
Article:
de Oliveira, Claire orcid.org/0000-0003-3961-6008, Gatov, Evgenia, Rosella, Laura et al. (17 more authors) (2022) Describing the linkage between administrative social assistance and health care databases in Ontario, Canada. International Journal of Population Data Science. ISSN: 2399-4908

https://doi.org/10.23889/ijpds.v7i1.1689

International Journal of Population Data Science

Journal Website: www.ijpds.org

Describing the linkage between administrative social assistance and health care databases in Ontario, Canada

Claire de Oliveira[1,2,3,4,*], Evgenia Gatov[1], Laura Rosella[1,2,5,6,7], Simon Chen[1], Rachel Strauss[1], Mahmoud Azimaee[1], Elizabeth Paterno[8], Astrid Guttmann[1,2,9,10,11], and Ministry of Children, Community and Social Services-ICES Working Group (Nelson Chong, Peter Ionescu, Sean Ji, Alexander Kopp, Annie Lan, Charlotte Ma, Miranda Pring, Priyanka Raj, Steven Ryan, Refik Saskin, Fiona Wong)[1,8]

Submission History

Submitted:	19/08/2021
Accepted:	20/01/2022
Published:	03/03/2022

[1]ICES, Toronto, Ontario, Canada
[2]Institute of Health Policy, Management and Evaluation, University of Toronto, Ontario, Canada
[3]Institute for Mental Health Policy Research, Centre for Addiction and Mental Health, Toronto, Ontario, Canada
[4]Centre for Health Economics and Hull York Medical School, University of York, York, United Kingdom
[5]Dalla Lana School of Public Health, University of Toronto, Ontario, Canada
[6]Institute for Better Health, Trillium Health Partners, Mississauga, Ontario, Canada
[7]Public Health Ontario, Toronto, Ontario, Canada
[8]Business Intelligence and Practice Division, Ministry of Children, Community and Social Services, Toronto, Ontario, Canada
[9]Division of Paediatric Medicine and Child Health Evaluative Sciences, Hospital for Sick Children, Toronto, Ontario, Canada
[10]Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada
[11]Edwin S.H. Leong Centre for Healthy Children, University of Toronto, Toronto, Ontario, Canada

Abstract

Background
The linkage of records across administrative databases has become a powerful tool to increase information available to undertake research and analytics in a privacy protective manner.

Objective
The objective of this paper was to describe the data integration strategy used to link the Ontario Ministry of Children, Community and Social Services (MCCSS)-Social Assistance (SA) database with administrative health care data.

Methods
Deterministic and probabilistic linkage methods were used to link the MCCSS-SA database (2003-2016) to the Registered Persons Database, a population registry containing data on all individuals issued a health card number in Ontario, Canada. Linkage rates were estimated, and the degree of record linkage and representativeness of the dataset were evaluated by comparing socio-demographic characteristics of linked and unlinked records.

Results
There were a total of 2,736,353 unique member IDs in the MCCSS-SA database from the 1st January 2003 to 31st December 2016; 331,238 (12.1%) were unlinked (linkage rate = 87.9%). Despite 16 passes, most record linkages were obtained after 2 deterministic (76.2%) and 14 probabilistic passes (11.7%). Linked and unlinked samples were similar for most socio-demographic characteristics (i.e., sex, age, rural dwelling), except migrant status (non-migrant versus migrant) (standardized difference of 0.52). Linked and unlinked records were also different for SA program-specific characteristics, such as social assistance program, Ontario Works and Ontario Disability Support Program (standardized difference of 0.20 for each), data entry system, Service Delivery Model Technology only and both Service Delivery Model Technology and Social Assistance Management System (standardized difference of 0.53 and 0.52, respectively), and months on social assistance (standardized difference of 0.43).

Conclusions
Additional techniques to account for sub-optimal linkage rates may be required to address potential biases resulting from this data linkage. Nonetheless, the linkage between administrative social assistance and health care data will provide important findings on the social determinants of health.

Keywords
data linkage; administrative social assistance data; administrative health care data; Ontario

*Corresponding Author:
Email Address: claire.deoliveira@camh.ca (Claire de Oliveira)

https://doi.org/10.23889/ijpds.v7i1.1689

Introduction

In Canada, universal health care is delivered through provincial and territorial publicly funded health care systems, which, in turn, collect administrative data that reflect patients' interactions with the health care system across multiple sectors (e.g., inpatient and ambulatory care) and over time. In Ontario, Canada's most populous province with over 14.5 million residents, the Ontario Ministry of Health collects data on the health care utilisation of all legal residents eligible for public health care insurance. Under section 45 of Ontario's *Personal Health Information Protection Act* (PHIPA) [1], ICES, an independent non-profit research institute, is a prescribed entity whose legal status under Ontario's health information privacy law allows it to collect, use, and disclose personal health information from health information custodians, without consent, for the evaluation, planning and/or monitoring of the health system. To ensure the privacy and protection of data, ICES implements a series of physical and logical controls to govern access to information, like the use of secure zones within ICES facilities, complex passwords, and encryption. The use of these data has enabled scientists to answer important policy-relevant questions across different disciplines such as health services research, health economics, epidemiology and public health [2–5].

In recent years, there has been a growing interest in examining social determinants of health [6] defined by the World Health Organization as "the conditions in which people are born, grow, live, work and age" [7], which lie outside of the health care sector. The social and economic conditions of an individual are known to substantially impact health outcomes [8, 9]. Within this context, researchers have been interested in understanding the relationship between the receipt of social assistance, typically provided to an economically disadvantaged segment of the population, and their health (such as injuries and substance use) and health care use [10, 11]. However, the lack of reliable and comprehensive data in most regions in Canada has made it difficult to examine the characteristics and outcomes of social assistance recipients. Most Canadian research on social assistance has relied on self-reported population survey data, such as the Survey of Labour and Income Dynamics and the Canadian Community Health Survey [12]. These data are limited by poor response rates, potentially unreliable responses due, in part, to social desirability bias, and biased samples, due to possible underrepresentation of respondents with lower socioeconomic status who may not have the means to participate in surveys [13]. At least one other Canadian province, Manitoba, has been successful in linking administrative health care data to administrative social assistance data [14], which has resulted in work characterising health outcomes in social assistance recipients [15] and the evaluation of a number of health care programs in this population, such as the impact of an unconditional prenatal benefit initiative [16].

