International Journal of Population Data Science

Journal Website: www.ijpds.org

Linking national immigration data to provincial repositories: The case of Canada

Marcelo L. Urquia[1,2,3], Randy Walld[1], Susitha Wanigaratne[2,4], Nkiruka D. Eze[1], Mahmoud Azimaee[2], James Ted McDonald[5], and Astrid Guttmann[2,3,4,6]

Submission History

Submitted:	09/10/2020
Accepted:	01/04/2021
Published:	25/05/2021

[1]Manitoba Centre for Health Policy, Rady Faculty of Health Sciences, University of Manitoba, 408–727 McDermot Ave, Winnipeg, Manitoba, Canada R3E 3P5
[2]ICES, 2075 Bayview Avenue Toronto, Ontario, Canada M4N 3M5
[3]Dalla Lana School of Public Health, University of Toronto, 155 College Street, 6th Floor, Toronto, Ontario, Canada, M5T 3M7
[4]SickKids, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8
[5]NB Institute for Research, Data and Training, University of New Brunswick, Keirstead Hall 304G, Fredericton, New Brunswick, Canada E3B 5A3
[6]Leong Centre for Healthy Children, University of Toronto, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8

Abstract

Background
Canadian health data repositories link datasets at the provincial level, based on their residents' registrations to provincial health insurance plans. Linking national datasets with provincial health care registries poses several challenges that may result in misclassification and impact the estimation of linkage rates. A recent linkage of a federal immigration database in the province of Manitoba illustrates these challenges.

Objectives
a) To describe the linkage of the federal Immigration, Refugees and Citizenship Canada Permanent Resident (IRCC-PR) database with the Manitoba healthcare registry and b) compare data linkage methods and rates between four Canadian provinces accounting for interprovincial mobility of immigrants.

Methods
We compared linkage rates by immigrant's province of intended destination (province vs. rest of Canada). We used external nationwide immigrant tax filing records to approximate actual settlement and obtain linkage rates corrected for interprovincial mobility.

Results
The immigrant linkage rates in Manitoba before and after accounting for interprovincial mobility were 84.8% and 96.1, respectively. Linkage rates did not substantially differ according to immigrants' characteristics, with a few exceptions. Observed linkage rates across the four provinces ranged from 74.0% to 86.7%. After correction for interprovincial mobility, the estimated linkage rates increased >10 percentage points for the provinces that stratified by intended destination (British Columbia and Manitoba) and decreased up to 18 percentage points for provinces that could not use immigration records of those who did not intend to settle in the province (New Brunswick and Ontario).

Conclusions
Despite variations in methodology, provincial linkage rates were relatively high. The use of a national immigration dataset for linkage to provincial repositories allows a more comprehensive linkage than that of province-specific subsets. Observed linkage rates can be biased downwards by interprovincial migration, and methods that use external data sources can contribute to assessing potential selection bias and misclassification.

Keywords
data linkage; immigration; Canada; provincial; national; linkage rate; correction of linkage rate; linkage bias; linkage methods

*Corresponding Author:
Email Address: marcelo.urquia@umanitoba.ca (Marcelo L. Urquia, PhD)

https://doi.org/10.23889/ijpds.v6i1.1412

Introduction

The Canadian publicly-funded health care system guarantees universal access to basic health care services to all legal residents. The delivery of health care services is the responsibility of the provincial governments, which maintain their own registries. Individuals who register with the provincial or territorial health insurance plans are issued a provincial unique health card number. Because of the near universal coverage, provincial health insurance registries are deemed to be the most comprehensive population rosters. Unlike health care, immigration is managed at the federal level in Canada. Immigration, Refugees and Citizenship Canada (IRCC) is the federal immigration agency that maintains nationwide databases of all applications for immigration to Canada, including those of who were granted permanent residence. In the last decade, four Canadian provincial data repositories [Population Data BC (PopData BC) in British Columbia, Manitoba Centre for Health Policy (MCHP) in Manitoba, the NB Institute for Research, Data and Training (NB-IRDT) in New Brunswick and ICES (formerly the Institute for Clinical and Evaluative Sciences) in Ontario] entered into individual ongoing data sharing agreements with the federal immigration agency to link the national IRCC Permanent Resident database (IRCC-PR database) to provincial health care registries for health services research. This article describes the methods used to link the national IRCC-PR database with provincial health care rosters and its challenges.

Although the national IRCC-PR database has been linked at the national level to tax data, hospital data and a community survey [1], the linkage to provincial health insurance records makes it possible to conduct intersectoral research involving data from health, education, social assistance, justice, and other services that are administered by provincial agencies. Provincial research data repositories use an encrypted version of the health card number to relate de-identified information of multiple linked administrative databases for research in compliance with strict privacy, ethical and legal protocols. The linkage process consists in attaching a unique personal identifier to each individual in the incoming database (e.g. immigration data) to make it linkable to all other existing databases (e.g., hospitalisations, social services).

