NEW ADMINISTRATIVE DATA FOR RESEARCH ON IMMIGRANTS TO CANADA

Ted McDonald is a Professor of Economics at the University of New Brunswick in Fredericton. He holds a PhD in Economics from the University of Melbourne. Dr. McDonald has been PI or co-PI on over $9 million of research funding since 2009. He is the founding Director of the NB Institute for Research, Data and Training, the New Brunswick lead for the Maritime SPOR SUPPORT Unit, and Chair of the Canadian Research Data Centre Network Academic Council. His main areas of research include the health status and health services use of immigrants, rural residents, minority groups and other subpopulations.

Michael Haan (PhD, University of Toronto, 2006) is an Associate Professor and Canada Research Chair in Migration and Ethnic Relations at the University of Western Ontario. He is also academic director of the Western University Research Data Centre, and director of the Migration and Ethnic Relations Collaborative Graduate Specialization. His research interests intersect the areas of demography, immigrant settlement, labour market integration, and data development. Dr. Haan is widely consulted by provincial and federal governments for policy advice in the areas of immigration, settlement services, the Canadian labour market, and population aging. Dr. Haan is currently investigator or co-investigator on over $6 million of research focused on immigrant settlement, developing welcoming communities, and identifying the factors that predict successful retention of newcomers. Since receiving his PhD in 2006, he has already published over 50 articles and reports on these topics.

Much of the published literature on immigrants and immigration in Canada has been based on a narrow list of data sources: Censuses of Canada, Statistics Canada surveys, and some longitudinal datasets. More recently, researchers' access to new data sources has markedly expanded through the availability of administrative data. This article reviews the strengths and limitations of two new data sources that can be used to study the experiences of immigrants in Canada: 1) the IMDB, a file based on immigrant landing records plus tax filer information, and 2) Provincial Medicare Registry data. The article also briefly presents some results on the settlement and mobility decisions of recent immigrants using these datasets in order to illustrate their potential for new research.

Une grande partie de la littérature publiée sur les immigrants et l'immigration au Canada est basée sur une liste restreinte de sources de données: recensements du Canada, enquêtes de Statistique Canada et certains ensembles de données longitudinales. Plus récemment, l'accès des chercheurs à de nouvelles sources de données s'est considérablement élargi grâce à la disponibilité de données administratives. Cet article passe en revue les forces et les limites de deux nouvelles sources de données pouvant être utilisées pour étudier les expériences des immigrants au Canada: 1) la BDIM, un fichier basé sur les dossiers d'établissement des immigrants et des informations des déclarants fiscaux, et 2) les données des régimes d'assurance-maladie provinciaux. L'article présente également brièvement quelques résultats sur les décisions d'établissement et de mobilité des immigrants récents en utilisant ces ensembles de données afin d'illustrer leur potentiel pour de nouvelles recherches.

As Canada pursues an immigration policy that sets annual immigrant targets of more than 300,000 people, the need for timely evidence on its impacts for new immigrants, established immigrants, and non-immigrants alike, is high. Much of the published literature has been based on a narrow list of data sources: Censuses of Canada, Statistics Canada surveys, and some longitudinal datasets. More recently, researchers' access to new data sources has markedly expanded through the availability of administrative data. Although every data source has strengths and limitations, they are sometimes not well understood.