The linkage of records across administrative databases has become a powerful tool to increase the amount of information available on individuals for research and analytics, going beyond any individual data source in isolation [17, 18]. For example, in Ontario the linkages of administrative health care databases to the Immigration, Refugees and Citizenship Canada permanent residents data [19], the Office of the Registrar General's Vital Statistics Death Registry [19], and the federal Indian Register [20] have produced important evidence that can be used to inform policy. The Ontario Ministry of Children, Community and Social Services, which administers social assistance programs in the province, partnered with ICES to address the need for more comprehensive data to support decision-making, policy development, and service provision relevant to the health and well-being of individuals living in Ontario.

In this paper, we describe the data integration strategy used to prepare the Ministry of Children, Community and Social Services (MCCSS)-Social Assistance (SA) database for research through record linkage with the administrative health care databases held at ICES. We further evaluate the degree of record linkage and the representativeness of the dataset by comparing the socio-demographic characteristics of linked and unlinked records.

Methods

Data sources

Administrative health care data housed at ICES

The data repository at ICES consists of individual record-level, coded, and linkable health datasets. It includes data on most publicly funded health services for the Ontario population eligible for universal health care coverage since 1986 and is capable of integrating analytics-specific data, registries and surveys. These health service records reflect Ontarians' day-to-day interactions with the health care system, including physician claims submitted to the Ontario Health Insurance Plan, drug claims submitted to the Ontario Drug Benefit Program, discharge abstracts of hospital stays and emergency department visits, and records for home and long-term care. All databases collected from health information custodians can be linked using unique encoded identifiers, termed ICES key numbers (IKNs), which are generated using a secure ICES algorithm based on an individual's health card number.

Administrative social assistance data from MCCSS

Ontario has two social assistance programs, which provide income and employment support to single adults and families who are in financial need: Ontario Works (OW), which provides financial and employment assistance to help people move towards paid employment and independence, and the Ontario Disability Support Program (ODSP), which provides financial assistance and employment support to enable individuals with disabilities and their families to live as independently as possible in their communities. To qualify, generally, an applicant must be 18 years or older, meet a financial/asset threshold, be a legal resident of Canada, and live in Ontario in the geographic area where they applied for SA. In addition, ODSP applicants must meet the definition of a person with a disability as defined by the ODSP Act, 1998 (ODSP Act), or be a member of a "prescribed class" [21]. Unless the requirement is deferred or waived, adults receiving financial assistance under OW, ODSP dependent adults, and ODSP non-disabled spouses without caregiving responsibilities must agree to participate in approved employment assistance activities as a

condition of eligibility for assistance (e.g., job search). OW and ODSP also provide additional benefits (e.g., prescription drug coverage). Monthly financial assistance is paid to families or households, known as "benefit units" (BU), which include the SA applicant, spouse, and dependents residing with them (if any). In most cases, each record in the MCCSS database represents a month where SA was received (i.e., the individual or family was eligible to receive SA), with monthly records listed for all members of the BU.

In November 2014, the Social Assistance Management System (SAMS) replaced the Service Delivery Model Technology (SDMT) as the technology supporting the administration of social assistance in Ontario. With this change, some variables were either added or removed, and coding practices were modified. In December 2018, records on SA recipients in Ontario from 1st January 2003 to 31st December 2016 were transferred to ICES using multiple files grouped into broad categories: BU characteristics; characteristics of members of a BU; pay/income details (types and amounts covering a specific month); OW-specific variables (e.g., job-search activities); and ODSP-specific variables (e.g., disability indicator and associated diagnosis). To enable seamless linkage and analyses, ICES developed the MCCSS-SA standalone dataset, which contains a minimum set of variables required for analtytic purposes, including member characteristics (sex, age, marital status, member role - applicant/spouse/dependent), BU-level characteristics (family size and composition, postal code of residence and accommodation status), and administrative details, including program (OW and/or ODSP), and the amount of monthly financial assistance provided. Figure 1 describes the steps taken in the data pre-processing linkage process for the MCCSS-SA input file.

Record linkage methods

Spine-based record linkage involves matching records in a database to records in a population registry (i.e., the spine), and creating a unique encoded identifier. The spine-based record linkage model at ICES follows the Fellegi-Sunter method [22]. There are two common types of record linkage methods: deterministic linkage and probabilistic linkage. Deterministic linkage consists of exact matching on a single field (e.g., health card number), or a combination of fields, and typically yields about 70%–85% matches. When unique identifiers are not available in the data or deterministic record linkage is not possible, probabilistic record linkage may be used to obtain additional matches [23, 24]. Probabilistic linkage estimates the likelihood that two records belong to the same individual and is based on probability theory; it typically contributes about 10%-20% of matches.