The linkage rate is a measure of the success of the linkage process and the quality of the linked data. Estimating linkage rates between national immigration and provincial health care registries can be challenging due to different reasons. First, unlike Scandinavian countries and due to its federal political system, Canada does not use a unique personal identifier to link personal information across data systems at the national level, relying on name, date of birth and other variables. Second, the IRCC-PR database contains details of an immigrant's permanent resident application, including a field indicating the province of intended destination at the time of the application. A previous linkage of the IRCC-PR database in the province of Ontario was limited to those whose intended destination was Ontario (province-to-province linkage). However, restricting the linkage to those whose intended destination was Ontario may underestimate linkage rates by including in the denominator individuals who never resided in the province and therefore would not be found in the provincial health care registry. Likewise, restricting the linkage to those whose intended destination was Ontario may miss those who did not originally intend to settle there but eventually did so after having reported a different province of destination. Third, more recently provinces have used the national IRCC-PR database for linkage (country-to-province linkage). The advantage of this approach is not limiting eligible matches to those with a given province of intended destination. One challenge, however, is that the pool of the eligible matches, that is, the denominator for the linkage rate, is unknown. A related challenge is linking a file composed of two subsets of individuals with different probabilities of residing in the province; those intending to settle in the province (high probability) and those intending to settle in the rest of Canada (low probability with higher chances of false positives). The variability in immigration and retention patterns and data linkage experiences between four Canadian provinces offers valuable insights into the linkage of national and provincial data.

The objectives of this study were to A) describe the linkage of the IRCC-PR database with the Manitoba healthcare registry and B) compare linkage methods and rates between the IRCC-PR database and four provincial healthcare registries in Canada before and after accounting for interprovincial mobility of immigrants. Although parts A and B are relatively independent, for readers not familiar with Canadian linkages part A may serve as a background for part B.

Part A) Record linkage in Manitoba

Methods

Data sources

At the time of the linkage, the IRCC-PR database contained records of all immigrants who landed in Canada between January 1, 1985, and December 31, 2017. The records included variables from the permanent residence application (e.g., country of birth, immigration category, landing date, intended destination and sociodemographic characteristics). The Manitoba Health Insurance Registry is the core database or 'spine' to which all other databases are linked in the Manitoba repository. Each record in the original registry file, held at Manitoba Health, Seniors and Active Living (MHSAL) contained individual identifiers and a unique personal health identification number (PHIN) for over two million individuals registered with MHSAL at some point from 1970 to 31st March 2019. Semi-annual data snapshots are sent from MHSAL to MCHP and integrated with historical registry data to create a longitudinal population-based registry. Before the data is sent to MCHP, all records are stripped of personal identifiers and a scrambled PHIN unique to MCHP is assigned to each record. The goal of the linkage process in Manitoba was to attach the PHIN of a Manitoba resident to each individual in the incoming IRCC-PR database before the scrambling algorithm is applied to make IRCC-PR records linkable to the Registry and other data holdings in the MCHP data repository.

Linkage

The linkage was conducted by an MCHP analyst (RW) at the MHSAL offices in Winnipeg, Manitoba using blocking

schemes to substantially reduce the number of comparisons made during the linkage process. The IRCC-PR database was divided into two groups; records with Manitoba as the intended destination and records with another Canadian province as the intended destination, 3.6%, and 96.4% respectively. The IRCC-PR database and the Registry were partitioned into subsets defined by the blocking variables. For each surname in the IRCC-PR database, records in that file and the Registry sharing that surname were grouped into a distinct subset, and sequential searches for matches were conducted within each subset. Both groups were processed using the same blocking schemes for deterministic, probabilistic, and manual matching (Figure 1). Because records of those not intending to settle in Manitoba were more numerous than records of those who intended to settle in Manitoba, the threshold for distinguishing between links and non-links was increased in the probabilistic passes for non-Manitoba records to reduce the likelihood of false positives.

Deterministic linkage involved an exact match on personal identifiers (e.g., last name, DOB) while probabilistic linkage involved comparing records on additional attributes (e.g., sex, region of residence) using LinkPro, a record linkage package [2]. To reduce misspellings of names in LinkPro, phonetic matches on surnames were conducted using Soundex coding, a name encoding algorithm. Matching with Soundex code on surnames was only done after exact matches on other variables. The likelihood of valid links was formalised by creating a linkage score, which added up the agreement or disagreement of all the linkage keys, weighted by their ability to discriminate between valid and invalid links. The linkage weight score generated by LinkPro was the basis of deciding which links to accept, reject or review. Links with weight between the set thresholds were reviewed manually by MCHP data analysts with extensive data linkage experience. After the linkage process was completed, all personal identifiers were removed, and each record was assigned a scrambled unique personal identifier.

Data checks

Data quality assessments were conducted using the MCHP data evaluation framework to ensure that linked records corresponded to unique individuals, to identify likely false positives and that the data were accurate and plausible [3]. Duplicate landing records were removed by retaining the earliest landing record per individual, records with erroneous health insurance coverage (zero days or before birth), records with coverage that ended before landing date, and records with coverage that started after the end of our observation period (31st March 2019) (Figure 1). The observed linkage rate was calculated separately for the Manitoba and non-Manitoba records as the percentage of immigrants in the IRCC-PR database that linked to the same individuals in the Registry.

