HAZEL

DYLAN SAUNDERS

KATHERINE WALL

GUSTAVE GOLDMANN

RAYMOND THÉBERGE

ROBERT TALBOT

JACK JEDWAB

MUSAH KHALID

DOUG NORRIS

TABLE OF CONTENTS

LETTERS

Comments on this edition of *Canadian Issues*?
We want to hear from you!

Canadian Issues | ACS
850-1980, rue Sherbrooke Ouest
Montréal, Québec H3H 1E8

Or e-mail us at <miriam.taylor@acs-aec.ca>

@CANADIANSTUDIES

STATCAN'S CENTURY: A NATIONAL INSTITUTION, A FAMILY TRADITIONAND A CHALLENGING FUTURE

RANDY BOSWELL

Randy Boswell is an associate professor at Carleton University's School of Journalism and Communication. He is a long-time Ottawa journalist who developed a unique national history beat as a Postmedia News writer from 2003-13. He has published academic studies about Canadian environmental history and archaeological history, and writes regularly on Canadian history and politics.

In the spring of 1986, for a few weeks before and after the June 3 "Census Day" held across the country that year, I was Statistics Canada's door-knocking, head-counting, form-filling emissary in my home town of Grand Valley, Ontario — population 1,255.

As a 19-year-old census taker, I was also StatCan's mullet-wearing, Duran Duran-loving, Walkman-grooving guy in G.V., roaming the streets (all dozen of them or so) like the Big Shot government agent that I briefly was, clipboard in hand.

By design, the census captured a frozen moment in time in the life of our village in south-central Ontario. Only now, in the afterglow of StatCan's 100th anniversary last year — and at a stage in my life when even a bad haircut would be welcome on this rather barren, 52-year-old scalp — do I realize I was part of history, a bit player in an important national mission to take measure of Canada's citizenry.

Today, we know Grand Valley had 1,255 souls in '86 because I helped count them. My mother had done the same for the 1976 census, when the Valley's population was just 1,096. A few years ago, my daughter — also aged 19, but with an iPod and a hairstyle more likely to stand the test of time — was a StatCan rep for the 2016 census, chasing numbers at select apartment buildings throughout Ottawa (population 934,243 at the time).

This has been a true family affair — three generations of collecting basic data about our fellow Canadians, contributing in our own little way to the information infrastructure that underpins the country's social, cultural and economic policies and programs.

With some measure of pride, I can say we've had a hand in shaping Canada's sense of itself.

One more memory: on Feb. 8, 2012 — more than a

quarter-century after my gig as a census taker back home — I was literally locked up with dozens of other journalists in a StatCan building in Ottawa from midnight to 9 a.m. while swimming in an ocean of fresh statistics about Canada's population gathered during the 2011 census.

The media "lock up" — we were strictly forbidden from communicating beyond the walls of the government office complex overnight — allowed news writers to sift out highlights, localize storylines and get expert input from federal statisticians and demographers in the hours before the census results were officially, widely, publicly released.

On that occasion, reporting for newspapers across the country, I had the privilege of explaining to Canadians some big-picture changes to their nation, including the surging population in Western Canada (for the first time, there were more people living west of Ontario than east) and an overall number of inhabitants that had surpassed 33 million — thanks largely to immigration but also, to some extent, healthier and prolonged lifetimes that meant (among other things) an unprecedented number of Canadians were reaching the age of 100.

(We learned later in 2012, when the next batch of 2011 census stats was released, that the exact number of Canadians over 100 was a staggering 5,825, giving us a pretty impressive army of centenarians that will only grow much, much larger in the coming decades.)

Now Statistics Canada itself has joined those ranks. This edition of *Canadian Issues* was inspired by StatCan's recent 100th anniversary and the December 2018 conference "Statistics Canada: 100 Years and Counting," held in Ottawa and co-sponsored by the Association for Canadian Studies, The Canadian Institute for Identities and Migration, Environics Analytics, the Vanier Institute of the Family, and Statistics Canada itself.

The insights that emerged from that event, a number of which are collected here in essay form, not only remind us of StatCan's indelible imprint on Canada's progress and prosperity since the days of the Dominion Bureau of Statistics, but also point to the tremendous challenges — and opportunities — facing the agency in the age of Big Data, social media and heightened concerns about real and imagined threats to personal privacy.

Our invitation for contributions to this issue of CITC framed the task as follows: "As we reflect on the history of this vital agency, the future of data-gathering and the challenges we face as a nation, what issue (or issues) should most trouble or excite us? How should the work of Statistics Canada evolve? Where should the lines be drawn between privacy rights and the country's need for information to profile its current population and plan for its future?"

We are grateful for the range and depth of responses provided by such a distinguished array of contributors.

Sen. Donna Dasko, one of Canada's best-known pollsters before she accepted an appointment to the Senate in June 2018, provides an excellent overview of the evolving set of challenges facing both survey firms and federal statisticians in tapping truly random samplings of population data in the digital era. She also highlights StatCan's sterling reputation for accuracy and innovation and its "perfect record of protecting the confidentiality of its respondents."

Demographics expert Gustave Goldmann tackles the thorny issue of privacy protection in the world of information gathering and argues the solution to the "Gordian knot" facing public-sector data collectors is to better explain the immense value of their work in sustaining Canada's "national security, health care, social support, public safety, standards for housing, standards for food quality and safety, transportation security, to name but a few."

In his essay, Michael Haan, Canada Research Chair in Migration and Ethnic Relations at Western University, warns about the critical need for social scientists to thoughtfully, continually embrace new methods of data collection and analysis — to confront the challenges posed by Big Data, machine learning and more — to remain relevant in an exciting but uncertain future.

ACS president Jack Jedwab illuminates the emerging problems around the terminology of identification — particularly the mounting challenges to the term "visible minority" as a useful descriptor in gathering population data and constructing public policy. "Identities and identification are in flux — as is the terminology we use to define ourselves," he writes. "In pluralist democracies, citizens define themselves in multiple ways and census takers face considerable challenges in measuring identities in highly charged contexts."

Jan Kestle, the president and founder of Environics Analytics, underscored the importance of building and maintaining trust among Canadians when it comes to data-gathering: "Clearly," she states, "we in the data community are not doing a good enough job at helping Canadians understand why good data are so important to their lives."

In an excellent demonstration of the importance of skilful analysis of health, socio-economic and other information collected in Canada, economist Musah Khalid debunks the notion that this country is suffering the same "deaths of despair" phenomenon widely cited in the United States.

Census expert Doug Norris, chief demographer at Environics Analytics, points to StatCan's recent use of crowd-sourced data on cannabis prices as an example of the kind of innovation in data collection techniques that will be increasingly vital as the federal agency adapts to technological and social change to carry out its mandate in the future.

In their examination of the changing ways in which labour statistics have been gathered and interpreted in Canada, StatCan analysts Martha Patterson, Myriam Hazel and Dylan Saunders explain how traditional tracking of work activity is being augmented by new approaches. "Statistics Canada continues to build on these existing labour market measures to better capture the impact of globalization, technological change and digitalization on the nature and quality of work," they observe. "In 2017, it was one of the first national statistical agencies to release survey-based employment estimates on the 'gig' economy."

Meanwhile, Raymond Théberge and Robert J. Talbot — respectively the Commissioner of Official Languages of Canada and the research manager in the commissioner's office — explain how shifting demographics have long shaped public discussion and policymaking around official languages. They also warn that "the recent rise of populist majoritarian discourse," in Canada and around the world, has "implications for linguistic minorities everywhere."

Finally, StatCan analyst Katherine Wall builds a strong case for the use of administrative data gathered from postsecondary institutions and other sources to shed light on education and employment trends among women in the STEM fields (science, technology, engineering and mathematics and computer science) — an important indicator of ongoing progress in the struggle for gender equality in Canada. She concludes: "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."

Collectively, the essays published in this volume serve as a testament to the fundamental contribution Statistics Canada has made to the country's development over the past century, and emphasize the urgent need for ongoing innovation, adaptation and modernization as Canada's chief data-gathering agency charts the course of its next 100 years.

Perhaps a future grandson or great-granddaughter of mine — whether still knocking on doors to gather bits of information directly from Canadians, or analyzing data streamed automatically and instantly to StatCan via the neural networks of a deeply digitized world — will carry on our family tradition of helping to take the measure of the nation.

SURVIVING THE NEXT AVALANCHE: SKILL DEVELOPMENT AND THE BRAVE NEW WORLD OF ADMINISTRATIVE DATA

MICHAEL HAAN

Michael Haan, PhD, is Associate Professor and Canada Research Chair in Migration and Ethnic Relations at Western University. His interests include migration, immigration and the Canadian labour market. He is also Director of the Collaborative Graduate Specialization in Migration and Ethnic Relations at Western, and Academic Director of Western's Statistics Canada Research Data Centre.

In a 1982 article, Canadian philosopher Ian Hacking documented staggering increases in the amounts of data being collected at the turn of the 19th century, referring to it as an "avalanche of printed numbers." Between roughly 1800 and 1830, he showed, the amount of statistical data increased exponentially, at a time when the written word was only rising in a linear fashion. Undoubtedly, this was a scientific revolution, consistent in nature and magnitude to what Thomas Kuhn described in *The Structure of Scientific Revolutions* (1962). Ways of seeing and thinking about the world changed; it not only became possible to collect and store reams of data, but it became desirable, even essential, for modern and modernizing nation-states to count and measure whatever they could. Consider that the United States (1790), the United Kingdom (1801), and France (1801) all held their first national enumerations around this time.

Unlike other scientific revolutions, however, this one did not initially yield material products. Sure, census results were published in print form, but the outputs were minimal compared to the data being collected. There was nothing like the creation of a printing press or steam engine. All we have from that era are stacks of now dusty books, with countless cross-tabulations. This is because a lot of the tools that we use today to understand statistics didn't yet exist. Although averages existed, medians, standard deviations, variances, correlations and regressions did not; even the word 'statistics' was

not part of the English language until the late 18th century, when it was adapted from the German word *statistik*.

The sheer volume of data being collected, along with a relative absence of techniques for working with these data, created the conditions for the pioneers of statistics to flourish. Brilliant scholars, such as Englishmen Francis Galton and Karl Pearson, developed a statistical infrastructure, much of which remains intact today.

What helped give the 19th century statistical data revolution traction was that it gave us a new way of thinking about the world, a new way of knowing. Gone were the days of indeterminism, as we entered a world of probabilities and likelihoods. Statistics became the ontological engine of the day.

I believe that we are currently at a similar point in history. We now have more data than ever before, and the tools required to analyze these data are either absent entirely, or not commonly used in many disciplines. What this means for social scientists is that we will either need to update our toolkits, or watch as other disciplines speed past us.

THE BIG (OFTEN ADMINISTRATIVE) DATA REVOLUTION

As I argue above, the pace at which data are being collected has once again created a situation in which techniques lag behind data development. The new data regime holds considerable promise, by allowing us to ask and answer questions that were previously impossible. That said, these new opportunities aren't free; big and/or administrative data are generally not collected primarily for research purposes, and as a result are often messy, disparate and poorly documented.

Big data files are not only huge, but also exceedingly complex. The Longitudinal Immigration Database (IMDB), for example, is more than 30 gigabytes once the different tax years have been linked to temporary resident files and permanent resident landing records. Most software packages are simply unable to work with such large files; one does not merely fire up SPSS* and begin running crosstabs on the IMDB. When my students begin working with these files, they often quickly realize that they don't have the skills to work with such data, requiring months of remedial work. This is not a negative reflection on them, but it instead points to how they are not receiving the necessary skills in their coursework.

Users now need to know about complex forms of missing data, probabilistic versus deterministic data linkages, often with less than perfect documentation. Students, at least in my university, don't often receive this training, and are instead forced to learn these things on the job or not at all. Looking forward to the next 100 years, it is difficult to know with certainty what the future will bring, but below are some current trends that I believe will disrupt business as usual for many of us.