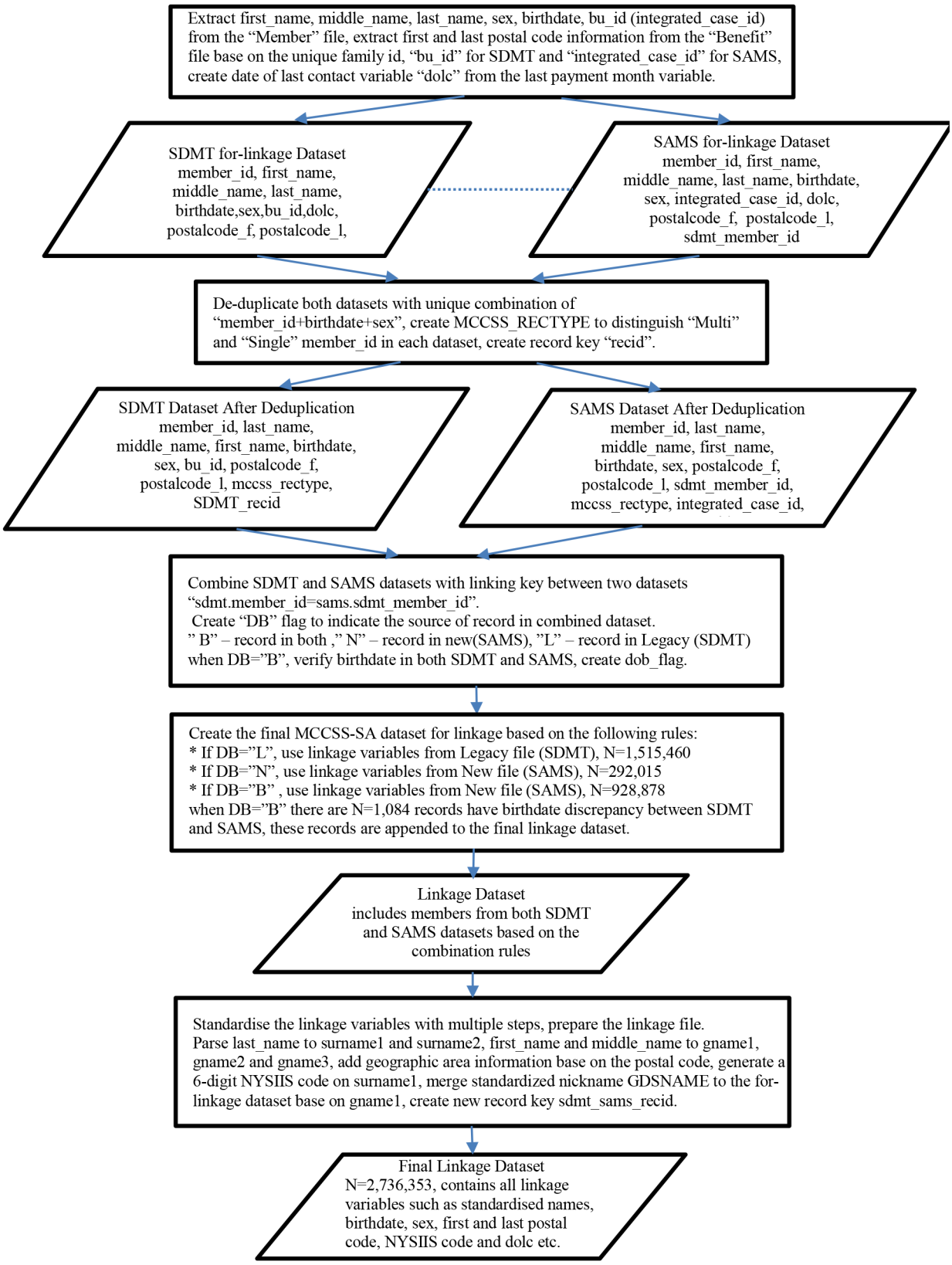
The Registered Persons Database is a population-based registry [25], which includes information on every unique individual ever assigned a health card number in Ontario, containing records for over 14 million individuals. The Registered Persons Database also contains data on demographics and personally identifiable information (e.g., surname, given names, sex, date of birth, earliest date of coverage, last date of contact with the health care system and residential postal code), which enables linkage across data holdings in the ICES data repository. To undertake data linkage, two types of variables are typically used: blocking variables, which consist of data fields that limit the number of comparisons by examining only records agreeing exactly on a given value of a blocking variable, and matching variables, which are those with common fields in both datasets and are used for comparing outcomes (e.g., agreement, disagreement, and partial agreements). Comparison outcomes contribute weights (agreements generate positive weight scores; disagreements generate negative weight scores), where the higher the weight, the higher the likelihood the record pair belongs to the same person. The success of record linkages is dependent on the quality of the individual data sources and identifiers as well as the accuracy of the record linkage process, which many times involves manual review. The goal is to reduce the number of mismatches and unlinked records and, in turn, reduce the potential for biases [26, 27], which may be created through the exclusion of unlinked records from study analyses and impact representativeness. The importance of reporting record linkage results has been highlighted in the RECORD reporting guidelines for studies using administrative health data [28].

The MCCSS-SA dataset was linked to the Registered Persons Database using a "many to 1" hybrid linkage matching approach [29], which allows multiple MCCSS-SA records to match to the same health card number using first a deterministic linkage approach followed by a probabilistic linkage approach. Surname, first and second given names, sex, date of birth (including day, month and year), date of death (where applicable), and residential postal code were used as blocking and matching variables. In some cases, extracted personal identifiers were used to match with additional data standardisation of surnames to increase record linkage rates through the implementation of the New York State Identification and Intelligence System phonetic conversion [30]. The flowchart in Figure 2 provides a description of the linkage process.