To assess differential selection bias resulting from the linkage, linkage rates were stratified according to various immigrant characteristics and standardised differences calculated between linked and unlinked individuals. This approach was used in the linkage of the IRCC in Ontario [4]. To assess whether immigrants to Manitoba were different from those settling in other parts of Canada, standardised differences were calculated between those linked to a Manitoba resident and those unlinked individuals who intended to settle in the rest of Canada.

Results

The IRCC-PR database contained 7,468,580 landing records which included 270,190 (3.62%) records with Manitoba as the intended destination. From the 270,190 records whose intended destination was Manitoba, 237,722 records were linked to PHINs in the Registry while 32,468 records were not found or could not be linked to a PHIN in the Registry. The linkage also showed that 40,161 records with non-Manitoba intended destinations ultimately ended up in Manitoba. Therefore, the total number of records linked to the Registry from the IRCC-PR database was 277,883. Deterministic matching accounted for 94.9% of matches, followed by 3.8% and 1.3% for the probabilistic and manual steps, respectively (Figure 1).

After the linkage, we excluded 14,172 (5.2%) records for not meeting the following criteria: duplicates (1,755), records not found in the MHIR (512), erroneous coverage (3,768), coverage ended before landing date (8,014) and coverage started after our observation period (123). After exclusion of these likely false positives, the total number of unique individuals linked from the IRCC-PR database to the MHIR was 263,711 (Figure 1).

Table 2 shows that the sociodemographic characteristics of linked ($229{,}025 + 34{,}686 = 263{,}711$) and unlinked individuals with intended destination Manitoba were generally similar, suggesting little or no bias associated with the linkage process for most characteristics. Large standardised differences (above 20%), indicating differences between unlinked and linked immigrants that are unlikely to be due to chance, were found among immigrants who intended to settle in Manitoba and were more common between 2005 and 2009. Other potential differences (standardised differences between 15% and 20%) included lower proportions of linked immigrants among those who landed in the 1990s, among refugees, and those born in Africa.

Likewise, the sociodemographic characteristics of the unlinked immigrants who intended to settle in the rest of Canada were broadly similar to the linked immigrants who intended to settle in Manitoba (Table 1). Large standardised differences indicate that family and economic immigrants and those who landed in the period 2010–2017 were more common in Manitoba than in the rest of the country, while the opposite applies to those who landed in the 1990s. Other potential differences include those who landed in the early 2000s and those born in the Americas.

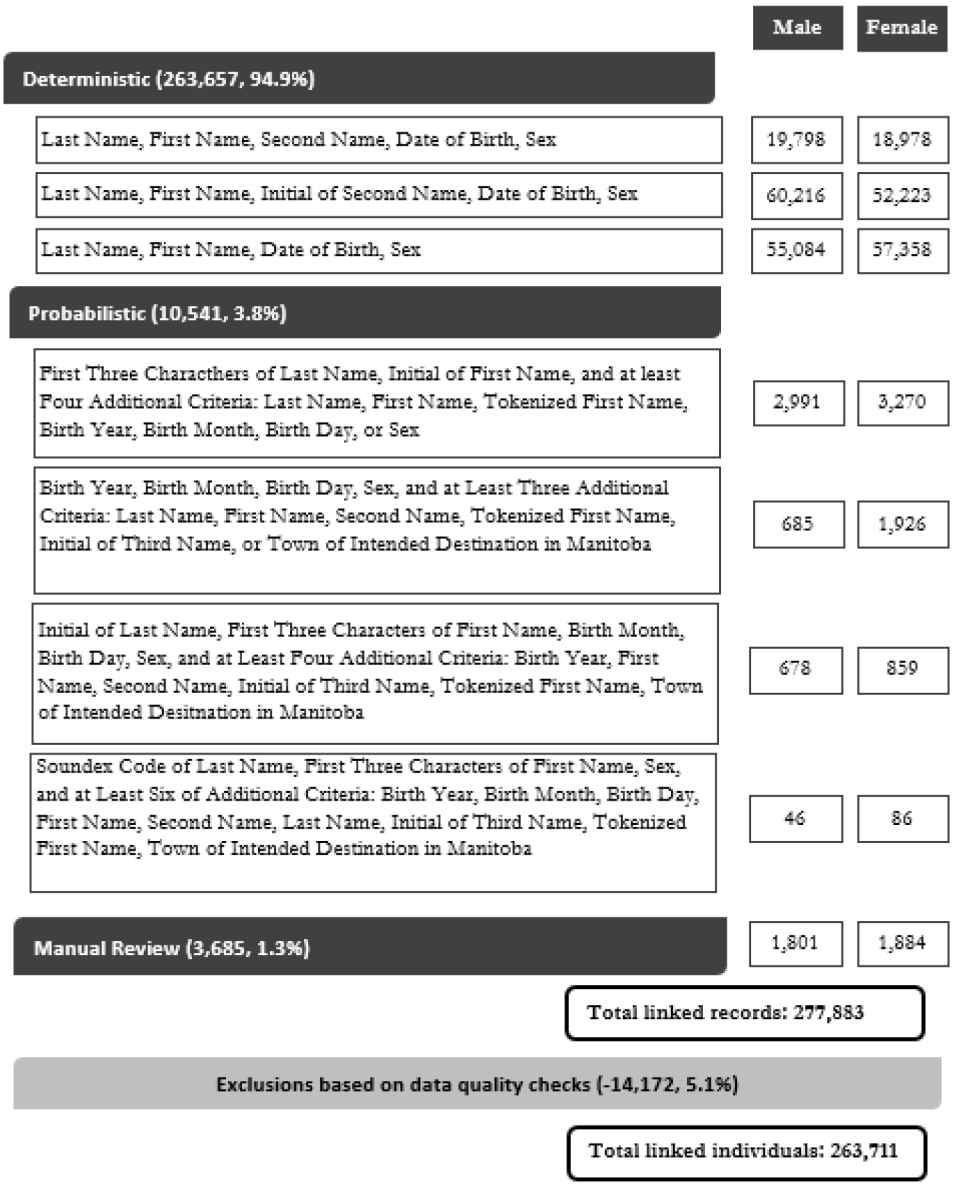
Part B) Comparison of the IRCC-PR database data linkage in four Canadian provinces