ANSWERING RESEARCH QUESTIONS WITHOUT AN EXPLICIT THEORETICAL FRAMEWORK

Big data often allow us to ask questions that we have no theoretical guidance on. For example, what (eerily accurate) shows will Netflix recommend next, when nearly the only thing they know about you is what you've watched in the past? In the old days, when all we had were the techniques

developed during and after the Avalanche, we would have used the age, sex, place of residence, income, and just about anything else we could get about a viewer to develop a predictive framework for suggesting shows.

Now, most of this information is either used tangentially, or not at all. If Netflix's algorithms do a better job at making suggestions than a well-grounded empirical framework (and, anecdotally, at least for me, they seem to), then numerous disciplines need to fundamentally rethink their curriculum. What role does a theory course play in a social science discipline when there is no obvious link between a theoretical framework and the methods that social scientists will use to get answers?

Although this seems counter-intuitive to how social science research is conducted, it can also be seen as an opportunity to think differently about what research is.

WHITHER INFERENTIAL STATISTICS?

It is increasingly likely that future datasets will represent entire populations, not just samples of interest. What this means is that the results we derive will necessarily be an accurate representation of a population, removing the need for the use of inferential statistics. By definition, any trend derived from population data are statistically significant, because they include everyone that we're trying to generalize from.

MACHINE LEARNING

Simply put, machine learning will likely put some of us, or our students, out of work. Machine learning

definitions vary, but they all seem to hold autonomous learning by computers as the centrepiece. This means that humans will no longer need to tell computers what to do, once they've loaded all the necessary algorithms. Reams of data will be analyzed without much input from a human analyst. Without sounding too much like science fiction, computers are beginning to teach themselves, and this will impact how data are collected and analyzed in the future.

There is considerable excitement around machine learning/artificial intelligence, with uses that could range from employing a machine learning algorithm to providing recommendations for law and policy, to finding the best place to live. Judges could make fairer decisions if they are advised on outcomes. Immigrant retention might increase if they can calculate what country they'd be most likely to succeed and be happy in.

COMPETITION FROM OUTSIDE THE SOCIAL SCIENCES

Most of us, as social scientists, have lived fairly peaceful lives within our disciplines. Each of us has a subject area, and everyone else more or less leaves us alone. Individuals? Everybody knows that the study of individuals belongs to psychology. Nations? Political science gets that one. Society? That's sociology; everybody knows that. But what if sociology could do more — take the stage for once?

If you believe that only 0.5 per cent of all data are currently being analyzed (Marr, 2015), and that Google alone handles and saves the parameters of 40,000 searches every second, it would seem that social science analysts will be more outmatched by data than ever before in the big data era. There are

two paths: we can let this opportunity slip away and remain in our individual worlds, or we can engage with this new frontier and use it demonstrate the value of social science.

CONCLUSION

Statistics Canada recently celebrated its 100th birthday, and should be congratulated for its many contributions in the fields of data collection, management and analysis. We in Canada not only have a centenarian statistical agency, but also an internationally recognized one. At the same time, this is hardly the time to pause: private data collection agencies are extremely active, and it's only a matter of time before we render ourselves out of existence. We also have a fairly widespread agreement that statistics are an essential component of modern democracy (just think of the backlash to the Harper government's decision to cancel the 2011 Long Form Census as proof of this). What we now need is a collective increase in the skills required to thrive in this space, to ensure that it and we are also able to survive the next avalanche.

REFERENCES

*SPSS — STATISTICAL PACKAGE FOR THE SOCIAL SCIENCES, a leading software package used for analyzing statistical information.

HACKING, IAN. 1982. "Biopower and the avalanche of printed numbers," *Humanities in Society* 5:279–295.

KUHN, THOMAS. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

MARR, B. (2015). "Big Data: 20 mind-boggling facts everyone must read," *Forbes* magazine. www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#39c4da3a17b1

STATISTICS CANADA. "History of the Census." www150.statcan.gc.ca /n1/edu/power-pouvoir/ch2/history-histoire/5214912-eng.htm. Retrieved May 7, 2019.

STATCAN@100: FROM PAST TO PRESENT TO FUTURE

SEN. DONNA DASKO

Donna Dasko, PhD, is a member of the Senate of Canada and is former Senior Vice-President of Environics Research Group, a leading survey research firm. She is a long-time member of the Statistics Canada Advisory Committee on Social Conditions.

It is an honour to have been asked to reflect on the role and contribution of Statistics Canada (StatCan) on the 100th anniversary of its founding. We live in a time that offers great challenges — and opportunities — when it comes to how surveys are conducted.

As a former survey research professional, I believe that the most difficult issues today involve the sampling process, i.e., how we find people to survey, and whether those people will answer our questions. Developments in this area have changed the entire industry.

We all know that the random sample is the "gold standard" of survey research: if a sample is chosen randomly from a population, that sample will have the characteristics of that population, within a margin of error. When I began my career in the 1980s, we were able to get excellent random samples through random-digit telephone dialing. Almost every household in Canada had a land line, creating potential access to almost the entire population. Furthermore, response rates were high because people were interested in answering our surveys, especially those about public affairs. Conditions for practising our trade were well-nigh ideal, but this golden age was not to last.

The first warning signs came in the 1990s, when response rates started to drop with the growth of recorded messages, call display, fax machines and cell phones. Further declines occurred as the novelty of being wooed by pollsters wore off. I recall response rates of up to 80% in the telephone polls we did in the 1980s. A decade later, they had dropped into the 40- to 50-per-cent range. Today, response rates to traditional phone polls (i.e., with live interviewers) are 10% or lower.

The growth in the use of cell phones created a new

sampling problem. In 2017, according to Statistics Canada, only 67% of households had land lines, with 28% of households using cell phones exclusively. As a result, researchers no longer have the advantage of near-universal coverage from a single source. Granted, cell phone users can be added to the mix, but this increases costs significantly.

The rise of Interactive Voice Response (IVR) and Internet technologies have been a mixed blessing for the industry. IVR involves telephone interviews conducted by computer, using random-digit dialing. While it is dramatically less expensive than traditional interviewing, response rates are absurdly low — in the 2- to 3-per-cent range. Internet polling, as its name suggests, administers the questionnaire over the Internet, typically using panels of respondents assembled through cold calling, banner ads, list purchases from multiple providers, and so on — methods that bear no relationship to random sampling.

Statistics Canada has likewise faced the challenges heretofore described. Although StatCan's response rates have always been higher than those in the private sector, theirs have declined, as well. For example, General Social Survey (GSS) response rates that used to be around 80% now hover between 50 and 60% and, in fact, dropped as low as 38% in the latest (2015) cycle.

StatCan's response to these serious sampling challenges has been extensive. I remember the multi-mode experiments with the GSS in 2013, recruiting respondents by phone and then offering the option of online survey completion. That made good use of Internet technology while maintaining the crucial random sampling. Now, multiple-response channels are offered to respondents: paper, telephone and online. They can be combined to deal with various sampling issues. The record Internet response rate to the 2016 census (68.3%) is a prime example of the success of this strategy.

Statistics Canada now has better and more streamlined communications with respondents, in large part because of the following:

– Its messages are adapted to different segments of the population;

– It uses "nudging," i.e., applying insights from behavioural sciences to encourage more people to respond;

– It has reduced the burdensome household roster;

– It has increased the use of linkages of survey to census and administrative data so that only the information that cannot be found anywhere else is obtained by contacting Canadians. The census serves as a good example: its linkages to tax and immigration data have decreased respondent burden and increased the quality of the information.

These efforts need to be expanded in the future:

– Statistics Canada should continue to multiply response channels as personal computers become less and less relevant. Potential respondents need to be reached on tablets and cell phones with proper applications (apps). Canadians are already using apps to monitor such things as their vital signs, physical exercise, nutrition, time use and budget. StatCan

must develop ways to use such apps.

– StatCan must also find ways to tap into the information that Canadians publicly post on the web. The personal information available on the web presents its own methodological and ethical issues and will require the use of artificial intelligence (AI). StatCan will have to tread carefully, but there are opportunities here.

We cannot turn back the clock on technological change. However, I believe that the survey vehicle will continue to be a vital part of the research process because, thus far, no one has found an adequate replacement.

From my many years on the Advisory Committee on Social Conditions, I have seen — first hand — how Statistics Canada has moved forward on the measurement of social issues and phenomena. It has been a leader in developing new definitions, standards and topics in so many areas; a number of examples come to mind from my time on the Committee:

– In 2000, StatCan embarked on a landmark survey to measure Internet use among Canadians — at a time when nobody knew what we might even call this new thing: Was it the World Wide Web, the Internet, or something else?

– StatCan grappled with the measurement of ethnic ancestry and ethnic identity in the 2002 Ethnic Diversity Survey, in concert with considerable analysis of census data. We have learned that ethnic identity — and Indigenous identity, especially — are fluid and changeable, subject to societal influence and even to the

wording of questions on the census.

– New content accounts for 25% of each cycle of the GSS, and new themes have been launched throughout the years, including the 2016 theme, "Canadians at work and home."

– StatCan led the way in creating significant longitudinal surveys on health, children, youth, labour and income. These were innovative but also challenging, both analytically and in terms of retention of respondents. What happened to those surveys? Can linked administrative data nowadays provide longitudinal views?

– I recall much discussion on the concept of disability, which is very difficult to measure in a reliable and standard way; it can vary considerably with question wording. As a participant in the United Nations *Washington Group* since its inception at the end of the 1990s, StatCan has been a leader in the development of a standard survey instrument to measure disability rate by type and level of severity. In 2017, the Canadian Survey on Disability (CSD) included the full implementation of a new set of Disability Screening Questions (DSQ) used to identify persons with disabilities; this was the culmination of many years of work.

– The measurement of sex and gender is now more complicated. In the old days, sex was the easiest and simplest variable on any survey: there were just two categories. Today, four concepts have emerged: sex, sexual orientation, gender identity and gender expression. With the recent additions of gender identity and gender expression to the Canadian Human

Rights Act and the Criminal Code, as well as some sources of administrative data changing from sex to gender, StatCan has adopted new definitions and standards on the concepts of sex, gender identity and gender expression for use in its surveys. It has also created a new Centre for Gender, Diversity and Inclusion Statistics, which will act as a gender-based analysis data hub to support future, evidence-based policy development and decision-making.

Even the small subset of topics listed above indicates that Statistics Canada has evolved in its research program to reflect significant social change. In my view, it has done well, as indicated by:

- Its solid reputation for the quality of its statistics;

- Its perfect record of protecting the confidentiality of its respondents;

- Its continuous innovation throughout its history;

- The success of the 2016 Census, with its return to the mandatory requirement of the long form (which itself garnered an impressive response rate of 97.8%);

- And changes to *The Statistics* Act to reinforce the independence of the Chief Statistician in determining the most appropriate methodology for sound statistics.

As we move forward, let me suggest some areas that will require special attention:

- There are significant challenges in forging a "Nation-to-Nation" relationship with Indigenous peoples who want to own and control the information collected about them. Can StatCan be a partner that transfers knowledge and expertise to Indigenous organisations? Or are the cultural differences and lack of trust just too difficult to overcome?

- How can StatCan make it easier for the public to understand its data? In today's world of fake news, skepticism vis-à-vis scientific evidence, polarization of society on many issues, and cynicism vis-à-vis the government, can StatCan assist in promoting a more evidence-based environment?

- On a related topic, how can StatCan make its information even more accessible to all? The web site remains somewhat difficult to navigate, and access to micro-data is still cumbersome for researchers.

- How can StatCan help better illuminate environmental and climate change issues? Its project on the Transition to a Low-Carbon Economy is a first step; it will address the need for information on the role of clean technology in Canada as the country develops policy in this area. StatCan needs to track the evolution of this sector, along with its social and economic impacts. Social impact will be a key component of the project because the success of the transition will require the mobilization of society as a whole.

- Can StatCan develop better equity indicators? Particularly with respect to income inequality, to help keep track of the more vulnerable segments of the population?