Statistical analysis

After record linkage between the MCCSS-SA data and the Registered Persons Database was completed, health card numbers, retrieved from the record linkage process, were encoded as ICES key numbers (IKNs) and all direct personal identifiers (e.g., names, health card numbers, addresses) were removed to produce a less identifiable dataset. To calculate linkage rates, we examined the number of records linked by deterministic and probabilistic record linkages in each step of the process, as well as the linkage rates over time. If a member identifier in the MCCSS-SA dataset was attached to multiple records in the latest month, the first applicant record was kept, followed by the spouse, the dependent adult, and finally the dependent child. If there was more than one record with the same member role, the ODSP record was prioritised. Cases where an ICES unique identifier (i.e., IKN) could not be attached to the record through linkage to the Registered Persons Database were considered unlinked. To assess the representativeness of the linked dataset, we then examined the socio-demographic and program-specific characteristics (contained in the minimum MCCSS-SA dataset, as described above) of linked versus unlinked individuals. Given the very large sample sizes, p-values

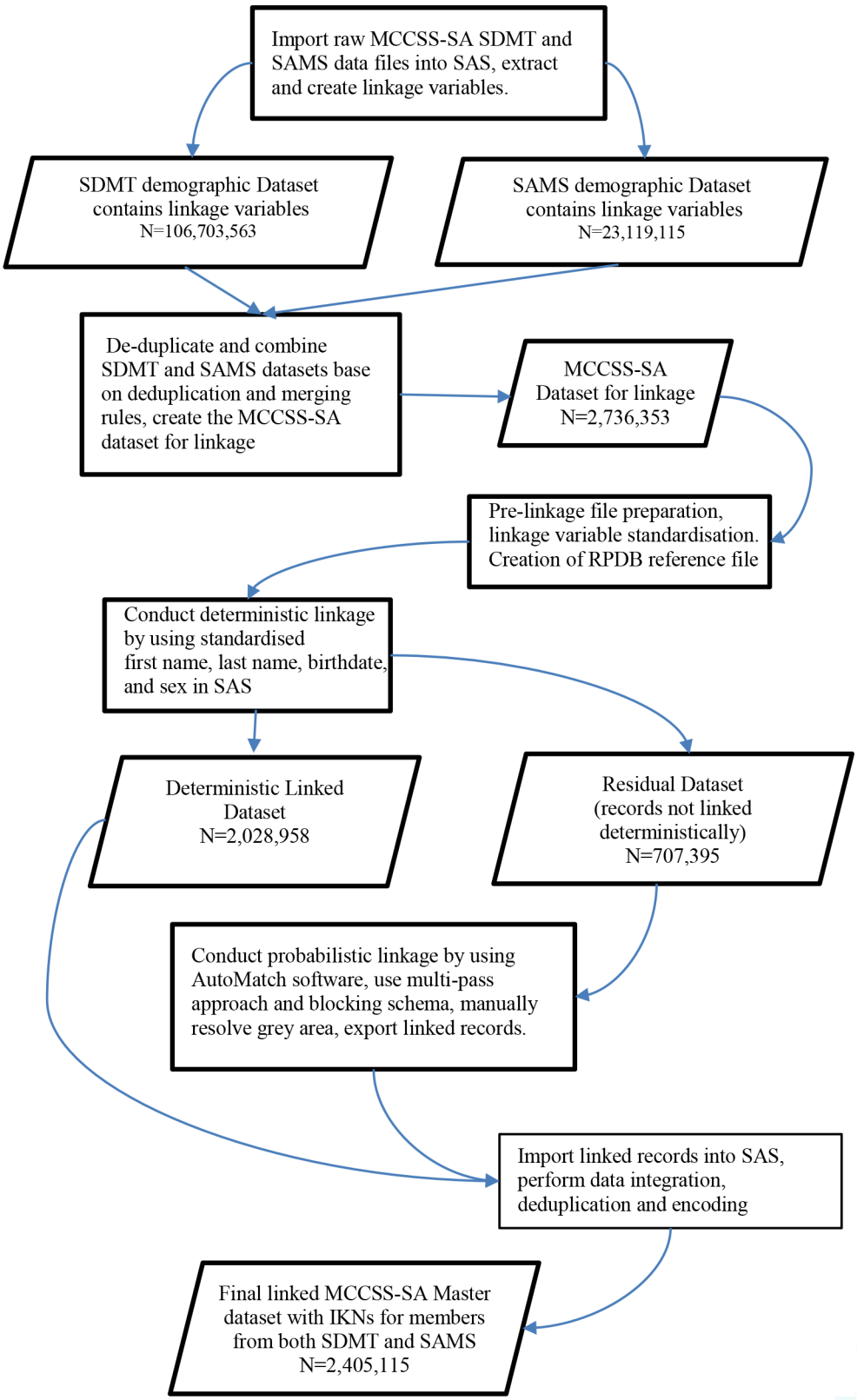
Figure 1: Pre-processing linkage steps for the Ministry of Children, Community and Social Services – Social Assistance input file



Legend: bu – benefit unit; SDMT – Service Delivery Model Technology; SAMS – Social Assistance Management System; NYSIIS – New York State Identification and Intelligence System; dolc – date of last contact; dob – date of birth.

were not used for statistical testing; instead, prevalence estimates between the linked and unlinked samples were compared using standardized differences to assess systematic bias as suggested by Cohen [31], with 0.2, 0.5 and 0.8 representing small, moderate and large standardized differences, respectively.

Figure 2: Linkage process for Ministry of Children, Community and Social Services – Social Assistance data



Legend: MCCSS – Ministry of Children, Community and Social Services; SA – Social Assiatance; SDMT – Service Delivery Model Technology; SAMS – Social Assistance Management System; RPDB – Registered Persons Database; IKN – ICES key number.

Results

There were a total of 2,736,353 unique member IDs in MCCSS-SA dataset from the 1st January 2003 to 31st December 2016, where 331,238 (12.1%) unique member IDs were unlinkable, for a total linkage rate of 87.9% (Table 1). Three quarters (76.2%) of records were obtained through deterministic linkage while the remaining 11.7% of records

Table 1: Socio-demographic and program-specific characteristics of linked and unlinked individuals in the Ministry of Children, Community and Social Services-Social Assistance – Registered Persons Database linkage (1st January 2003 – 31st December 2016)