Methods

Characteristics of provincial linkages

Linkage methods were similar across provinces; however, the IRCC-PR file received by provinces varied in terms of coverage

Figure 1: Linkage blocking schemes used to Link the IRCC-PR database and the Manitoba Health Insurance Registry



and coverage years. Other differences include software used, methods for data quality assessment, and linkage completion date (Table 2).

British Columbia

On August 14, 2014, Population Data BC completed a probabilistic record linkage to identify common individuals between the IRCC-PR database and the British Columbia Ministry of Health Medical Service Plan Registrations. The IRCC-PR database included all landing records of individuals who were granted permanent residence in Canada from 1985 to 2012. The British Columbia Ministry of Health Medical Service Plan Registrations included identifiers from registrations from 1985–2014. A custom in-house linkage code written in C was used, along with the Jaro-Wrinkler phonetic algorithm. A hybrid probabilistic/deterministic linkage approach started with a probabilistic method to generate the weights and outcome strings, followed by a rule-based system to choose the links. Consistency checks were performed on the results of each rule. Identifiers referenced for linkage included surname, given name, middle name(s), gender, and birth date. Accepted links were refined with a review. Some scenarios during review considered the intended destination of immigrants, particularly the non-BC destined immigrants, with lower probabilities of settling in the province.

New Brunswick

Despite receiving the national dataset from IRCC, the first linkage between the IRCC-PR Database and the New Brunswick Medicare Registry was restricted to immigrants who indicated NB as their intended destination and those who did not have a specified destination. This decision was made to minimise false positives in the context of resource constraints. Because the Medicare Registry is the most complete, accurate, and up-to-date list of all NB residents across all government departments in the province, deterministic matching was the primary method used, followed by some manual matching of records that were categorised as having NB as the intended

Table 1: Linkage of Immigrants from the IRCC-PR database to the Manitoba Health Insurance Registry by Intended Destination and Sociodemographic Characteristics

	Intended destination				Standardised differences	
	Manitoba (%)		Rest of Canada (%)			
Sociodemographic characteristics	Linked (N = 229,025) (a)	Unlinked (N = 41,165) (b)	Linked (N = 34,686) (c)	Unlinked (N = 7,163,704) (d)	Manitoba (%) (e) = (a-b)	Rest of Canada (%)* (f) = (a-d)
Age at Landing (Years)						
0–14	26.34	22.58	23.44	20.75	8.76	13.21
15–24	15.43	15.34	17.21	15.23	0.25	0.54
25–44	46.30	50.71	49.55	48.44	8.83	4.29
45–64	9.91	9.14	8.40	11.98	2.62	6.64
65–85	1.92	1.94	1.37	3.43	0.11	9.34
85 and Older**	0.09	0.29	0.02	0.16	4.44	1.98
Sex						
Female	49.59	49.61	47.92	51.33	0.05	3.47
Male	50.41	50.33	52.08	48.66	0.15	3.50
Landing Year						
1985–1989	8.25	10.15	8.46	9.25	6.59	3.55
1990–1994	8.98	14.24	12.96	16.11	16.46	21.64
1995–1999	5.93	10.65	14.75	13.90	17.18	26.92
2000–2004	10.49	9.25	21.09	15.76	4.16	15.67
2005–2009	21.25	12.71	19.54	16.60	22.88	11.88
2010–2017	45.10	43.00	23.20	28.37	4.23	35.25
Immigration Category	20.25	17.03	25.12	31.07	8.28	24.96
Family						
Economic	64.66	61.49	56.44	54.28	6.58	21.26
Refugee	14.79	20.54	17.70	13.08	15.12	4.94
Other	0.30	0.94	0.67	1.54	8.19	13.01
Birth Region						
Americas	8.72	9.01	10.19	13.63	1.04	15.65
Europe	16.85	15.47	14.78	15.63	3.75	3.30
Asia and Pacific	62.64	58.40	58.43	60.50	8.67	4.39
Africa	11.78	17.05	16.58	10.16	15.04	5.19
Marital Status						
Married or Common-Law	48.48	48.35	48.82	52.07	0.26	7.20
Separated, Divorced or Widowed	2.31	2.32	2.14	3.18	0.10	5.31
Single	48.84	48.75	48.42	44.00	0.19	9.71
Not Stated	0.37	0.58	0.61	0.75	3.01	5.05
Education Level						
Secondary or Less	38.59	38.63	36.04	39.14	0.08	1.14
Some Post-Secondary	16.68	16.68	14.47	16.80	0.01	0.31
Bachelor Degree or Higher	25.90	26.55	31.66	27.68	1.50	4.02
None or Not Stated	18.84	18.14	17.83	16.38	1.79	6.44
Occupational Skill Level						
Skilled	20.68	23.08	24.83	22.95	5.80	5.49
Unskilled	8.48	8.51	4.47	5.15	0.10	13.24
Other	70.84	68.41	70.71	71.90	5.27	2.35