– And, finally, how can StatCan be more timely? This is a world that does not want to wait long for information.

It is my hope that StatCan will continue to lead the way in this rapidly evolving environment. With sufficient resources to innovate and experiment, I am confident that it will.

WHY MORE AND BETTER DATA
ARE KEY TO AN EVIDENCE-DRIVEN CANADA

JAN KESTLE

Jan Kestle is a President and Founder of Environics Analytics, a leading provider of data and insights to Canadian businesses, governments and not-for-profits. She has worked in the statistics community in Canada for 50 years, is a member of the National Statistics Council and Director of the Board of the Canadian Marketing Association.

When the mandatory long-form census was cancelled and replaced with the voluntary survey in 2011, it was surprising to hear experienced researchers quoted as saying that we have so much big data we don't need a detailed census. The truth is that while many of the new sources can augment our production of data, official reliable statistics are required to build an evidence-driven economy and an effective social policy framework.

With the reinstatement of the long-form census, the dust has settled — at least for now. All agree that the results from the National Household Survey (NHS), which temporarily replaced the long-form census, were questionable at best. As a private company that builds small area data for use by businesses, governments and not-for-profits, we look back at the predictions from our models to see how they compared to what was subsequently measured. Most results from the NHS were not usable; for the data that were of sufficient quality, there were still obvious "breaks" in the series based upon historical review, despite applying best practices.

Hopefully our politicians have learned from this ill-conceived decision. But to ensure it doesn't happen again, the statistical community must convince voters that the census is the cornerstone of Canadian data. As StatCan moves into its second century, two things are essential. First, we must protect and expand the census, the economic accounts, the small area administrative data program and the key survey programs. Second, StatCan has to develop new methods of sourcing and producing data. The current modernization efforts that are underway, which include public-private cooperation, mining

big data and enhanced use of administrative data, are crucial to StatCan's future. But these initiatives will not succeed unless new approaches are combined with strong support for the fundamental programs.

A good example of the challenges and opportunities that StatCan can expect to encounter, as it tests and employs new approaches, emerged recently in another headline story. StatCan was going to conduct a pilot accessing detailed expenditures from personal bank records to test the use of these administrative bank records as a possible source to replace the Survey of Household Spending. Why? Because it's increasingly difficult to get Canadians to fill out extensive surveys, such as this one that is so crucial to the economy.

The plan was to obtain, under The Statistics Act, information about what people spend their money on. From a large roster, a sample would be developed and then the data would be aggregated and used to produce official statistics. A key test in the design of the pilot was to see if the random sample of bank records would be representative enough to replicate the survey results or if the data would be of better quality if they were weighted by the demographics that are typically used in survey design.

StatCan was planning to acquire the spending data along with personal identifiers in order to append select census attributes to the bank records. Those identifiers would be deleted after they were used to match the records. The anonymous records would be processed to produce the statistics. This approach, using appended census data, would be tested against using the same bank spending data without adding the census attributes for sample design and weighting. StatCan had discussed the

approach with the Canadian Bankers Association and the Office of the Privacy Commissioner. The fact that the data would be transferred under the protection of the *Statistics Act* would have ensured that no personal information would be disclosed or stored.

This seemed like a reasonable plan to get good quality data without a significant burden on respondents while ensuring the full protection of confidentiality for the data being used. But it didn't go well. The press and the opposition took great exception to this "spying" on Canadians without their permission, and eventually the pilot was shelved.

It's worth adding that StatCan is well equipped to handle large volumes of data with all the privacy and security provisions that come with them. However, whether you are working with a single record or millions of files, what matters is how those data were collected. We believe that analysts can draw geographically pertinent conclusions without collecting personal information.

WHAT DID WE LEARN?

For citizens to share these kinds of sensitive data with their government, they have to trust their government, which apparently they do not. One critical report cited the possibility that allowing the government to look at the bank records would make it possible to identify individuals who cheat on their taxes. Of course, this would not be possible since no data would be permitted to leave StatCan (i.e. it could not legally be shared with any other government department), but in addition the concern seems a little ironic.

Clearly, we in the data community are not doing a good enough job at helping Canadians understand why good data are so important to their lives — for community safety, compensation and employment policies, as well as research, innovation, health, education and social services planning.

Also, the fact that StatCan has long been rated one of the top statistical agencies in the world and can be trusted with confidential information does not seem to be well known or accepted. When I pointed out to people that the personal information would only be used to append other data and then removed, many people indicated clear disbelief — once StatCan gets the data, they said, "the government (Big Brother) will be watching us."

Lost in the debate was the fact that many countries use these kinds of administrative data sources combined with census and other data as a substitute for surveys. If an independent and world-class organization like StatCan is prohibited from accessing administrative and alternate data sources, our policymakers and business leaders will have to rely on a number of different organizations that have those types of data to tell us their version of what is happening in Canada. Better that these important data be enhanced and integrated into the national statistical framework with oversight from StatCan.

SO WHERE DO WE GO FROM HERE?

A lot of data, from anonymized mobile phone location data to the Internet of Things reports on things like energy consumption, can be leveraged to design, develop and measure government programs. But they need to be normalized, cleaned, weighted and aggregated using more traditional data to ensure accuracy and reliability. Official data can make big data usable and reliable.

The experimenting has to continue to ensure that we not only preserve what we have, but that we fill in the many gaps. Silos across federal departments and between the federal and provincial governments need to be overcome. Producing the volume and variety of data that are needed in our changing world will require significant changes in our approaches.

Citizens certainly have a right to have their personal data protected under the laws that enshrine privacy. But if the data producers can show that not only will the data be protected but will actually be used to make people's lives better, we will be able to move closer to evidence-driven decision making in Canada.

Politics aside, these new approaches must be explored. We are striving to decipher fact from fiction. Statistics Canada is the right organization to lead these continued modernization efforts. Canadians and their government should give the agency a chance to do that and make these next 100 years as successful as the previous century.

SOCIETY AND THE ECONOMY:
THE EVOLUTION OF LABOUR STATISTICS

MARTHA PATTERSON, MYRIAM HAZEL AND DYLAN SAUNDERS

This paper was adapted from presentations made during the panel "100 years of labour statistics" at the conference, Statistics Canada – 100 Years and Counting, Ottawa, December 6-7, 2018.

Martha Patterson is a Senior Analyst in the Labour Statistics Division at Statistics Canada. She holds a Bachelor of Arts in Economics from Trent University and a Master of Arts in Economics from Carleton University.

Myriam Hazel is a Senior Analyst in the Labour Statistics Division at Statistics Canada. She holds Bachelor's and Master's degrees in Economics from the Université du Québec à Montréal.

Dylan Saunders is an Analyst in the Labour Statistics Division at Statistics Canada. He holds a Master of Arts in Public Administration and a Bachelor of Commerce, both from Carleton University.

From the first census in present day Canada in 1666 — an enumeration in New France that identified professions and trades for nearly 800 people, there has been an interest in recording and analyzing the world of work. Today, Statistics Canada continues to foster methods and approaches that respond to and reflect the evolving dynamics of the Canadian labour market.

Labour statistics are somewhat unique in their ability to situate broad social changes in the context of long-term structural shifts in the economy. The Labour Force Survey (LFS), Statistics Canada's largest and longest-running household survey, was launched in 1945 in response to the urgent need for data related to the transition from a war-time to a peace-time economy.[1] Since then, the LFS has been

---

1   Usalcas, Janine, and Mark Kincack. 2017. *History of the Canadian Labour Force Survey, 1945 to 2016*. Statistics Canada catalogue number 75-005-M. Ottawa, Ontario. www150.statcan.gc.ca/n1/pub/75-005-m/75-005-m2016001-eng.htm

an essential source of information on the changing face of Canadian society and the Canadian economy.

SOCIO-DEMOGRAPHIC SHIFTS

One of the most significant changes in the labour force over the last century has been the increasing participation of women, which coincided with other related trends, such as increased participation in higher levels of education, delayed marriage and childbearing, and increased separation and divorce. The 1901 Census of Population recorded that 14% of women were participating in the labour force, compared with 78% of men.[2] By 1946, women's participation had risen to nearly 25%, and continued to increase by another 20 to 25 percentage points over the subsequent 40 years.[3] While the rate for all women has been between 60% and 63% since the early 2000s, the participation of women in the core working ages of 25 to 54 has continued to rise and reached a record level of 83.2% in 2018.[4]

Another key demographic change affecting the size and composition of the labour force is the aging population. Over the last 40 years, people aged 55 and older have accounted for an increasing share of the labour force. This reflects a number of factors, such as the entry of the baby-boom cohort into this age group, higher levels of educational attainment, longer life expectancy and improved health, and is driven by both choice and necessity.[5] Although the participation rate of older people is rising, it remains substantially lower than the rate for younger cohorts. As the population ages and a larger share of people fall into this lower-participation age group, the overall labour force participation rate in Canada continues to decline.

While the aging population puts downward pressure on the labour force, LFS data indicate that employment gains in recent years are predominantly accounted for by immigrants.[6] The LFS first included questions on immigrant status in 2006 and, since then, it has become an important source of information for studying the labour market outcomes of newcomers.

AN EVOLVING ECONOMY

The LFS also provides leading economic indicators, such as the official unemployment rate. With data released every month, just three weeks after the ref-

2   Statistics Canada. 1983. Historical Statistics of Canada. Statistics Canada catalogue number 11-516-X. Ottawa, Ontario. www150.statcan.gc.ca/n1/pub/11-516-x/sectiond/4057750-eng.htm. Table D107-122.

3   *Ibid*. Table D205-222.

4   Labour Force Survey (LFS); Labour force characteristics by sex and detailed age group, annual (14-10-0018-01) www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410001801

5   Fields, Andrew, Sharanjit Uppal and Sébastien LaRochelle-Côté. 2017. *The impact of aging on labour market participation rates*. Statistics Canada catalogue number 75-006-X. Ottawa, Ontario. www150.statcan.gc.ca/n1/pub/75-006-x/2017001/article/14826-eng.htm AND Hazel, Myriam. 2018. *Reasons for working at 60 and beyond*. Statistics Canada catalogue number 71-222-X. Ottawa, Ontario. www150.statcan.gc.ca/n1/pub/71-222-x/71-222-x2018003-eng.htm

6   Patterson, Martha, Myriam Hazel, and Dylan Saunders. 2019. *Annual Review of the Labour Market, 2018*. Statistics Canada catalogue number 75-004-M. Ottawa, Ontario. www150.statcan.gc.ca/n1/pub/75-004-m/75-004-m2019002-eng.htm

erence period, the LFS provides timely information on the health of the Canadian economy, and the impact of disruptive events, such as natural disasters.

This economic side of the LFS is complemented by the Survey of Employment, Payrolls and Hours (SEPH) and the Job Vacancy and Wage Survey (JVWS). These employer-based surveys specialize in providing labour data at the detailed industry and occupation level, respectively.

The modern version of the SEPH was first published in 1983, however, the origin of Canada's payroll employment survey dates back to 1918, when employment data were collected by the Department of Labour. The SEPH is based on administrative data from the Canada Revenue Agency as well as the monthly Business Payrolls Survey. SEPH data are an important input to other processes, including estimating Canada's Gross Domestic Product and calculating provincial equalization payments.

Launched in 2015, the Job Vacancy and Wage Survey is the newest survey on labour statistics. Each quarter, this survey gathers information about job vacancies and their characteristics, including the wage being offered, the level of experience required, and whether they are full- or part-time positions. Data from the JVWS help employers, policy makers, job seekers, and students understand the needs of the labour market in different regions of the country, and in different occupations.

These data sources have captured both short- and long-term industrial shifts in the Canadian economy. For example, SEPH and the previous payroll employment surveys show how Canada's economy transitioned from a predominantly goods-based economy to that of a service-based economy in the late 1950s. Since that time, payroll employment in services-producing industries has continued to rise, while that in goods-producing industries has remained relatively unchanged since the 1960s.