Socio-demographic and program-specific characteristics	Linked sample N	Linked sample %	Unlinked sample N	Unlinked sample %	Standardized difference
Overall	2,405,115	87.9	331,238	12.1	N/A
Program					
Ontario Works	1,630,744	67.8	254,119	76.7	0.20
Ontario Disability Support Program	774,371	32.2	77,119	23.3	0.20
Member role					
Applicant	1,433,505	59.6	149,495	45.1	0.29
Spouse	195,181	8.1	29,156	8.8	0.02
Dependent adult	120,574	5.0	22,782	6.9	0.08
Dependent child	655,855	27.3	129,805	39.2	0.26
Sex					
Male	1,210,680	50.3	127,502	38.5	0.24
Female	1,154,547	48.0	137,660	41.6	0.13
Unknown	39,888	1.7	66,076	19.9	0.62
Age					
Mean (SD)	31.01 ± 19.47		25.39 ± 18.17		0.30
Median (IQR)	29 (16-47)		22 (11-38)		0.30
Migrant status					
N/A (Canadian-born and long-term residents)	1,731,689	72.0	184,830	55.8	0.34
All other (immigrants and refugees)	673,426	28.0	146,408	44.2	0.34
Rural dwelling					
Yes	216,878	9.0	21,109	6.4	0.10
No	2,168,548	90.2	306,977	92.7	0.09
Missing	19,689	0.8	3,152	1.0	0.01
Family composition					
Single without children	1,000,286	41.6	99,077	29.9	0.25
Single with children	782,381	32.5	124,509	37.6	0.11
Couples without children	161,980	6.7	19,457	5.9	0.04
Couples with children	460,468	19.1	88,195	26.6	0.18
Accommodation status					
Homeless	20,785	0.9	3,767	1.1	0.03
Not homeless	2,384,330	99.1	327,471	98.9	0.03
Data entry system					
In SDMT only: January 2003 – October 2014	1,260,419	52.4	255,041	77.0	0.53
In SAMS only: November 2014 – December 2016	263,941	11.0	28,074	8.5	0.08
In both systems	880,755	36.6	48,123	14.5	0.52
Number of months on social assistance					
Mean (SD)	49.57 ± 50.02		30.70 ± 37.02		0.43
Median (IQR)	29 (10-77)		17 (6-38)		0.40

Legend: N/A – not applicable; SDMT – Service Delivery Model Technology; SAMS – Social Assistance Management System; SD – standard deviation; IQR – interquartile range.

was obtained through probabilistic linkage (Table 2). The vast majority of records were obtained after two deterministic passes (76.2%) and fourteen probabilistic passes (11.7%). In total, sixteen passes were required to obtain the total number of records (2,405,115), using a variety of matching and block schemes (e.g., surnames, given names, sex, date of birth, residential postal code).

On average, standardized differences between linked and unlinked samples were less than 0.1 for most socio-demographic characteristics (i.e., sex, age and rural dwelling), except migrant status (standardized difference of 0.52) (Table 1). The individuals that were successfully linked were different from those that were not for program-specific characteristics, such as program, where there were more individuals in the unlinked group enrolled in the OW and less enrolled in the ODSP compared to the linked group (standardized difference of 0.20 for each), data entry system, SDMT only, or both SDMT and SAMS (standardized difference of 0.53 and 0.52, respectively), and months on social assistance (standardized difference of 0.43) (Table 1). Compared to linked member ID, unlinked individuals were more likely to be on OW, in the SDMT system only, and to have a shorter duration of social assistance.

Table 2: Deterministic and probabilistic linkage schema used to link the Ministry of Children, Community and Social Services – Social Assistance database to the Registered Persons Database

Type	Total number of records	Linkage type		Total number of linked records
		Deterministic	Probabilistic	
Unique member ID SDMT + SAM	2,736,353 (100%)	2,083,864 (76.2%)	321,251 (11.7%)	2,405,115 (87.9%)

Pass #	Linkage Type (D = Deterministic P=Probabilistic)	Number of records linked		Matching and blocking variables
		Males	Females	
1	D	1,071,584	983,389	Matching on: Surname 1 + Given Name 1 + Sex + DOB, Alternate with Given Name 2 (RPDB) and Standardized Given Name (MCCSS)
2	P	57,711	52,265	Blocking on: Surname 1 first-3 characters + Given Name 1 first-3 characters + DOB + Sex Matching on: Surname 1 + Given Name 1 + Given Name 2 + Given Name 3
3	P	25,625	25,460	Blocking on: Surname 1 initial + Given Name 1 initial + DOB + Sex Matching on: Surname 1 + Standardized Given Name (MCCSS)/Given Name 1 (RPDB) + Given Name 2 + Given Name 3
4	P	11,782	10,631	Blocking on: DOB + Sex + Surname 1 initial Matching on: Surnames + Given Names + Postal Codes
5	P	10,753	10,133	Blocking on: Surname 1 initial + Given Name 1 initial + Birth Year + Sex Matching on: Surnames + Given Names + Birth Month + Birth Day + Postal Codes
6	P	20,814	2,435	Blocking on: Surname 1 initial + Given Name 1 initial + Birth Month + Birth Day + Sex Matching on: Surnames + Given Names + Birth Year + Postal Codes
7	P	6,478	575	Blocking on: NYSIIS code of Surname 1 + Birth Year + Sex Matching on: Surnames + Given Names + Birth Month + Birth Day + Postal Codes
8	D	14,439	14,452	Matching on: DOB + Surname 1 + Given Name 1
9	P	5,240	3,298	Blocking on: DOB + Surname 1 initial + Given Name 1 initial Matching on: Surnames + Given Names +Postal Codes
10	P	2,784	67,136	Blocking on: DOB + Sex Matching on: Surnames + Given Names + Postal Codes
11	P	1,220	394	Blocking on: Birth Year + Sex Matching on: Surnames + Given Names + Birth Month + Birth Day + Postal Codes
12	P	546	293	Blocking on: Birth Month + Birth Day + Sex Matching on: Surnames + Given Names + Birth Year + Postal Codes
13	P	1,972	364	Blocking on: Surname 2 initial (MCCSS)/Surname 1 initial (RPDB) + DOB Matching on: Surname 2 (MCCSS)/Surname 1 (RPDB) + Given Names + Postal Codes
14	P	51	37	Blocking on: Surname 2 Initial + DOB Matching on: Surnames + Given Names + Postal Codes
15	P	1,986	898	Blocking on: Birth Year + Given Name 1 initial + NYSIIS code of Surname 1 Matching on: Surnames + Given Names + Birth Month + Birth Day + Postal Codes

Continued.