Crude percent and standardised differences, 1985–2017.
* Individuals in Group c were not included in this calculation because they are interprovincial migrants who eventually settled in Manitoba and may not represent those who chose to settle in other provinces.
**Fewer than 6 records (under 1%) were missing age in the Permanent Resident Database.

destination but could not be matched after various processes of deterministic matching were performed with the Power Query linkage software. To differentiate unmatched individuals due to matching errors from those who may have moved out of

Table 2: Characteristics of Provincial Linkages between the IRCC-PR database and Provincial Healthcare Registries

Linkage Province Characteristics	British Columbia	Manitoba	New Brunswick	Ontario
Source IRCC-PR file	National	National	National	Provincial: subset composed of those whose intended destination was Ontario
Period	1985/01/01 to 2012/12/31	1985/01/01 to 2017/12/31	1985/01/01 to 2019/02/05	1985/01/01 to 2017/05/31
Health Care Registry	Provincial	Provincial	Provincial	Provincial
Type of linkage	Country-to-province	Country-to-province	Province-to-province	Province-to-province
Linkage completion date	August 2014	September 2019	December 2019	First linkage: June 2012 Last update: May 2018
Linkage Software	In-house code written in C	LinkPro	Power Query	AutoMatch
Phonetic Algorithm	Jaro-Winkler	Soundex Coding	None, but string name variations	New York State Identification and Intelligence System
Data Quality Assessment	Consistency checks	MCHP Data Evaluation Framework	NB-IRDT Data Quality Framework	ICES Data Quality Framework
Stratification by intended destination	Yes	Yes	No, linkage limited to those intending to settle in the province	No, source file limited to those intending to settle in the province
Immigrant retention rates (2008-2013)*	90.4%	84.3%	63.9%	93.1%

* The percentage of immigrants who arrived between 2008 and 2013 who resided in their province of destination in the 2013 tax year [5].

province, if no other family members in the same immigrant household were found to be present the unmatched individual was counted as never arrived in the province. A study found that about 89% of unmatched primary applicants lived in a household where no family members could be matched [6]. Individuals who were linked from the IRCC-PR database to Medicare Registry records were assigned their corresponding unique scrambled identifier that enabled the file to be linked with other data files held at NB-IRDT. Duplicate records were removed by retaining the record with the earliest landing date.

Ontario

The IRCC-PR database contained the records of over three million immigrants who landed in Ontario between January 1985 and December 2016, while the Ontario Registered Persons Database (RPDB) comprises the base population file of 13.5 million Ontario residents eligible for provincial health care coverage. ICES performed the record linkage between the IRCC-PR database and the RPDB. A blocking technique, similar to the Manitoba linkage was implemented which involved deterministic, probabilistic, and manual review. Automatch, a probabilistic record linkage program, was used for matching and to augment the record linkage process. The New York State Identification and Intelligence system was used for phonetic conversion. To improve the accuracy of the manual review process, the Statistics Canada Postal Code conversion file was used to generate the corresponding Ontario city location to compare geographic information between the IRCC-PR database and the RPDB. Individuals who were linked from the IRCC-PR database to the RPDB were assigned unique scrambled identifiers, derived from individual health card numbers. Duplicate records were removed by retaining the record with the earliest landing date. More details about the linkage process in Ontario can be found elsewhere [4].

Correction of observed linkage rates

To approximate the actual unknown denominators for the provincial linkage rates, we accessed the Longitudinal Immigration Database (IMDB), a linkage between the IRCC-PR database and the "T1 Family file" based on personal income tax returns from 1980 to 2016 provided to Statistics Canada by the Canada Revenue Agency [7]. Immigrant tax filing behavior from the IMDB was used as a proxy of the number of immigrants who ever resided in a province, and thus provides a more accurate linkage rate compared to the observed raw rate obtained from the IRCC-PR database alone. Using tax filling information from the IMDB as a proxy for residence, however, may miss some immigrants who temporarily resided in a given province but did not file taxes in that province. In addition, the tax record linkage of the IMDB is subject to error [1].