Since the turn of the 21st century, SEPH data illustrate how payroll employment changes from the two largest goods-producing sectors have nearly offset one another, with large declines in manufacturing jobs, and strong gains in construction.[7] Over the same time period, growth in the services-producing sectors was led by health care and social assistance, in keeping with the aging population. According to the JVWS, the health care and social assistance sector was among the sectors with the most job vacancies available, on average, in 2018, behind accommodation and food services and retail trade.[8]

ADAPTING FOR THE FUTURE

The economic and the social domains come together when looking at the changing nature of work in Canada. Advances in technology are disrupting traditional production processes and contributing to an acceleration of globalization and the digitalisation of the economy. Technological progress is

---

7   Survey of Employment, Payrolls and Hours (SEPH); Employment by industry, annual (14-10-0202-01) www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410020201

8   Job Vacancy and Wage Survey (JVWS); Job vacancies, payroll employees, job vacancy rate, and average offered hourly wage by industry sector, quarterly, unadjusted for seasonality (14-10-0326-01). www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410032601. Calculated using the annual average of the four quarters.

also changing the nature of work across industries and occupations by changing the skills employers seek and introducing new ways of exchanging labour. These ongoing changes have led to questions about impacts on the quality of employment experienced by individual workers.

Since the 1990s, the LFS has included content on some of the dimensions affected by these changes, such as temporary employment arrangements and unionization. The JVWS is helping to shed light on changing labour demand, including the level of education sought by employers for current vacancies. In addition, the Census of Population provides information about commuting patterns, which contributes to our understanding of the evolving relationship between labour supply and demand.

Statistics Canada continues to build on these existing labour market measures to better capture the impact of globalization, technological change and digitalization on the nature and quality of work by pursuing a number of initiatives in these broad areas. In 2017, it was one of the first national statistical agencies to release survey-based employment estimates on the "gig" economy, with a focus on peer-to-peer ride services such as Uber and private accommodation services such as Airbnb.[9]

Statistics Canada recently conducted one-time surveys on the reasons for temporary work and self-employment; reasons for continuing to work beyond age 55 (choice versus necessity);[10] and the activities of people, including youth, not participating in the labour market.[11] In addition, the LFS will soon include content on emerging data priorities such as pension and benefit coverage, multiple job holders, and earnings of the self-employed. Data linkages with the LFS are also being piloted (e.g. with tax, Census, and administrative-based immigration data) to derive more detailed indicators on labour market outcomes, and small-area estimation methods are being refined in an effort to meet the needs of users seeking more local data.

Statistics Canada is also involved in discussions with other federal departments on the classification of digital skills. Looking forward, the agency will play a leading role in modifying international standards for measuring forms of employment to ensure that new and changing work arrangements are reflected.

At the same time, Statistics Canada is leveraging the use of new technologies in its own production and dissemination processes. Adapting to technological change is not a new challenge — LFS interviewers started collecting information by telephone in the 1970s, and began recording data on computers in 1994.[12] More recent areas of innovation and

9   Statistics Canada. 2017. "The sharing economy in Canada." The Daily. February 28. Statistics Canada Catalogue no. 11-001-XIE. www150.statcan.gc.ca/n1/daily-quotidien/170228/dq170228b-eng.htm.

10   Hazel, Myriam. 2018. *Reasons for working at 60 and beyond*. Statistics Canada Catalogue no. 71-222-X. Ottawa, Ontario.

11   Bourbeau, Emmanuelle, and Rachelle Pelletier. 2019. *Young people not in employment, education or training: What did they do in the past 12 months?* Statistics Canada Catalogue no. 71-222-X. Ottawa, Ontario. www150.statcan.gc.ca/n1/daily-quotidien/190213/dq190213b-eng.htm

12   Usalcas, Janine, and Mark Kincack. 2017. *History of the Canadian Labour Force Survey, 1945 to 2016*. Statistics Canada catalogue number 75-005-M. Ottawa, Ontario. www150.statcan.gc.ca/n1/pub/75-005-m/75-005-m2016001-eng.htm

exploration include the use of electronic question-naires, the development of interactive visualization tools, an increased presence on social media (e.g. Reddit), and the use of machine learning to assist with rote tasks such as the coding of occupations and industries.

In 2018, Statistics Canada held a year-long cele-bration of the agency's 100 years of existence with high-profile events and activities. While we cele-brate these past achievements, we look forward to ongoing modernization and meeting the challenges of the next 100 years.

The last century of labour statistics has seen the development of new methods, adjustments to address changing priorities, and the adoption of new technologies. What will the next century of work bring?

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.

PRIVACY RIGHTS VERSUS THE NEED FOR INFORMATION — SOCIETY'S GORDIAN KNOT

GUSTAVE GOLDMANN

Dr. Goldmann is a senior researcher and social demographer specializing in the demography of Aboriginal Peoples, the settlement and absorption of immigrants and issues related to ethnic groups and ethnicity. He has published extensively on these topics in Canadian and international journals. After a long career with Statistics Canada he accepted an academic appointment as a Visiting Professor in the Faculty of Social Sciences at the University of Ottawa.

A search through any repository of peer-reviewed journals yields thousands of articles dealing with information collection and privacy. For example, a recent search through the University of Ottawa's holdings using the keywords "information", "individual" and "privacy" resulted in a listing of over 1,200 peer-reviewed articles. Many of the articles focus on the impact of social network sites such as Facebook and on the impact of cloud computing using large databases. Information extraction technology has advanced to the point where it is possible to analyze massive amounts of data in order to extract information that identifies specific patterns and characteristics in a population. It is clear that information is one of the most precious and sought-after resources in most societies. Planners, policy makers, industry organizations, health institutions, educational institutions and countless others are continuously seeking information on the individuals they serve, on the individuals they would like to serve and on the contexts in which they live and function. There are no geopolitical, social or economic boundaries that define or limit the need for information. However, the methods by which the information is obtained varies across all sectors (public and private) of society and they are often not entirely visible to the subject from whom this information is obtained. If we add the question of an individual's rights to privacy and control over their information, we introduce both a paradox and the components of a Gordian Knot — an insolvable problem.

Public sector organizations that collect data on individuals are governed by ethical standards and by the rule of law. Effective, responsive and responsible government requires very detailed information on the social, economic and demographic situation of the people it serves in order to develop, implement and evaluate policies and programs that are the foundation of any society. The collection

and dissemination of this information is very tightly controlled to ensure that the confidentiality and the privacy of respondents, in the broadest sense of the concept, are not compromised in any way. These controls are applied at every stage of the data collection process — from conception to dissemination.

Private sector organizations are not necessarily bound by the same ethical standards and operational constraints as public sector organizations. Their motivation is often based on economic benefit. For instance, they seek to learn as much as they can about their clientele, whether actual or potential, with a view of maximizing profits (which includes minimizing costs). The methods that they use to gather personal information range from formal or informal surveys to collecting data at the point of sale, often without the knowledge of the respondent. Herein lies the fundamental element of the paradox referred to earlier. Consumers willingly provide this information with or without their knowledge because they want to acquire the good or service that is being offered. Approval is implied as soon as they purchase the product or service. Concerns over individual rights to privacy and control over their information are occasionally voiced, but they are rarely acted upon because the need for the product or service trumps the concerns over privacy. In a most draconian sense, consumers are held hostage if they wish to purchase the product or service.

At the present time, concerns over individual privacy are generally voiced only with respect to data and information collected by public organizations. The debate and discussion are often centred on the governments' need for the information and on the potential violation of individual privacy. One

rarely hears a debate or discussion on the need for the product or service that is made possible through the collection of these data. Herein lies a possible solution to undo the Gordian Knot. The products and services provided by the public sector include national security, health care, social support, public safety, standards for housing, standards for food quality and safety, transportation security, to name but a few. The discussion needs to shift from the need to collect the data to the need for the service or product. If the service or product is deemed to be of value and importance to the members of society, the focus of the discussion should be centred on how the data make it possible to provide the service or product. The data are not required if the service or product is not deemed to be of value or importance. Granted, this is a simplification of the data collection process, but it serves to illustrate that it may be possible to untie the Gordian Knot.

How the information is obtained is another important component of this discussion. Obtaining information directly from respondents is but one of the means employed by statistical agencies. They are actively developing and using methods of extracting information from existing sources, such as data collected from the administration of social and economic programs, as a substitute or to supplement data obtained through direct surveys. These data contain information on individuals, on the economic activity of small, medium and large enterprises, on the movement of goods and on the provision of services. Three primary objectives motivate such an approach: to obtain data on a timelier basis, to minimize the burden on respondents and to be more efficient in collecting the required information.

In Canada, the privacy of respondents and the protection of their personal information is governed by legislation such as the Privacy Act, the Statistics Act and the enabling legislation under which many of the federal departments and agencies function. The ways in which their personal information and privacy are protected are not always clear to respondents, thereby opening the door to some of the public debates that we have witnessed over the past decade.

Individuals have the right to ensure that their privacy is not compromised in any way by the organizations that obtain their data, regardless of the means by which the data are collected. Organizations that use the information have the obligation to ensure that they do so responsibly and that they respect the privacy of the individuals. The first point is not in dispute. However, current debates suggest that some doubt exists among individuals concerning the second point. The challenge for all organizations that use the information on individuals is to openly demonstrate how they value and protect individual privacy, thereby building the trust required to resolve the Gordian Knot.

BEYOND THE NUMBERS GAME: OFFICIAL LANGUAGES AND THE CENSUS, PAST AND PRESENT

RAYMOND THÉBERGE AND ROBERT J. TALBOT

Raymond Théberge is the Commissioner of Official Languages of Canada. He has a PhD in linguistics from McGill University, a master's degree in applied linguistics from the University of Ottawa and a bachelor's degree in history from the Collège universitaire de Saint-Boniface. He held a number of leadership positions across Canada, in government and in the scholarly community, prior to coming to the Office of the Commissioner of Official Languages. Commissioner Théberge is the author of peer-reviewed publications, three books, one bibliography and forty-five research projects, which reflect his interest in official language minority communities, including *Demain, la francophonie en milieu minoritaire?*

Robert J. Talbot has been the Research Manager at the Office of the Commissioner of Official Languages of Canada since 2017. From 2013 to 2018, he was Professor of History (part-time) at the University of Ottawa, English Language Secretary of the Canadian Historical Association and Co-Editor of the CHA Bulletin/Intersections. He completed a SSHRC postdoctoral fellowship at the University of New Brunswick and a PhD in History at the University of Ottawa, specializing in Francophone/Anglophone relations in Canada. His book, *Negotiating the Numbered Treaties*, won the Manitoba Historical Society Award for Scholarly History.

The year 2018 marked the 100th anniversary of Statistics Canada. The year 2019 marks another key anniversary: the 50th anniversary of the *Official Languages Act*. The significance of these two anniversaries and the subject areas to which they relate — language and demography — are intimately linked. Indeed, our linguistic regime and Can-adians' conversations around official languages have long been shaped and driven by the numbers, or more precisely, by the shifting demographics.

A quick look back at the past serves to illustrate the point. It also provides a few cautionary lessons for today's scholars and decision-makers as they con-

sider how the evolving demo-linguistic situation should inform current and future language policy. The significance of these lessons becomes all the more apparent when we consider the recent rise of populist majoritarian discourse, both in Canada and around the world, and its implications for linguistic minorities everywhere.

TYING THE RIGHTS OF MINORITIES TO THE SIZE OF MAJORITIES – A "CATCH-22"

The rights of Canada's official language minority communities (English-speaking in Quebec and French-speaking elsewhere) have long been tied to their size relative to the majority. It was no coincidence, for example, that at the time of Confederation, the two provinces with bilingual legislatures and judiciaries, Quebec and Manitoba, were home to the most influential and, in proportionate terms, the largest official language minorities; in Quebec in 1867, the English minority represented around 20% of the population, and in Manitoba in 1870, French and English each represented about half of the population. A century later, following decades of growth of its Francophone minority, New Brunswick would also recognize the equality of English and French at the provincial level.[1]

While the relative growth of a minority can make it more politically possible to attain rights, the opposite can also occur. Looking back through history, we can see how the relative decline of minorities has on occasion contributed to a context in which it became politically possible to *remove* recognition of the minority language. Coupled with a rise in populist majoritarian discourse, the removal of minority rights became not only politically possible, but also politically desirable from the perspective of those seeking to govern. This is what occurred in 1890, when Manitoba adopted its *Official Language Act* (English), and in 1974, when Quebec adopted its own *Official Language Act* (French). Note the use of "language," singular.