Table 2: Continued

Pass #	Linkage Type (D = Deterministic P=Probabilistic)	Number of records linked		Matching and blocking variables
		Males	Females	
16	P	258	112	Blocking on: Birth Month + Birth Day + Given Name 1 initial + NYSIIS code of Surname 1 Matching on: Surnames + Given Names + Birth Year + Postal Codes
Linked Total		1,233,243	1,171,872	
		2,405,115 (87.9%)		

Legend: SAMS – Social Assistance Management System; SDMT – Service Delivery Model Technology; DOB – date of birth; RPDB – Registered Persons Databas; MCCSS – Ministry of Children, Community and Social Services; NYSIIS – New York State Identification and Intelligence System.
Notes: Surnames – Array variable of surname; element contains Surname 1 and Surname 2.
Given Names – Array variable of given name; element contains Given Name 1, Given Name 2 and Given Name 3.
Postal Codes – Array variable of postal code; element contains member's first historic postal code and most recent postal code.
Standardized Given Name – standardized nickname from Given Name 1.

The deterministic linkage rates within the SDMT system ranged from 74.2% in 2003 to 79.8% in 2014, while these rates in the SAMS system were 79.3% in 2015 and 77.8% in 2016 (Figure 3). The probabilistic linkage rates within the SDMT system ranged from 11.6% in 2009 to 12.2% in 2003 while in the SAMS system they were 11.4% in 2015 and 11.2% in 2016 (Figure 3). The proportion of unlinked records was typically higher in the SDMT system (Figure 3).

We also looked at these differences by program (OW/ODSP) and by year; the differences were quite consistent (see Appendix Tables A1, A2 in the Appendix).

Discussion

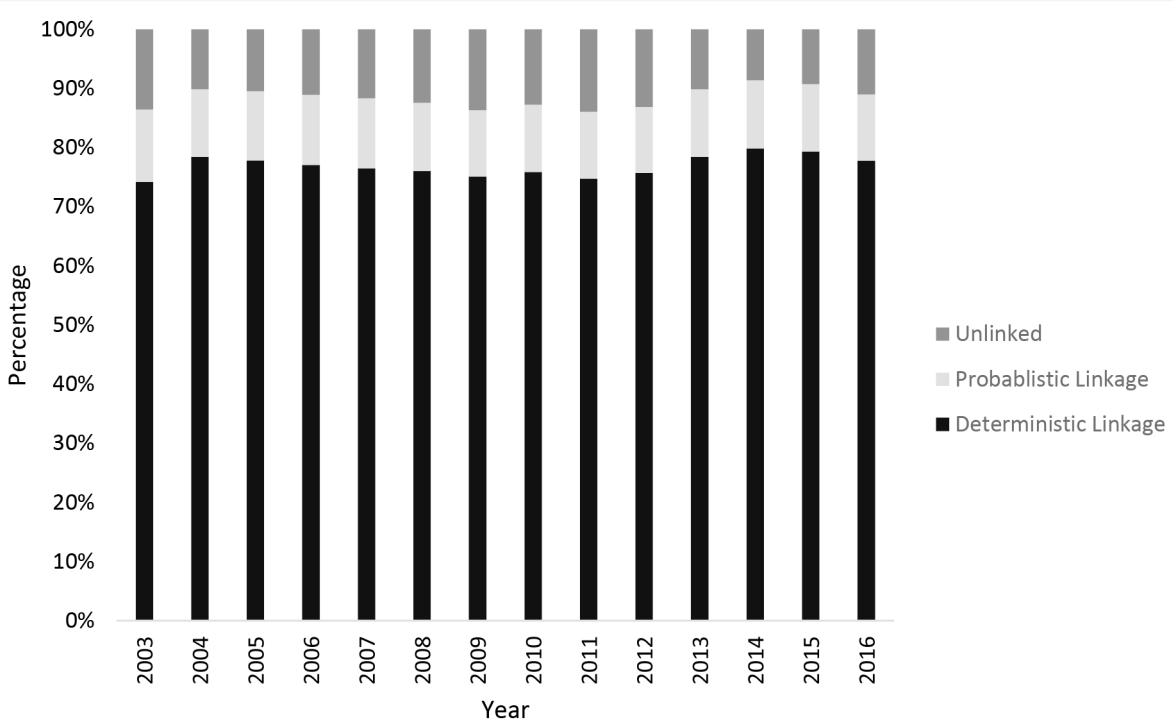
The linkage between the MCCSS-SA and the Registered Persons Database resulted in a high number of MCCSS-SA participants successfully linked to health administrative records (linkage rate = 87.9%) and the linked sample was reasonably representative of the original MCCSS-SA sample. However, there were large differences in the linked and unlinked samples with regard to migrant status such that there was a greater proportion of migrants (i.e., immigrants and refugees) in the population receiving SA that were unable to be linked. There were smaller differences in the linked and unlinked samples by program and system entry characteristics as well as by duration on social assistance, such that the OW records were more likely to be unlinked as was the case for individuals with SDMT records only, and unlinked individuals were on social assistance for a shorter period of time.