We estimated corrected linkage rates using the IMDB database in each of the four provinces separately for immigrants who intended to settle in the province of linkage

and immigrants who intended to settle in the rest of Canada by multiplying the denominators by a province-specific correction factor. The correction factor approximates the proportion of immigrants in the IRCC-PR database who filed taxes in the province of linkage, by intended destination. The correction factor for those who intended to settle in the linkage province was (b-a)/b, reflecting retention of the original immigrants to the province, and for those who intended to settle in the rest of Canada was c/d, reflecting attraction of immigrants originally destined to other provinces, where:

- a = immigrants who intended to settle in the linkage province, but never filed taxes in that province
- b = total immigrants who intended to settle in the linkage province
- c = immigrants who did not intend to settle in the province of linkage, but who filed taxes in that province
- d = total immigrants who intended to settle in any other Canadian province except the province of linkage (Rest of Canada)

Results

The observed provincial linkage rates for immigrants who intended to settle in the linkage province ranged from 74.00% in New Brunswick to 86.68% in Ontario (Table 3), which aligns with provincial retention rates shown in Table 2. Observed linkage rates for immigrants who intended to settle in other Canadian provinces could only be calculated for British Columbia and Manitoba; 5.21% and 0.48% respectively.

Using the corrected denominator, the estimated linkage rates for immigrants who intended to settle in the linkage province increased by 5 to 20 percentage points across provinces. For immigrants who intended to settle in other Canadian provinces, linkage rates increased to 99.20% and 87.61 in British Columbia and Manitoba. The ability to perform the linkage with the subset of those who intended to settle in the rest of Canada resulted in a gain of about 260,000 immigrants in British Columbia and of 40,000 in Manitoba, accounting for 22% and 14% of all immigrants to the respective provinces. The inability to include this subset in New Brunswick and Ontario resulted in the loss of about 12,000 (23%) and 415,000 (11%) immigrants to these provinces, respectively.

Discussion

Main findings

The linkage between the IRCC-PR and the Manitoba provincial healthcare Registry

The linkage between the Manitoba portion of the IRCC-PR and the Manitoba provincial Registry, mainly accomplished by deterministic matching, resulted in a linkage province-to-province rate of 84.76%. After stratifying the linkage process according to the province of intended destination and accounting for interprovincial mobility (country-to-province) the linkage rate was estimated to be 96.14%. Comparison between linked and unlinked individuals suggested little or no differential selection bias due to the linkage process according to most immigrant characteristics. Similar findings were observed in the Ontario linkage [4].

Comparison of the linkage in the four provinces

Consistent with the case of Manitoba, the observed crude linkage rates for the subset of immigrants intending to settle in the province where the linkage was conducted ranged from 74% to 87% in the other three provinces. Similarly, after correction for interprovincial mobility, resulting linkage rates were higher, in the range 88 to 99%.

Country-to-province versus province-to-province linkage

The opportunity to include all immigrants who intended to settle in the province where the linkage was conducted and the rest of Canada (country-to-province linkage) resulted in sizeable gains in the number of immigrants matched to a resident in British Columbia and Manitoba because many of those who intended to settle in the rest of Canada eventually settled in the province of linkage. By contrast, New Brunswick and Ontario (province-to-province linkage) missed 23% and 11% of all immigrants to the province, respectively, for not being able to use the national immigration file. These unlinked immigrants were those who intended to settle in the rest of Canada and could not be included in the province-to-province linkages. The use of the national file ensured that all immigrants to Canada are eligible for matching, irrespective of their original intended destination.

Interpretation

The observed raw linkage rates for the subset of immigrants intending to settle in the province of linkage, ranging from 74% to 87% across provinces, are misleading from various perspectives. Firstly, they cannot be used to compare the success of the provincial linkages. Despite some methodological differences, the lowest linkage rate in New Brunswick (74%) and the highest in Ontario (87%) reflect the settlement patterns of immigrants rather than the efficiency of the linkages. Immigrant retention rates are also lowest in New Brunswick (64%) and highest in Ontario (93%). In other words, those who declared their intention to settle in the province of linkage but ended up settling in another province and therefore could not be found in the provincial registries were much more common in New Brunswick than in Ontario. This suggests that the crude observed linkage rates are not directly comparable from a methodological standpoint, since they are heavily influenced by immigrant retention [8]. Secondly, these raw linkage rates are underestimated. Because immigrant retention is not fully achieved in any province, the denominators used to compute the linkage rates are inflated by including a varying proportion of immigrants who are not eligible to be matched to a resident. The lower the retention rate of immigrants the higher the underestimation rate. Thirdly, the previous considerations are warranted by the finding that after correction for interprovincial mobility the linkage rate was higher for New Brunswick (94%) than for Ontario (91%).