Figure 1 captures the dramatic proportional decline of Western Canada's ethnic "French" population, from 1870 to 1921. It bears pointing out, however, that this constituted a *relative* decline, and not an absolute decline. Even though the minority had grown in absolute numbers, it still was deprived of its rights, in part because the majority had grown at a faster rate.

While politicizing the "decline" of a minority can lead to the denial of rights and recognition, politicizing the growth of an official language minority community can also lead to a more charged and contentious discourse around language rights, wherein the majority asserts a threat of being overtaken by the minority in its midst. This is what happened in

---

1   Around 16% of N.B.'s population was Acadian according to the 1871 census. By 1971, it was around 34%. See Joel Belliveau, "L'évolution de la situation démolinguistique des Acadiens du Nouveau-Brunswick: La situation actuelle replacée dans le temps long," presentation delivered to the colloquium *Les enjeux de la migration en contexte acadien et francophone* organized by the Canadian Institute for Research on Linguistic Minorities (Moncton, N.B., August 2010) (https://icrml.ca/fr/site_content/item/38-enjeux-migration-contexte-acadien-et-francophone).

The figures referred to here are in relation to ethnicity (French and British, the latter comprising English, Scottish, Welsh and, more contentiously, Irish — in the case of Manitoba, the 'French' and 'English' populations in the data include French- and English-speaking Métis). This data is used to approximate the French- and English-speaking populations, since data on language is generally not available for the 19th century.

FIGURE 1[2]: PERCENTAGE OF POPULATION OF WESTERN PROVINCES BY ETHNICITY, 1870, 1921

British   French   Others, Uncertain
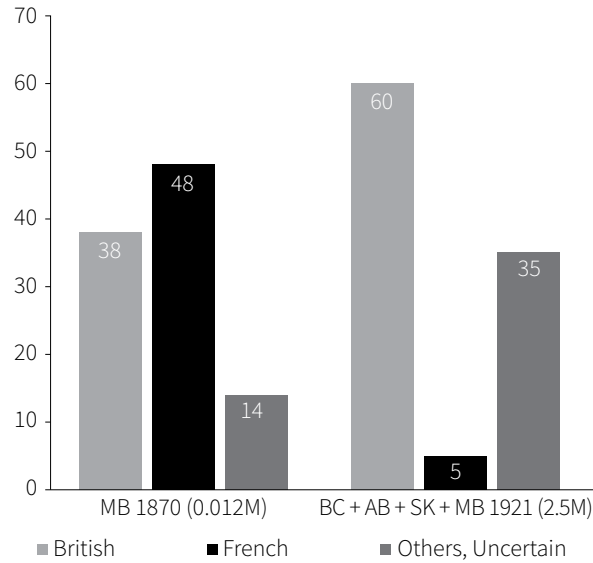
Ontario when, in 1911, the census confirmed the continuing growth of the province's ethnic "French" population and, as we can see in Figure 2, the relative growth of the French population across central and eastern Canada. "The battle waged for a century will have been lost!" warned Toronto Mayor Horatio Hocken. "The barrier that Ontario has for so long

opposed to the oncoming tide of French settlement will have been swept away. All that would mean to the destiny of Canada cannot be readily imagined. It would almost inevitably mean French domination and papal supremacy."[3] One year after the census, in 1912, the provincial government imposed Regulation 17, effectively removing the French-language education rights of the Franco-Ontarian minority.
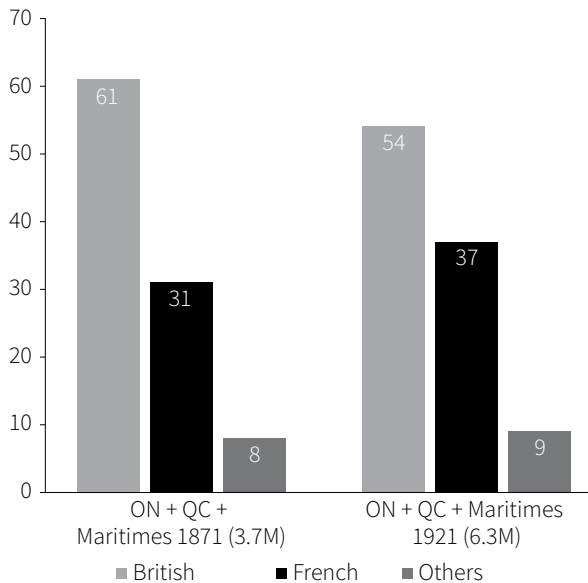
More recently, in August 2017, the release of erroneous 2016 census data on language served to exaggerate the growth of the English-speaking minority in Quebec, leading to gloomy predictions about the future of French in the province. (Fortunately, the hard work of Statistics Canada's staff, and their openness in consulting with the Office of the Commissioner of Official Languages and with other key stakeholders in the days following the initial release, helped to bring about an effective correction.)

If the 2016 census mix-up taught us anything, it is that we must be cautious when making assumptions about how Canada's linguistic landscape will evolve into the future, and we must ensure that our linguistic regime of tomorrow is not bound by the assumptions of today.

This last point is particularly salient as we reflect on the 50th anniversary of Canada's *Official Languages Act*, and as we look forward to the next 50

---

2   Note that Manitoba, the Northwest Territories and British Columbia were not included in the Canadian census of 1871, but a separate census was conducted in Manitoba in 1870. The methodology of that Manitoba census was distinct – the percentages used here for 1870 are thus approximations (we have included the Métis populations of "British" and of "French" origins). The figures for the 1870 Manitoba "Archibald" census are discussed online at Norma Jean Hall, "Provisional Government of Assiniboia" (https://hallnjean2.wordpress.com/the-red-river-resistence/red-river-censuses/). For the figures from 1921, see Canada, *Sixth Census of Canada, 1921*, Vol. I: Population: Number, Sex, and Distribution — Racial Origins — Religion (Ottawa: F.A. Acland, Printer to the King, 1924).

3   In Robert J. Talbot, "Moving Beyond Two Solitudes: Constructing a Dynamic and Unifying Francophone/Anglophone Relationship, 1916-1940," Doctoral dissertation in History (University of Ottawa, 2014), p.68.

FIGURE 2[4]: PERCENTAGE OF POPULATION OF EASTERN PROVINCES BY ETHNICITY, 1870, 1921

the advancement of language rights would help to maintain the vitality of official language minority communities. This, we would argue, has turned out to be the case, although serious challenges and legislative deficiencies remain. Second, it was based on a rather optimistic assumption that the natural increase of the French-speaking population would help it to keep up with the English-speaking population, which benefited significantly from immigration. One researcher for the Royal Commission on Bilingualism and Biculturalism (1963-1969) even speculated that, in New Brunswick, for example, "la minorité sera peut-être un jour la majorité"![5] Meanwhile, statisticians in the 1960s estimated that, by 2020, Canada's Francophone majority province, Quebec, would be home to some seventeen million souls.[6] Something tells us that we're not going to get there by the next census!

Moreover, the problem with the 10% minority benchmark was that the minority's vitality was, once again, being tied to the size of the majority. Even if the minority was a vibrant or a growing community, if it wasn't growing as fast as the majority it risked losing its rights. Recent events in Ontario, where the French-speaking minority is growing but at a slower pace than the majority, provide a case in point. In their rationale for removing the independent Office of the French Language Services Commissioner and for cancelling funding for a French-language university,

years. Among other things, the 1969 Act allowed for the creation of "bilingual districts" across the country, whenever the local minority grew to 10% of the population. As it turns out, these districts were never created. Why? In part, because their creation was predicated on two assumptions — one accurate, the other less so. First, it was based on the hope that

---

4   See Canada, *Census of Canada, 1871*, Vol. I (Ottawa, I.B. Taylor, 1873); Canada, *Sixth Census of Canada, 1921*. Vol. I: Population: Number, Sex, and Distribution - Racial Origins - Religion.

5   Archives Radio-Canada, "Bilinguisme, la position du Nouveau-Brunswick," interview with Premier Louis Robichaud, 25 June 1963 (http://archives.radio-canada.ca/sante/langue_culture/clips/2397/).

6   See House of Commons, Second Session – Twenty-Sixth Parliament, 1964, Special Joint Committee of the Senate and of the House of Commons on old age pensions, Minutes of Proceedings and Evidence, 11 December 1964 (Ottawa: Duhamel, 1964).

Ontario officials sought to minimize the relative size of the Francophone minority, and pointed out that Ontario is home to multiple linguistic minorities.[7] This, of course, implies that different minority groups are to be weighed one against the other — not for the purposes of *granting* rights and recognition to one group or another, but rather to *remove* rights and recognition that had previously been secured. In such a race to the bottom, the only group left benefiting from language rights and recognition will inevitably be the majority itself.

TOWARDS A NEW APPROACH: A NEW OFFICIAL LANGUAGES ACT, RENEWED RESEARCH, AND THE 2021 CENSUS

So, what can we learn from these lessons, past and present? Fifty years of experience tells us that we need a modernized *Official Languages Act* and new regulations that are more robust and not simply tied to cold percentage points, left vulnerable to the constant ebb and flow of population increases and decreases. In order to make this feasible, we must ensure that decision-makers have a more nuanced understanding of community vitality, of community conditions and of communities' needs — all of which researchers can measure through advanced data analyses that link language with the multiplicity of factors that affect and that speak to community vitality, including: youth, gender, mobility, immigrant status, institutional support, employment, income, occupation and education, among others.

The full implementation of the current and future *Official Languages Act* and of the language rights guaranteed under the *Charter of Rights and Freedoms* requires access to complete, accurate and rigorously analyzed data. To that effect, the federal government needs to provide stronger guarantees for research funding and engage in a new post-censal survey on the vitality of official language minority communities. We also need new questions in the census on minority language education, in time for 2021, and not after. Official language minority schools are a critical vehicle for the transmission of language, culture and identity. It is therefore crucial for minority education rights holders to be fully counted in every corner of the country in order to guarantee that parents and guardians can have access to education for their children in the language, as is their right under the Charter.

This would not be the first time that we have added questions to the census out of recognition of the complexity of Canadians' linguistic realities:

- In 1901, questions were added on knowledge of language and on mother tongue, out of recognition that language and ethnicity are not one and the same;

- In 1971, questions on the language(s) spoken in the home were added, out of recognition that the mother tongue is not necessarily the language of use; and

- In 2001, questions on language of work were

---

7   As a point of fact, as of 2016 in Ontario, French remained by far the most common language after English – in terms of knowledge of language, mother tongue, home language, and language of work. See Statistics Canada, "Focus on Geography Series, 2016 Census" (www12.statcan.gc.ca/census-recensement/2016/as-sa/fogs-spg/Index-eng.cfm).

added, out of recognition that the importance of a language is not limited to the home.[8]

In 2021, will we add questions related to the language of education, out of recognition that schools are essential to the continued vitality of communities and to linguistic duality in Canada in general? Will Statistics Canada continue to provide innovative and proactive leadership in the world of research to help give voice to our linguistic duality, in all its diversity and fullness?

Let's hope so. For the strength and health of our official languages, from coast to coast to coast, is best understood in terms of the social, economic, cultural, linguistic and institutional vitality of the minorities that speak them, and not only in terms of their size relative to the majorities alongside whom they live.

---

8   Statistics Canada, "Census questionnaire content and derived variables since Confederation" (2009) (www12.statcan.gc.ca/census-recensement/2006/ref/dict/app-ann001-eng.cfm#02).