The difference in the linked and unlinked samples for migrants may be due to a number of reasons. For example, some refugees, such as asylum seekers, are eligible for social assistance prior to obtaining permanent resident status and provincial health care benefits; this could explain why these individuals had a record in the MCCSS-SA data but not in the ICES data. This difference could also be due to cases where names were recorded incorrectly. The OW records were more likely to be unlinked likely because these records are of shorter duration. This may also be related to the fact that shorter spells of SA had poorer linkage. In addition, individuals with SDMT records only were more likely to be unlinked. This is not surprising given that the SDMT, the old data entry system, was likely not as good at recording information on the variables used for linkage; for example, it includes more free text fields, which can introduce recording errors. In terms of the SA program variables, while there was not complete representativeness, to date there has been more interest in individuals with developmental disabilities [32, 33] and the ODSP data, which provides information on SA provided to these individuals. The extent to which these differences may introduce selection bias will depend on the specific question of study and the specific population generated from the database. The high linkage rate may help mitigate against some of the impacts of these biases, though this will be dependent on the amount of bias and how focused it is. Thus, it is important to acknowledge the potential for selection bias and researchers should seek to address this in their analyses, if/where applicable.

The linkage between the Registered Persons Database and the MCCSS-SA data linkage has a few limitations. The data transferred to ICES are currently only available from 2003 onwards, which limits the how far back researchers can go to examine SA. We were not able to examine the representativeness of all variables (e.g., education), as these were not entered reliably in the SA data (they are not mandatory for entry). Furthermore, given the switch in data systems, researchers should be aware that some comparisons over time may not be possible, in particular when undertaking longitudinal analyses, which cover the transition year (i.e., from 2014 to 2015). Finally, while there are currently limitations on the use of the SA data, there are proposed legislative changes that will enable their broader use in Ontario [34].

Nonetheless, despite these limitations, this high linkage rate will enable scientists to examine one of the many social determinants of health and answer a series of questions that have not been possible until now. Few jurisdictions have been able to undertake this type of data linkage; examples include the Canadian province of Manitoba [35, 36], Scotland [37] and Scandinavian countries, such as Sweden [38]. Future work will

Figure 3: Deterministic linkage, probabilistic linkage and unlinked rates and percentage of unlinked records for the Ministry of Children, Community and Social Services – Social Assistance by year (2003–2016)



seek to examine the health service use of SA clients and their characteristics using ICES data as well as explore the inclusion of additional MCCSS variables into the SA minimum dataset available to researchers.

Conclusion

We found a high linkage rate between the MCCSS-SA and ICES administrative health care databases; furthermore, the linkage was fairly representative of the population of social assistance recipients. However, given sub-optimal linkage rates for migrant recipients of social assistance, there is potential for selection bias. Nonetheless, this linkage represents a significant advancement in understanding the social determinants of health and will enable scientists to answer relevant research questions in the future, recognising the limitations of the data.

Acknowledgments

This research was supported by ICES, which is funded by an annual grant from the Ontario Ministries of Health (MOH) and the Ministry of Long-Term Care (MLTC). Parts of this material are based on data and information compiled and provided by the MOH. The analyses, conclusions, opinions, and statements expressed herein are solely those of the authors and do not reflect those of the funding or data sources. No endorsement by ICES, the Ontario Ministry of Health, the Ministry of Long-Term Care or the Ontario Ministry of Children, Community and Social Services is intended or should be inferred.

Conflicts of interest

None to be declared.

Ethics statement

The use of data in this project was authorised under section 45 of Ontario's Personal Health Information Protection Act and, as a result, informed consent and ethics review was not required.

References

1. Personal Health Information Protection Act, 2004, S.O. 2004, c. 3, Sched. A. Ontario. 2020. https://www.ontario.ca/laws/statute/04p03 (Accessed 29 April 2021).

2. Toulany A, Stukel TA, Kurdyak P, Fu L, Guttmann A. Association of Primary Care Continuity With Outcomes Following Transition to Adult Care for Adolescents With Severe Mental Illness. *JAMA Netw Open*. 2019;2(8):e198415. https://doi.org/10.1001/jamanetworkopen.2019.8415.

3. de Oliveira C, Cheng J, Rehm J, Kurdyak P. The Economic Burden of Chronic Psychotic Disorders in Ontario. *J Ment Health Policy Econ*. 2016; 19(4):181–192.

4. O'Neill M, Buajitti E, Donnelly PD, Lewis J, Kornas K, Rosella LC. Characterising risk of homicide in a population-based cohort. *J Epidemiol Community Health*.

2019; 74(12):1028–1034. https://doi.org/10.1136/jech-2019-213249.

5. Manuel DG, Perez R, Sanmartin C, Taljaard M, Hennessy D, Wilson K, Tanuseputro P, Manson H, Bennett C, Tuna M, Fisher S, Rosella LC. Measuring Burden of Unhealthy Behaviours Using a Multivariable Predictive Approach: Life Expectancy Lost in Canada Attributable to Smoking, Alcohol, Physical Inactivity, and Diet. *PLoS Med*. 2016; 13(8):e1002082. https://doi.org/10.1371/journal.pmed.1002082.

6. Raphael D. Addressing the social determinants of health in Canada: bridging the gap between research findings and public policy. *Policy Options*. 2003;24(3):35–40.

7. Marmot M, Allen J, Bell R, Bloomer E, Goldblatt P, Consortium for the European Review of Social Determinants of Health and the Health Divide. WHO European review of social determinants of health and the health divide. *Lancet*. 2012;380(9846):1011–1129. https://doi.org/10.1016/S0140-6736(12)61228-8.

8. Ahnquist J, Wamala SP, Lindstrom M. Social determinants of health–a question of social or economic capital? Interaction effects of socioeconomic factors on health outcomes. *Social Science & Medicine*. 2012;74(6):930–939. https://doi.org/10.1016/j.socscimed.2011.11.026.

9. Walker RJ, Smalls BL, Campbell JA, Williams JL, Egede LE. Impact of social determinants of health on outcomes for type 2 diabetes: a systematic review. *Endocrine*. 2014;47(1):29–48. https://doi.org/10.1007/s12020-014-0195-0.

10. Naper SO. All-cause and cause-specific mortality of social assistance recipients in Norway: a register-based follow-up study. *Scand J Public Health*. 2009;37(8):820–825. https://doi.org/10.1177/1403494809347023.