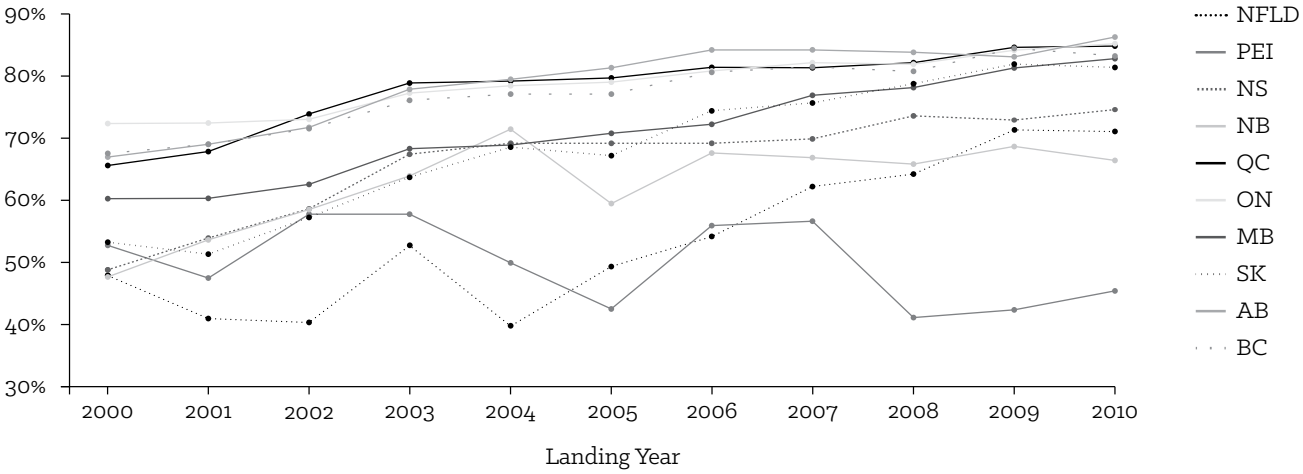
ADMINISTRATIVE DATA SOURCES

When it comes to administrative data for studying immigration, the Longitudinal Immigrant Database (IMDB), is probably the most important file currently available. The original IMDB consisted of landing records linked to the family tax-filer data. The landing records contain all of the information that a newcomer must submit to Immigration, Refugees, and Citizenship Canada when they land in Canada. This includes admission category, education, intended destination, knowledge of official languages, size of immigrating unit, and host of other useful pieces of information. When linked to the tax files, annual information on earnings, union dues, self-employment and investment income, as well as marital status, number of dependents, and postal code of residence, are available.

The IMDB is promising for a number of reasons. First, it is the first long-term longitudinal file that can be used to study immigrants in great detail. Second, it contains 100 percent of Canada's foreign-born tax-filing population, enabling detailed analysis of populations that are small, in remote locations, or are otherwise hard to track. With a host of planned future linkages, the strengths of the IMDB will only become more manifest. That said, the IMDB has three main limitations. First, only data on immigrants are included so that comparisons to native-born Canadians require data from other sources and a synthetic cohort approach. Second, the postal code provided for individuals is not necessarily always an individual's personal residence but rather the address for tax purposes; there are occasions where it appears as though the residence of the accountant is listed. This probably does not matter for studying some aspects of migration (such as inter- and many types of intra-provincial moves), but it may make movements within low levels of geography difficult to study. Finally, there are a few noteworthy substantive pieces of information that are not on the file. These include improvements in education after landing, occupation, and visible minority information, all of which are important for understanding the immigrant experience.

One of the insights that can be gained from the IMDB is investigating how many people go to their intended destination because the landing records contain information on where people plan to live, and the tax files provide the postal code of where they actually end up living. Figure 1, generated by Jonathon Amoyaw and Yoko Yoshida of Dalhousie University, presents the percentage of new arrivals that file their first Canadian tax return in the same province that was their intended destination in the landing records.

FIGURE 1: PERCENTAGE OF NEW ARRIVALS THAT FILE THEIR FIRST CANADIAN TAX RETURN IN INTENDED PROVINCE OF DESTINATION



Source: IMDB

For most provinces, a majority of immigrants file their first tax return in their intended destination province. Although this was already true for 2000, a growing number of newcomers headed to their intended destination province over

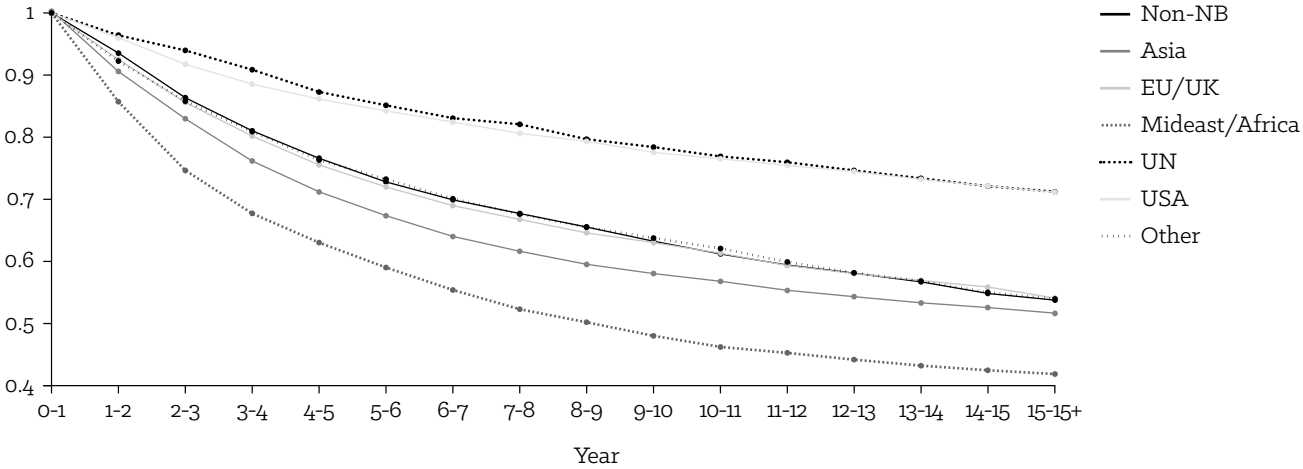
time, so that by 2010 six provinces had match rates of over 80 percent. In three Atlantic Provinces (all but Prince Edward Island, PEI, discussed below), match rates vary from 66 percent in New Brunswick to 75 percent in Nova Scotia. PEI had the lowest rate of landings from those that stated they were heading there but ended up elsewhere, and it also had the most dramatic fluctuations over the period.