VISIBLE MINORITY, RACE AND RACIALIZATION IN CANADA AND THE UNITED STATES: CHALLENGING CONCEPTS AND MEASUREMENTS

JACK JEDWAB

Jack Jedwab is the President and CEO of the Association for Canadian Studies and the Canadian Institute for Identities and Migration. Holding a PhD in Canadian History from Concordia University, he has taught at Université du Québec à Montréal and McGill University. He taught courses on the history of immigration in Quebec, on ethnic minorities in Quebec, on official language minorities in Canada and on sport in Canada. He has also written essays for books, journals and newspapers across the country, in addition to being the author of various publications and government reports on issues of immigration, multiculturalism, human rights and official languages. He is a frequent media commentator on politics, public policy, history and culture.

The government of Canada's 1995 Employment Equity Act (henceforth the Act) defines visible minority as referring to persons, other than aboriginal peoples, who are non-Caucasian in race or non-white in colour. The term "non-white" is used in the wording of the Act and in employment equity questionnaires for government applicants and employees. As required by the Act, since 1996 the census of Canada has included a question on visible minority. The results are used primarily by governments, business, community groups, health care providers, researchers and others promoting employment equity. The census responses to visible minority questions include the following groups: "Chinese, South Asian, Black, Filipino, Latin American, Southeast Asian, Arab, West Asian, Korean, Japanese, Visible minority, n.i.e. ('n.i.e.' means 'not included elsewhere'), and Multiple visible minority."[1] However, a few exceptions are applied to some groups.

Efforts to have the term *race* dropped from the United States census have not been successful. On the basis of guidelines provided by the Office of Management and Budget (OMB), the United States Census Bureau collects racial data to help promote civil rights and specifically to support

---

[1] www12.statcan.gc.ca/census-recensement/2016/ref/guides/006/98-500-x2016006-eng.cfm

equal employment opportunities as well as to assess racial disparities in health and environmental risks.

Based on self-identification, census officials contend that "…racial categories…generally reflect a social definition of race recognized in this country and not an attempt to define race biologically, anthropologically, or genetically."

By contrast with classification in Statistics Canada's visible minority census question, the U.S. census lists five minimum categories: White, Black or African American, American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander.

The use of terms *race* and *visible minority* have come under fire from many academics, civil society representatives and policy-makers in North America and Europe who insist that such terminology reinforces stereotypes and obscures meaningful distinctions between different minority groups.

That which follows will review recent developments in Canada and the United States as regards the terminology that seeks to categorize 'skin color' and to determine how best to measure the economic and social outcomes associated with its classification. Particular attention is directed at how analysts deal with the question of mixed identification.

Since the introduction of this uniquely Canadian term, the identification and/or classification of visible minority status has faced mounting criticism and remains an object of ongoing controversy. The idea of inquiring about visible minority status is increasingly part of a global debate about the respective use of the terms *race* and *racialization.*

The term *visible minority* has become a lightning rod, with perhaps the main criticisms involving the composition of "visible minorities", where the selected groups often have little in common with each other, as some groups encounter greater vulnerability than others and yet get rolled up into a singular category for policy purposes. Others note that *visible minorities* may form a majority of the population as is already the case in Vancouver and Toronto. "When we use the aggregated visible-minority groups," StatCan analyst Laetitia Martin said, "we tend to homogenize the experiences of the different minority groups and we divert the attention from where the needs are really the greatest."

In the academy, civil society and amongst a number of policy makers in North America and Europe, the term *racialization* or *racialized minorities* is increasingly used instead of visible minority or racial formation (the former terms aim to replace the term race).

Analysts insist that simply asking such questions reinforces the view that there is such a thing as race. Hochman (2019) contends "… if race is best defined as a biological concept which does not refer to anything in the world – then we should probably cease using 'race' as a descriptive term." Segall (2002) remarks that "belief in the existence of 'race', coupled with widely held beliefs in behavioral differences, is a powerful sociological phenomenon that impacts on intergroup relations." He argues that simply treating the social construct of "race" *as if it were* a biological reality is itself "racist" and should be resisted vigorously.

Hochman states that the term *racialized* offers a way for groups that have been understood and treated

as inferior "races" to assert and defend themselves collectively, while rejecting the biologization and inferiorization associated with "race." Others argue that the use of the term race encourages racism by acknowledging the existence of race.

But Hochman acknowledges that there has been a backlash against the use of the term *racialized*. Some suggest that it is an awkward substitute for 'racial formation' and creates confusion when speaking about racism. It might be argued that there has not been much of a backlash against the term since — outside of the academy and civil society — there is little evidence of considerable public uptake of the word *racialized*.

The 2020 U.S. census form will state that: "race and ethnicity categories are socio-political constructs; should not be interpreted as scientific or anthropological in nature". United States census officials have considered alternatives to the term *race*, and in 2015 experimented with new ways to ask people about their background. Test questions asked respondents which "category" best describes them, and then list a series of races and origins including white, Hispanic and black, among others. The census differentiates *between* race and origin. That Hispanic is not regarded as a race has been cause for some confusion.

The U.S. census bureau ultimately determined that the 2020 census would keep the same racial and ethnic categories used in the previous census. However, under the check box for "Black or African American," a new space will be added on the questionnaire to enable participants to write in their non-Hispanic origins with such examples offered as "African American," "Jamaican" and "Nigerian".

FROM 'RACES' TO VISIBLE MINORITIES TO 'ETHNICITIES'?

There appears to be a growing interest in moving away from the collection of information on visible minorities in favor of using data on ethnic origins so as to identify economic and social inequality. In 2012, the United Nations warned Canada that the term visible minorities risked homogenizing experiences of different ethnic groups. They declared that "its lack of precision may pose a barrier to effectively addressing the socioeconomic gaps of different ethnic groups." Visible minority categories (including the 'not a visible minority' category) can indeed mask a diversity of ethnic origins.

Preliminary findings of a Statistics Canada analysis revealed that income and employment outcomes by ethnic origins gave rise to differing outcomes in the labour market. Statistics Canada has been assessing the feasibility of disaggregating data sets on ethnic origins with a view to exploring whether the visible-minority concept "is still relevant in measuring inequalities in today's labour market." But any such changes would require amending Canada's Employment Equity Act, which would by no means be straightforward.

THE MIX UP?

In Canada and in the United States, complications arise in the process of classification of those cases where census respondents identify as both a visible minority and a non-visible minority. In Statistics Canada's population group, variable data are included for multiple responses such as 'White and South Asian' and 'White and Black'. But the same multiple responses are counted differently in

the visible minority variable because of the need that it be derived in accordance with the government's employment equity requirements. In the visible minority variable, persons who reported 'South Asian,' 'Chinese,' 'Black,' 'Filipino,' 'Southeast Asian,' 'Korean,' or 'Japanese,' in combination with 'White' on a write-in response, are included in the visible minority count for the specific visible minority group reported. For example, respondents who checked 'Black' and 'White' are included in the 'Black' count. But persons reporting 'Latin American' and 'White,' 'Arab' and 'White,' or 'West Asian' and 'White' are classified as 'Not a visible minority'. An important challenge to the use of the *term visible minority* is that not all those identifying as such may be "visually" identifiable. For example, many persons that consider themselves of various Latin American or Arabic ethnic origins choose not to self-identify as a visible minority in the corresponding categories in the census question.

In the United States, the multiple responses to questions about racial identification are often referred to as "mixed race". In the United States, "mixed race" brings to mind a biracial experience of having one black parent and another white one, or perhaps one white and the other Asian. But for many U.S. Latinos, mixed-race identity takes on a different meaning — one that is tied to Latin America's colonial history and which commonly includes having a white and indigenous, or "mestizo" background somewhere in their ancestry. A 2014 report by Pew Research reveals that one-third of U.S. Hispanic adults identify as "mestizo," "mulatto" or some other mixed-race combination. The term "mixed race" has yet to find a satisfactory substitute in Canada. Rarely does the term *multiple visible minority* get employed in public discourse when referring to

persons who identify with more than one group. University of Toronto geographer Minelle Mahtani has recently written an important monograph in Canada entitled *Mixed Race Amnesia: Resisting the Romanticization of Multiraciality.*

According to a Pew Research survey of U.S. Hispanics, those who identify as mixed race, mestizo or mulatto are more likely than Latino adults who do not identify as mixed race to be non-Mexican (45 vs. 36%) and to have a higher educational attainment (45% have some college or more, versus 27%). The U.S. Census Bureau encounters important challenges when measuring Hispanic 'racial identity'. Indeed, a considerable number of Hispanics do not choose a standard census race category such as white, black or Asian. Instead, about four in ten select the 'some other race' and some two-thirds of U.S. Hispanic adults regard being Hispanic as part of their racial background, not just an ethnicity (Gonzalez-Barrera, 2015).

The census of Canada makes it possible to look at the relationship between ethnic identification(s) and visible minority status. In the case of a number of ethnic groups that identify as Latin American (as well as Arabic) there are varying responses to the census question on visible minority status. As observed in the Table below, those who ethnically identify as Argentinean and Brazilian do not identify with a visible minority, a modest majority of Mexicans do and an important majority of Colombians do.

When looking at income data on the basis of the intersections between the four aforementioned Latin American ethnic origins and visible minority status, one observes important differences between those groups identifying as visible min-

TABLE 1: PERCENTAGE DECLARING VISIBLE MINORITY STATUS BY SELECTED ETHNIC GROUP, CANADA, 2016

	Male	Female
Argentinian	41%	42%
Brazilian	45.8%	48.2%
Colombian	86%	86.2%
Mexican	57.2%	59.5%

SOURCE: STATISTICS CANADA, SPECIAL TABULATIONS, CENSUS OF CANADA, 2016

ority Latin American and not identifying as a visible minority. In the table below, we can see a 15-per-cent income difference amongst ethnic Argentinians that respectively identify as Latin American and Not Visible Minority, and the difference rises to 38% for ethnic Brazilians on the basis of whether they define as Latin American or not a visible minority.

CONCLUSION

Identities and identification are in flux — as is the terminology we use to define ourselves. In pluralist democracies, citizens define themselves in multiple ways and census takers face considerable challenges in measuring identities in highly charged contexts. Segall (2002) insists that it is imperative to spread the word that "race" is a cultural construct that varies over space and time and differs across countries. He maintains that groups that are anywhere perceived as "races" are better described as ethnic groups, simultaneously shaped by and reshaping their own particular cultures, and influencing other groups with whom they come in contact. But the continued effort to spread the word and get North Americans to use such terms as 'racialized' as a substitute for race, racial formation or visible minorities does not seem to have made substantial progress outside of the academy, and the time may be right for a reset.