Table 3: Observed and Estimated Linkage Rates in British Columbia, Manitoba, New Brunswick and Ontario

Province of Linkage and Immigrants' Intended destination	Observed			Estimated		
	Records in IRCC (denominator) (a) N	Records Linked to Provincial Database (b) N	Raw Linkage Rate (c) = b/a %	Correction Factor (d)	Corrected denominator (e) = a*d N	Corrected Linkage Rate (f) = b/e %
British Columbia (1985–2012)						
British Columbia	1,010,845	825,382	81.65	0.9204[1]	930,381	88.71
Rest of Canada	5,083,689	264,754	5.21	0.0525[2]	266,893	99.20
Overall	6,094,534	1,090,136	—	—	1,197,274	91.05
Manitoba (1985–2017)						
Manitoba	270,190	229,025	84.76	0.8687[1]	234,714	97.58
Rest of Canada	7,198,390	34,686	0.48	0.0055[2]	39,591	87.61
Overall	7,468,580	263,711	—	—	274,305	96.14
New Brunswick (1985–2019)						
New Brunswick	54,680	40,550	74.00	0.7854[1]	42,910	94.50
Rest of Canada	8,030,170	105[3]	—	0.0015[2]	12,156	—
Overall	8,084,850	40,655	—	—	53,346	76.21[4]
Ontario (1985–2016)						
Ontario	3,530,278	3,060,086	86.68	0.9494[1]	3,351,645	91.30
Rest of Canada	3,642,025	—	—	0.1100[2]	415,190	—
Overall	7,182,099	3,060,086	—	—	3,766,835	81.24[4]

1. Correction factor for immigrants intending to settle in the linkage province was calculated as (Total intended province IMDB - Did not file taxes in province IMDB) / Total intended province IMDB.
2. Correction factor for immigrants NOT intending to settle in the linkage province was calculated as (Ever filed taxes in province IMDB / Total intended rest of Canada IMDB).
3. Represent individuals with an unknown province of intended destination. Those intending to settle in the rest of Canada were not included in the NB linkage process.
4. Corrected overall rates for New Brunswick and Ontario are not true linkage rates because these provinces did not use the national file. However, their calculation illustrates the impact of not using the national file.

The use of the nationwide IRCC dataset in British Columbia and Manitoba resulted in substantial gains in the number of immigrants linked to provincial residents. The gains were greater in British Columbia because this province attracts a much larger proportion of immigrants to Canada from other provinces than Manitoba.

Limitations

The lack of a national registry limited our ability accurately account for interprovincial mobility of immigrants. To circumvent this limitation, we used the IMDB. However, since the IMDB results from a linkage between federal income tax files and the IRCC-PR database (83% linkage rate) [1] the use of these external data to obtain estimated linkage rates for the provinces may have resulted in some degree of error. Tax filling may underestimate residence, since some immigrants may face challenges to file taxes, particularly shortly after arrival due to language and systemic barriers and low income. Although more accurate and realistic than the raw linkage rates, our linkage rate correction represents an approximation that accounts for interprovincial migration and should not be interpreted literally as "correct". Since the focus of this study was on the impact of provincial versus national datasets in the estimation of linkage rates among immigrants, we did not attempt to measure other sources of variation between provinces. Such an exercise could be done when the source IRCC-PR file is widely available and data linkage approaches are more standardised across provinces. Interprovincial comparisons of corrected linkage rates must be done with caution, given the various sources of variation between provinces. Finally, the limitation of not using the national dataset prevented New Brunswick and Ontario from matching interprovincial migrants. Consequently, many immigrants who intended to settle in the rest of Canada but did so in Ontario (∼415,000; 11% of all immigrants to the province) and New Brunswick (∼12,000; 23%) were not eligible to be linked to the provincial registries, many of whom may have remained misclassified as non-immigrants. Such misclassification may introduce substantial bias in comparisons between immigrants and non-immigrants in research studies. Future updates of these linkages will likely involve the nationwide IRCC file and overcome the noted limitations.

Conclusions

Linking national datasets to provincial repositories poses challenges in estimating linkage rates in the absence of a unique national personal identifier. These challenges and workarounds were illustrated with the case of the linkage

between a national immigration database and provincial heath care registries in Manitoba and other Canadian provinces. Despite provincial differences in linkage methods, observed linkages were relatively high across provinces. Our findings indicate that linkage rates based on a provincial subset of immigrants (province-to-province linkage) are biased downwards by interprovincial migration and failure to use a nationwide dataset (country-to-province linkage) may result in both under-capture and substantial misclassification of immigrants as non-immigrants. These findings, seen in the context of immigrant mobility, suggest that researchers may benefit from collaborating across provinces to capture a full picture when studying immigration data collected at the national level. Despite some limitations, methods that use external data sources containing information on interprovincial mobility can be used to assess potential selection bias resulting from the province-to-province linkage of immigration data. However, the use of a country-to-province approach to linkage is recommended whenever possible.

Acknowledgments

The authors acknowledge the Manitoba Centre for Health Policy (MCHP) for the use of data contained in the Manitoba Population Research Repository under project #2019–012. The results and conclusions are those of the authors and no official endorsement by the MCHP, Manitoba Health, or other data providers is intended or should be inferred.