There are also multiple administrative files collected by the provinces. Each province in Canada is responsible for the provision of health, education and social assistance services, and in doing so, they collect large amounts of data on the recipients of those services. With few exceptions, provincial health insurance is universal for residents of each province. In New Brunswick, the Medicare Registry includes data on an individual's date of birth, preferred language, gender, country or province of previous residence, year of arrival to the province, Medicare eligibility status, citizenship status, and postal code. It is therefore possible to identify immigrants based on citizenship status and country of previous residence. Since Medicare coverage ceases when an individual moves out of the province, Registry data will record the date of departure from New Brunswick and so enable examination of immigrant retention. Furthermore, by linking Registry data with other administrative datasets, it is also possible to study issues as diverse as the educational outcomes of immigrant children, the health service use of older immigrants, and the extent to which immigrants rely on social assistance.

Medicare data offer both advantages and disadvantages relative to survey, Census and the IMDB. Advantages include the fact that almost all residents of a province are covered; the data is longitudinal since Medicare data date back to the 1970s in New Brunswick and changes of address within New Brunswick are recorded; data on entry to and exit from New Brunswick do not rely on self-reported information; and data are made available to researchers substantially faster than is the case for other datasets. Disadvantages mainly centre around what is missing – there is no information on entry class for permanent residents; country of last residence is self-reported and country of birth is not available; and the date at which immigrants leave NB and Canada for another country will not be known – only the date at which Medicare registration lapsed. As well, given privacy laws, analysis also must be conducted on a province by province basis and linkage across provinces is still a long way off.

NB invests significant resources in the recruitment and retention of new immigrants in an effort to address the challenges of an aging population and stagnant economic growth. One of the key questions for its policymakers is therefore what factors are likely to improve immigrant retention. Using NB Medicare registry data, it is possible to examine immigrant retention without the need for linkages to other datasets by relying on the duration for which immigrants were enrolled in Medicare. McDonald, Cruickshank and Liu (2018) take this approach and define immigrants as individuals who came to NB from another country and who were listed as non-citizens at the time of registration into Medicare. In Figure 2, statistical survival curves are plotted that give the proportion of initial arrivals who are still in NB after the number of years they have been in the province, disaggregated by region of last residence. Retention rates are the highest for immigrants coming from the US and humanitarian immigrants (classified as UN immigrants) and lowest for immigrants from the Middle East and Africa. For comparison, the Non-NB category references Canadian-born but not-native New Brunswickers and their likelihood of leaving NB.

FIGURE 2: PROPORTION OF INITIAL ARRIVALS STILL IN NB AFTER A NUMBER OF YEARS



While this result generally accords with results using other datasets, what is notable is that five-year retention statistics – approximately 75 percent of initial arrivals on average – are substantially higher than what is often estimated using the

IMDB. The difference reflects the number of individuals who legally 'land' in New Brunswick based on intended province of destination compared to the number who actually arrive in NB and are resident at least long enough to register for health care. Thus, the Medicare statistics appear significantly more encouraging for immigrant retention.

MULTIPLE LINES OF INQUIRY NEEDED

Recently available datasets are expanding the range of policy-relevant questions that can be answered, leading to a better understanding of the outcomes of Canadians and thus to more effective policy development. This includes not just the datasets described here but the increasing array of linked Census-survey-administrative datasets being provided by Statistics Canada through the Research Data Centre network. However, every dataset comes with strengths and limitations, and this can limit the generalizability of results and require caveats be noted alongside those results. The best approach is to undertake an analysis using multiple lines of inquiry – defining consistent research approaches applied to multiple complementary datasets. In that way the robustness of results can be assessed and notable differences in results can highlight policy issues warranting further investigation.

REFERENCES

McDonald, James Ted, Cruickshank, Brent and Liu, Zikuan. 2018. "Immigrant Retention in NB: An Analysis Using Administrative Medicare Registry Data." *Journal of Population Research 35*, no. 4: 325–41.