TABLE 2: INCOME STATISTICS FOR SELECTED LATIN AMERICAN ORIGINS, CANADA, MALE, 25-54 YEARS OF AGE, AVERAGE WAGES, SALARIES AND COMMISSIONS IN 2015 ($) AMONG RECIPIENTS UNIVERSITY CERTIFICATE, DIPLOMA OR DEGREE AT BACHELOR LEVEL OR ABOVE

2016	Argentinian	Brazilian	Colombian	Mexican
Total - Visible minority	100,325	67,489	63,870	69,239
Latin American	91,840	56,405	60,453	66,610
Not a visible minority	104,037	77,734	82,725	79,225
Approximate Difference between Latin American and NVM	15%	38%	35%	20%

SOURCE: STATISTICS CANADA, SPECIAL TABULATIONS, CENSUS OF CANADA, 2016

REFERENCES

BERENSON, TESSA, "The 2020 Census May Not Use the Word Race, *Time magazine*, June 18, 2015. http://time.com/3927002/census-race-ethnicity/

GONZALEZ-BARRERA, ANA, "'Mestizo' and 'mulatto': Mixed Race Identities among U.S. Hispanics" Pew Research Center, July 10, 2015. www.pewresearch.org/fact-tank/2015/07/10/mestizo-and-mulatto-mixed-race-identities-unique-to-hispanics/

GRANT, TAVIA AND BALKISSOON, DENISE. 'Visible minority': Is it time for Canada to scrap the term?' *Globe and Mail*, Feb. 6, 2019 www.theglobeandmail.com/canada/article-visible-minority-term-statscan/

HOCHMAN, ADAM (2019). "Racialization: A defense of the concept," *Ethnic and Racial Studies*, Vol. 42, Issue 8. pp. 1245-1262. DOI: 10.1080/01419870.2018.1527937

SEGALL, M. H. (2002). "Why is There Still Racism if There is No Such Thing as 'Race'? " *Online Readings in Psychology and Culture, 5*(1). http://dx.doi.org/10.9707/2307-0919.1045

www12.statcan.gc.ca/census-recensement/2016/ref/guides/006/98-500-x2016006-eng.cfm

www.census.gov/content/dam/Census/newsroom/press-kits/2018/paa/2018-paa-presentation-race-question.pdf

U.S. OFFICE OF MANAGEMENT AND BUDGET, "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity", Federal Register / Vol. 62, No. 210 / Thursday, October 30, 1997 / Notices

# UNPACKING CANADIAN 'DEATHS OF DESPAIR'

MUSAH KHALID

Musah Khalid is a Senior Specialist, Canada Mortgage and Housing Corporation, and a recent graduate of the Economics PhD program at the University of Manitoba.

"Deaths of despair" have become a widely cited and accepted phenomenon occurring in the United States. A large spike in mortality by suicide, drug overdose and alcohol-related disease beginning in the late 1990s altered the downward trajectory, from the 1970s onwards, of all-cause mortality rates of white, middle-aged working-class Americans. During this same time period, Hispanic and Black non-Hispanic mortality rates continued to fall (Case & Deaton, 2017). More specifically, Case and Deaton found that from 1999-2015, white non-Hispanic men and women aged 50-54 with an education of high school or less had a significant increase in mortality rates (Figure 1) (2017). The rate over time is illustrated with a comparison with other OECD countries whose mortality rates remain much lower during this time period (although they argue a similar trend is slowly emerging in some countries such as Canada, the UK and Australia). Case and Deaton ascribe the reversal in mortality rates to the cumulative disadvantage experienced by the white working-class as a conse-

DRUG, ALCOHOL AND SUICIDE MORTALITY, MEN AND WOMEN AGES 50-54

quence of social and economic changes brought on by globalization and technological change leading to crushed aspirations, fueling risky health choices and producing higher risk health behaviours.

FIGURE 1: DEATHS OF DESPAIR ALL AGES, MEN FROM 2000 – 2013
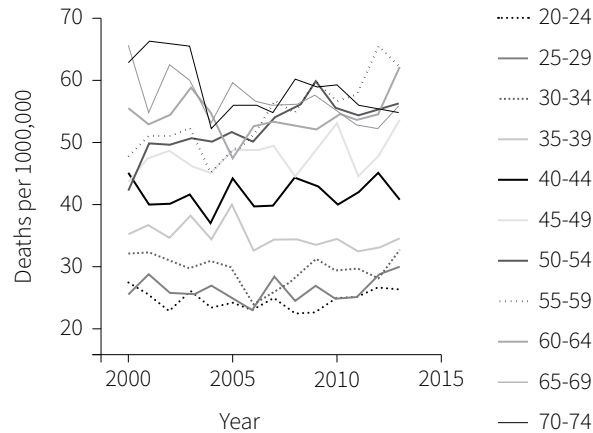
Although some literature has identified dubious evidence indicating the presence of deaths of despair in Canada (Case and Deaton, 2017; Probst & Rehm, 2018), to the authors' knowledge no peer-reviewed studies analyzing whether or not the phenomenon has occurred in Canada have been published. This study fills the gap by examining linked mortality data to discover that middle-aged Canadians are not dying disproportionately due to deaths of despair attributable to economic change. In fact, all ages groups are experiencing increases in mortality rates due to drug poisoning, with middle-aged women's rates increasing at a faster rate than those of men.

METHODS
STUDY DESIGN

We conducted a retrospective, population-based, longitudinal cohort study of mortality due to alcoholic-liver cirrhosis, drug poisoning and suicide in adults aged 20-74 in Canada from 1979 to 2013. We constructed Canada-level five-year age groups to ensure our estimates would be comparable to the Case and Deaton findings (2015; 2017) and analyzed the change in mortality rates by age group, gender, and cause over time.

SETTING AND DATA SOURCES

We used mortality data from the Canadian Vital Statistics Death Database (CVSD) (1979-2013) and the pre-linked Canadian Community Health Survey (CCHS) (cycles 2000-2011) to the aforementioned database, both from Statistics Canada.

The CVSD is a census with a cross-sectional design; the data are extracted from administrative files. The dataset includes demographic and cause of death information collected annually from all provincial and territorial vital statistics registries on all deaths in Canada. The registration of all deaths within provincial and territorial jurisdictions is mandatory. Prior to 2010, some of the data include Canadians who died in the United States. The cause of death is classified according to the World Health Organization International Statistical Classification of Disease and Related Health Problems (ICD).

The linked dataset includes survey iterations from the CCHS from 2000 to 2011. The CCHS is a cross-sectional survey; it collects health information of the non-institutionalized household population aged 12 and older. It excludes members of the Canadian Forces, people living on reserve and some remote areas of Canada (approximately 4 per cent of the total population). Linkage involved: 1) data preparation; 2) record

FIGURE 2: DEATHS OF DESPAIR ALL AGES, WOMEN FROM 2000 – 2013
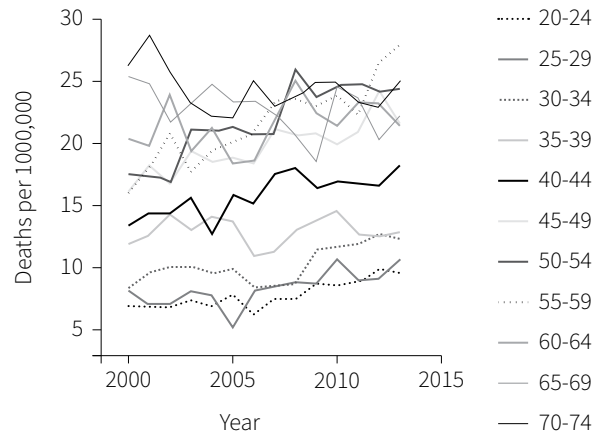
linkage; and 3) quality assessment. The linkage was approved by Statistics Canada's Executive Management Board. Respondents to the CCHS who consented to have their information linked to provincial and federal ministries of heath (85.3 per cent, n=614,774) are included in this dataset. Sampling weights were created by Statistics Canada to adjust for the number of participants who did not consent to linkage.

PARTICIPANTS

In our analysis of the CVSD, we included adults aged 20-74 who died due to alcohol-liver cirrhosis, drug poisoning and suicide in Canada from 1979 to 2013. In the subsequent analysis of the linked dataset to CCHS, we included minors and adults aged 15 and older who died due to alcoholic-liver cirrhosis, drug poisoning and suicide in Canada from 2000 to 2011 (n=925).

PRIMARY OUTCOME

The primary outcome was mortality rates due to drug poisoning, suicide, and alcoholic-liver cirrhosis. Causes of death were based on the ICD9 (1979-1999) and the ICD10 (2000-2013) codes. Categorization of cause of death was determined using the following codes:

- Drug poisoning was characterized by ICD9 codes: 850-860, 980 and ICD10 codes: X40-45, Y10-15, Y45, Y47, Y49;

- Suicide was characterized by ICD9 codes: 950-959 and ICD10 codes: X60-84, Y87.0;

- Alcoholic Liver Disease and Cirrhosis was characterized by ICD9 codes: 571 and ICD10 codes: K70, K73, K74.

For comparison purposes these are the same codes used in the Case and Deaton study (2017).

STATISTICAL ANALYSIS

We first constructed Canada-level age groups, and gender specific five-year alcoholic-liver cirrhosis, drug poisoning and suicide mortality rates for 20-74 year olds for the years 1979 to 2013. We split the years this way to ensure our estimates are comparable to the Case and Deaton (2017) findings. We analyzed time trends of mortality rates of deaths of despair (all three causes combined) by age cohort and gender. We then repeated the process disaggregated by cause. Mortality rates over time were compared between age groups, gender and cause.

Using the linked dataset to the CCHS (2000 –

FIGURE 3: ALCOHOL MORTALITY, MEN AND WOMEN, 2000 – 2013
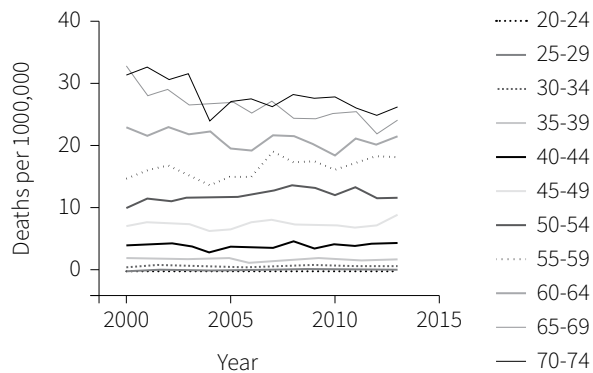
2011), we analyzed education and income source of respondents ages 15 and older by cause. Education was categorized as the highest completed level of school: primary school, secondary school and post-secondary school. Income source was categorized as wages, pension or social assistance.

RESULTS

We found that mortality rates attributable to deaths of despair increased for 45-64 year olds from the late 1990s onwards with the biggest increases being among 55-59 year olds. When disaggregated by gender, for men, the biggest increase occurs among 45-64 year olds (see Figure 1). For women, this is among 40-59 year olds (see Figure 2). When disaggregated by cause of death, there was no significant change in mortality rates due to suicide or alcohol from 2000 to 2013 for either men or women in all age groups (see Figures 3 & 4). However, we did find an increase in mortality rates due to drug poisoning across all age groups, increasing faster for

45-64 year olds (see Figure 5). Although absolute numbers remain higher for men, both genders see increasing rates due to drug poisoning, with a larger increase in the rate for women. Taken together, Canada does not have a 'deaths of despair' problem, but rather a drug poisoning problem. The implications of this result are discussed below.

Furthermore, we looked at the source of income and education levels of individuals who died from drug poisoning between 2000 and 2011 (see Figures 6 & 7). We found that a majority have secondary education with income sources being wages or social assistance.

This study has concluded that middle-aged Canadians are not dying disproportionately due to deaths of despair caused by economic change. The increase in mortality rates due to deaths of despair for 45-64 year olds from the late 1990s onwards is owed to an increase in drug poisoning mortality rates for men and women across all age groups.