We acknowledge Heather Prior, Gilles Detillieux, Ina Koseva, Dr. Randy Fransoo, and Dr. Chelsea Ruth from MCHP who contributed to the linkage process. We are grateful for the comments provided by Dr. Maria Chiu (ICES) on an earlier version of the manuscript, and our Advisory Group members (Anna Bird and Jamie Matwyshyn from Manitoba Education & Training, Karen Serwonka and Marc Silva from Manitoba Health Seniors and Active Living and Dr. Lori Wilkinson from the University of Manitoba). We thank Immigration, Refugees and Citizenship Canada, especially Lorna Jantzen and Sophie Qu, for their help in facilitating this collaborative research. We thank Population Data BC for the provision of British Columbia linkage-related information and Population Data BC staff (Harold Yip, Brent Deere and Tim Choi). We thank Zikuan Liu of the New Brunswick Institute for Research, Data and Training, for data clarifications.

We acknowledge Manitoba Health, Seniors and Active Living, Immigration, Refugees and Citizenship Canada, and Statistics Canada Research Data Centers for the use of their data.

Funding

This study was supported through funding from Manitoba Health, Seniors and Active Living and the Canadian Foundation of Innovation. MLU holds a Canada Research Chair in Applied Population Health.

Statement on conflicts of interest

The authors have no conflicts to declare.

Ethics statement

Ethics approval for this study was granted by the University of Manitoba Health Research Ethics Board (HS22881 (H2019:216)) and was reviewed for data privacy and approved by the Government of Manitoba's Health Information and Privacy Committee (File No. 2019/2020-17). Use of the provincial data was approved by Manitoba Health, Seniors and Active Living. Use of the Immigration Longitudinal Database was approved by the University of Manitoba Health Research Ethics Board (HS23221 (2019:362)).

References

1. Evra R, Prokopenko E. Longitudinal Immigration Database (IMDB) Technical Report, 2016 in Analytical Studies: Methods and References [Internet]. 2018. Available from: https://www150.statcan.gc.ca/n1/en/pub/11-633-x/11-633-x2018019-eng.pdf?st=_pXdQbW_shttps://www.statcan.gc.ca/eng/rdc/cencchs-imdbhttps://www.statcan.gc.ca/eng/rdc/liddad
2. Roos LL, Wajda A, Sharp SM, Nicol JP. Software for health care analysts: A modular approach. J Med Syst [Internet]. 1987 Dec;11(6):445–64. https://doi.org/10.1007/bf00993011
3. Smith M, Lix LM, Azimaee M, Enns JE, Orr J, Hong S, et al. Assessing the Quality of Administrative Data for Research: a Framework from the Manitoba Centre for Health Policy. J Am Med Informatics Assoc [Internet]. 2018 Mar 1;25(3):224–9. https://doi.org/10.1093/jamia/ocx078
4. Chiu M, Lebenbaum M, Lam K, Chong N, Azimaee M, Iron K, et al. Describing the linkages of the immigration, refugees and citizenship Canada permanent resident data and vital statistics death registry to Ontario's administrative health database. BMC Med Inform Decis Mak [Internet]. 2016;16(1):135. https://doi.org/10.1186/s12911-016-0375-3
5. van Huystee M. Interprovincial Mobility: Retention Rates and Net Inflow Rates 2008–2013 Landings [Internet]. 2016. Available from: https://www.canada.ca/content/dam/ircc/migration/ircc/english/pdf/research-stats/r33-2016_mobility-eng.pdf
6. Leonard P, McDonald T, Miah P. Analysis of unmatched immigrants in the BizNet Database. [Internet]. Fredericton, NB; 2020. Available from: https://www.nbirdt.ca/nbirdt-files/publications/79/59
7. Statistics Canada. Longitudinal Immigration Database (IMDB) Technical Report, 2018 in Analytical Studies: Methods and References [Internet]. Ottawa, ON; 2019. Available from: https://www150.statcan.gc.ca/n1/en/pub/11-633-x/11-633-x2019005-eng.pdf?st=1XQ95Ksr
8. van Huystee M, St Jean B. Interprovincial Mobility of Immigrants in Canada 2006–2011 [Internet]. 2014. Available from: https://www.canada.ca/content/dam/ircc/migration/ircc/english/resources/research/documents/pdf/mobility2006-2011.pdf

Abbreviations

ICES: Institute for Clinical and Evaluative Sciences (formerly)
IMDB: Longitudinal Immigration Database
IRCC: Immigration Refugees and Citizenship Canada
IRCC-PR: Immigration Refugees and Citizenship Canada Permanent Resident database
MCHP: Manitoba Centre for Health Policy
MHSAL: Manitoba Health Seniors and Active Living
MHIR: Manitoba Health Insurance Registry
NB-IRDT: New Brunswick Institute for Research, Data and Training
PHIN: Personal Health Insurance Number
PopData: Population Data BC
RPDB: Ontario Registered Persons Database