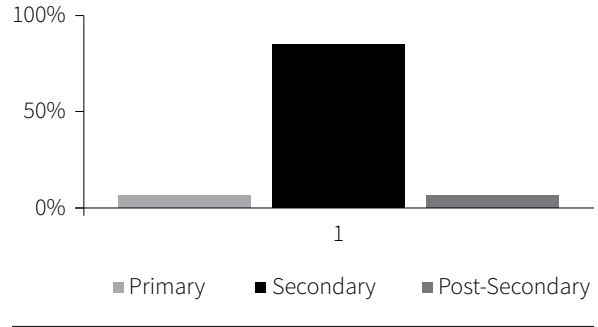**FIGURE 4: SUICIDE MORTALITY, MEN AND WOMEN, 2000 – 2013**
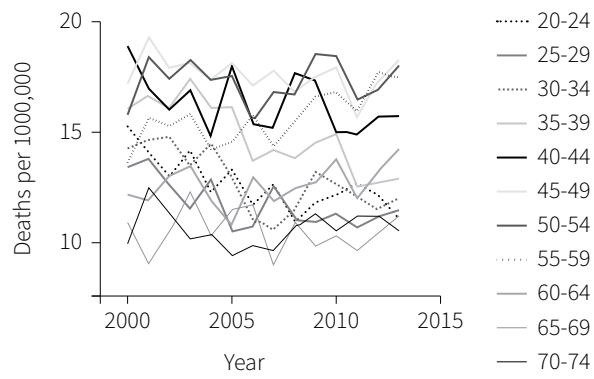
FIGURE 5: DRUG MORALITY ALL AGES, MEN AND WOMEN 2000 – 2013

FIGURE 6: LEVEL OF EDUCATION: DRUG MORTALITY AGE 15 AND OLDER, 2000-2011
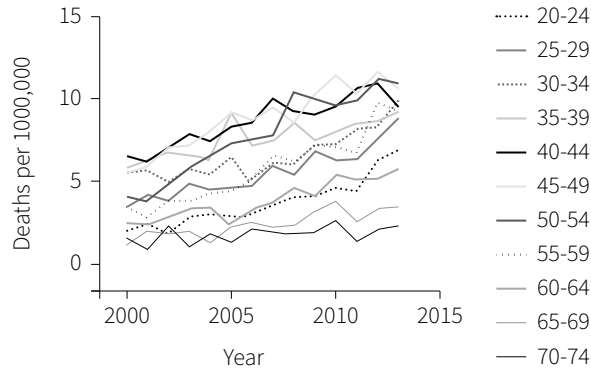
The rates have increased faster for 45-64 year olds with a larger increase observed in women and a disproportionate number of people on social assistance.

While a recent cross-country comparative study of drug-related mortality rates determined that the U.S. was the only country experiencing a drug-related mortality crisis, their analysis failed to include Canada (Barbieri, 2018). It is clear from this study that Canada is facing a similar crisis. This has important policy implications. Instead of concentrating on solving the economic issues of the "disenfranchised middle-aged white working class", policies should aim to curb the increasing trajectory of drug-related deaths. For instance, prescribed pharmaceuticals should be examined more closely given that people living on social assistance have pharmaceutical coverage and a disproportionate number of drug-related deaths were of people on social assistance. It is possible this is contributing to the opioid crisis.

The term "deaths of despair" has become a generally accepted phenomenon. The language/notion/idea has seeped into our everyday way of thinking. It can be found in media sources, and is referred to

FIGURE 7: SOURCE OF INCOME: DEATHS OF DESPAIR AGE 15 AND OLDER, 2000-2011
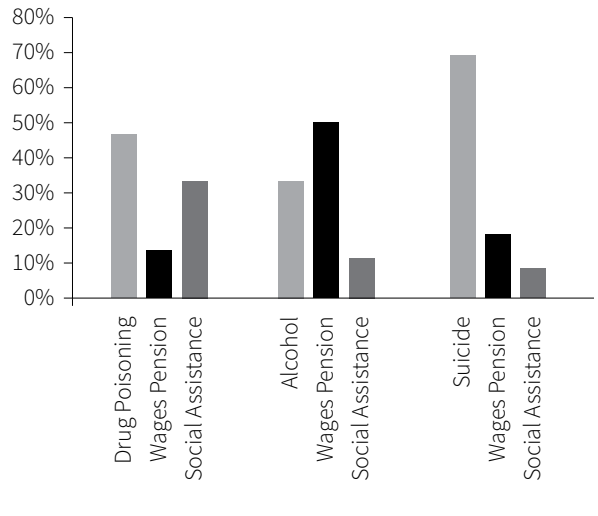
in explaining certain cultural phenomena of the current "uncertain" era. Given alcohol and suicide deaths follow completely different patterns from drug-related deaths, their aggregation into a "deaths of despair" theory as presented by Case and Deaton does not reflect the Canadian reality. It is possible our results reveal the beginnings of the drug crisis. Prescribed pharmaceuticals need to be examined more closely. The most important takeaway message is that middle-aged Canadians are not dying disproportionately due to deaths of despair. To confuse the overwhelming increase in drug-related deaths with the notion of deaths of despair due to economic change is to confuse possible/appropriate approaches/solutions/ to solving the problem.

In conclusion, even though we find that the drug mortality rate is increasing faster for women, more statistical analysis would be necessary to confirm this. It is possible that the deaths are not categorized properly — that some of the cases of drug poisoning would otherwise be categorized as suicide.

REFERENCES

BARBIERI, M. (2018). "The contribution of drug-related deaths to the US disadvantage in mortality," *International Journal of Epidemiology*.

CASE, A., & DEATON, A. (2015). "Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century," *Proceedings of the National Academy of Sciences, 112*(49), 15078-15083.

CASE, A., & DEATON, A. (2017). "Mortality and morbidity in the 21st century," *Brookings Papers on Economic Activity, 2017,* 397.

REHM, J., & PROBST, C. (2018). "Decreases of life expectancy despite decreases in non-communicable disease mortality: the role of substance use and socioeconomic status," *European Addiction Research, 24*(2), 53-59.

FUTURE CHALLENGES AND OPPORTUNITIES FOR STATISTICS CANADA

DOUG NORRIS

Doug Norris, PhD, is a Senior Vice President and Chief Demographer at Environics Analytics and one of Canada's leading experts on the census.

Statistics Canada has come a long way from its origins as a paper-based organization that did tabulations manually. Today, it is collecting and disseminating increasingly complex data and analysis electronically.

Over the past 100 years, the statistical agency has overcome many challenges, but with every test, it has evolved and grown stronger. That experience will be crucial in the coming years as it is about to be tested once again. The rapid advances of technology and the associated rise of "big data" have ushered in a new era in which evidence-based decision-making is the new mantra.

For many organizations, data and analytics are increasingly expected to shape policy and business decisions. Adapting to a wide range of increasing demands won't be easy, but it could mean an even stronger agency.

NEW CHALLENGES

One of Statistics Canada's primary roles is to help all levels of government craft better policies by supplying them with vital information about Canadians. While that mandate has not changed, the questions our leaders are being asked to answer have changed — and they're becoming increasingly complex as Canadian society undergoes a significant shift. For instance, Canada's population is aging, while rural and small towns across the country are shrinking. Overall, the country is becoming more diverse and our leaders are trying to cope with the challenges related to inclusion.

To compound matters, the economy is undergoing a change due to the rapid growth of the digital economy and the adoption of automation, which will compel the future labour force to adopt new skills. This economic change could also exacerbate

inequality in the population.

Governments need higher quality data, delivered in a timely manner and with more granularity, to address these multifaceted issues. It's not just governments who need these data; not-for-profit organizations and private sector businesses are also looking to data to help them assess the changing needs for their products and services, as well as to improve the way they serve their customers. This increased awareness of the power of data has resulted in a renewed focus on "data gaps," which was recently highlighted in an extensive series in the *Globe and Mail*.[1]

The pressure to deliver data in a timely manner will be a particular challenge for Statistics Canada. Historically, the data development cycle was a long, often multi-year process. At times, data only became available after some unexpected event thrust an issue onto the public agenda. Technology has improved timeliness, but the organization still has a long way to go before it can deliver data in near real-time.

Fortunately, the agency is heading in the right direction. There is a recognition that non-traditional data sources can also play a role. A recent example occurred leading up to the legalization of cannabis. To help governments and businesses prepare for the policy change, there was a surge in demand for cannabis data. In response, Statistics Canada conducted a new survey and used "crowdsourcing" techniques to estimate the price paid for cannabis.

A second example involved the demand for data on the foreign ownership of condominiums, which were thought to be a significant factor contributing to the inflation of housing costs. Statistics Canada has been working with the Canada Mortgage and Housing Corporation (CMHC) and municipalities to develop new data on this issue quickly.

DATA COLLECTION

The increasing demand for data is happening at a time when traditional data collection approaches are being called into question. Much of the Statistics Canada social statistics program has been built around household surveys. However, response rates to surveys have been falling and it seems clear that it's time to consider a new approach.

The solution is to look to administrative or other data sources collected for reasons other than statistical purposes. For instance, large federal government administrative data sets, such as tax records and immigration records, have been used for some time as part of some statistical programs, but these represent only the tip of the iceberg.

Governments, the private sector and other organizations collect an enormous amount of data, not the least of which are social media data. These data present statistical challenges, including the coverage of the data set and the concepts and definitions used. Expanding data sources can also introduce new issues around comparability between data sets. This is a particular challenge for data related to health, education, justice and labour. Many of

---

1   *Globe and Mail*, "In the dark: The cost of Canada's data deficit," Jan. 26, 2019

those data fall under provincial jurisdiction where programs and privacy requirements differ from province to province. Municipalities also collect data that would be valuable, although the comparability issues are much greater.

Despite these challenges, the use of administrative data has several advantages. In many cases, the data will be more accurate. For example, the shift to using tax data for all respondents in the 2016 Census significantly improved the accuracy of the income data, especially at a small-area level. The use of postal codes or address coding also allows for the production of small-area level data — something that has been in very short supply outside of the census — which is important for businesses and local governments. Moreover, the use of administrative data makes it possible to combine data from different sources or over time, which allows for the development of a much richer data series.

While administrative data can provide great opportunities, all data development initiatives must be carefully balanced against issues of privacy, which are becoming more critical with each passing day. We are fortunate that Statistics Canada has a stellar reputation for safeguarding the privacy and confidentiality of personal data while maintaining or even expanding access to data to inform public policy. Privacy and security expectations will continue to rise, especially as our lives are stored electronically. StatCan needs to be more proactive in informing the public about the importance of the data it collects and its applications, but also about all of its safeguarding mechanisms. They need to

earn the public's trust.

ACCESS AND DISSEMINATION

In addition to the many challenges and opportunities on the data development side, Statistics Canada will also have to continue to improve access to and dissemination of data. Analysts have already seen some major improvements in this area in the post-cost recovery era. Today, all non-confidential data produced by Statistics Canada are available free of charge online.

The introduction and growth of Research Data Centres, which give researchers the ability to access detailed data in a controlled environment, has been a significant advancement in making data more accessible. A future step, on which some work is already being done, is to provide access to detailed microdata in real time to researchers with, of course, the appropriate safeguards on confidentiality and security of the datasets.

While technology has greatly facilitated data dissemination, it also presents larger challenges to maintaining privacy and confidentiality. For instance, while using technology to link datasets can make data more useful, it makes disclosure more complex. It is interesting that the U.S. Census Bureau recently decided to strengthen its data disclosure approach by moving to a technique called differential privacy,[2] which requires all data tabulations to include some random noise. Furthermore, at the time of writing, the Census Bureau has stated that because of this change it may not be able to

---

2   See https://assets.ipums.org/_files/mpc/MPC-Working-Paper-2018-6.pdf

produce some of its detailed tabulations and public use data files from the 2020 Census.

STAKEHOLDERS

A valuable by-product of the widespread focus on data is that there is a growing number of active stakeholders creating data. With much more data available, the opportunities for data development are greatly increased, but they will require new methodological techniques and approaches.

There needs to be much more cooperation among the many data players. Historically, much of Statistics Canada's work has been directed at meeting the needs of the federal government at the national and, to a lesser extent, provincial level, with less attention given to smaller area geographies. The involvement of more stakeholders means that mechanisms need to be established to better understand data needs, including sharing some of the data development work as well as the resulting output of those data.

Another challenge will be deciding how to prioritize data needs. In some cases, Statistics Canada may have to stop producing some of its lower priority data, thereby creating data gaps for some, to free up resources for higher priority needs.

Many opportunities and challenges lie ahead for Statistics Canada and more generally for all data players. Hopefully, the newly announced Canadian Statistics Advisory Council can provide Statistics Canada with guidance. More generally, the government and Statistics Canada have embarked on a data modernization program that not only looks for ways to replace existing data, but will also search for ways to develop new and more detailed data to address issues of the day.

We, at Environics Analytics, are excited about this government's commitment to big data and innovation. We believe that more collaboration between the data geeks in the public and private sectors will accelerate the process. We, together with our clients, some of the best and brightest data users in Canadian business, are committed to helping put data in action and make Canada a data-driven economy and society.