11. Martikainen P, Makela P, Koskinen S, Valkonen T. Income differences in mortality: a register-based follow-up study of three million men and women. *Int J Epidemiol*. 2001;30:1397–405. https://doi.org/10.1093/ije/30.6.1397.

12. Tarasuk V, Vogt J. Household food insecurity in Ontario. *Can J Public Health*. 2009;100(3):184–188. https://doi.org/10.1007/BF03405537.

13. Jang M, Vorderstrasse A. Socioeconomic Status and Racial or Ethnic Differences in Participation: Web-Based Survey. *JMIR Res Protoc*. 2019;8(4):e11865. https://doi.org/10.2196/11865.

14. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health*. 2011;32:91–108. https://doi.org/10.1146/annurev-publhealth-031210-100700.

15. Mustard CA, Derksen S, Kozyrskyj A. A Description of the Use of Insured Health Care Services by Income Assistance Recipients in the Province of Manitoba: A Pilot Study. Recipients of Income Assistance for Mental Health Disability. Winnipeg, MB: Manitoba Centre for Health Policy and Evaluation; 2000.

16. Brownell M, Chartier M, Au W, Schultz J. Evaluation of the Healthy Baby Program. Winnipeg, Canada: Manitoba Centre for Health Policy; 2010.

17. Karmel R, Anderson P, Gibson D, Peut A, Duckett S, Wells Y. Empirical aspects of record linkage across multiple data sets using statistical linkage keys: The experience of the PIAC cohort study. *BMC Health Serv*. Res. 2010;10:41. https://doi.org/10.1186/1472-6963-10-41.

18. Holman CDAJ, Bass JA, Rosman DL et al. A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system. *Aust. Health Rev*. 2008; 32: 766–777. https://doi.org/10.1071/ah080766.

19. Chiu M, Lebenbaum M, Lam 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*. 2016;16(1):135. https://doi.org/10.1186/s12911-016-0375-3.

20. Walker J, Pyper E, Jones CR, et al. Unlocking First Nations health information through data linkage. *Int J Popul Data Sci*. 2018;3(1):450. https://doi.org/10.23889/ijpds.v3i1.450.

21. Ministry of Children, Community and Social Services. Ontario Disability Support Program - Income Support. 2019. https://www.mcss.gov.on.ca/en/mcss/programs/social/directives/odsp/is/1_2_ODSP_ISDirectives.aspx (Accessed 29 April 2021).

22. Schull MJ, Azimaee M, Marra M, et al. ICES: Data, Discovery, Better Health. *Int J Popul Data Sci*. 2020;4(2):1135. https://doi.org/10.23889/ijpds.v4i2.1135.

23. Méray N, Reitsma JB, Ravelli ACJ, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *J. Clin. Epidemiol*. 2007;60:883–891. https://doi.org/10.1016/j.jclinepi.2006.11.021.

24. Harron K, Dibben C, Boyd J et al . Challenges in administrative data linkage for research. *Big Data Soc*. 2017;4: 2053951717745678. https://doi.org/10.1177/2053951717745678.

25. Government of Ontario. Registered Persons Database (RPDB). 2017. https://data.ontario.ca/dataset/registered-persons-database-rpdb (Accessed 29 April 2021).

26. Baldi I, Ponti A, Zanetti R, Ciccone G, Merletti F, Gregori D. The impact of record inkage bias in the Cox model. *J Eval Clin Pract*. 2010;16(1):92–96. https://doi.org/10.1111/j.1365-2753.2009.01119.x.

27. Bohensky MA, Jolley D, Sundararajan V, Evans S, Pilcher DV, Scott I, et al. Data linkage: a powerful research tool with potential problems. *BMC Health Serv Res*. 2010;10:346. https://doi.org/10.1186/1472-6963-10-346.

28. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med*. 2015;12(10):e1001885. https://doi.org/10.1371/journal.pmed.1001885.

29. Christen P, Goiser K. Quality and Complexity Measures for Data Linkage and Deduplication. In: Guillet F.J., Hamilton H.J. (eds) Quality Measures in Data Mining. Studies in Computational Intelligence, vol 43. 2007. Springer, Berlin, Heidelberg.

30. Taft R. Name search techniques: New York state identification and intelligence system. Albany: Special Rep; 1970.

31. Cohen J. Statistical power analysis for the behavioral sciences. New Jersey: Hillsdale; 1988.

32. Lunsky Y, Klein-Geltink JE, Yates EA, eds. *Atlas on the Primary Care of Adults with Developmental Disabilities in Ontario*. Toronto, ON: Institute for Clinical Evaluative Sciences and Centre for Addiction and Mental Health; 2013.

33. Lin E, Balogh RS, Durbin A, Holder L, Gupta N, Volpe T, Isaacs BJ, Weiss JA, Lunsky Y. Addressing Gaps in the Health Care Services Used by Adults with Developmental Disabilities in Ontario. Toronto, ON: ICES; 2019.

34. Ministry of Government and Consumer Services. Ontario Public Service Data Integration Data Standards. 2021. https://www.ontario.ca/page/ontario-public-service-data-integration-data-standards (Accessed 7 February 2022).

35. Roos NP, Roos LL, Freemantle J. Administrative data and the manitoba centre for health policy: some reflections. Données administratives au Centre des politiques de santé du Manitoba: réflexions. *Healthc Policy*. 2011;6(Spec Issue):16–28.

36. Jutte DP, Roos LL, Brownell MD. Administrative Record Linkage as a Tool for Public Health Research. *Annual Review of Public Health*. 2011;32:91–108. https://doi.org/10.1146/annurev-publhealth-031210-100700.

37. Katikireddi SV, Leyland A. Establishing data linkage between welfare and health data in the UK: Overcoming barriers to linking government datasets. *IJPDS*. 2017;1(1):170, Proceedings of the IPDLN Conference (August 2016).

38. Dackehag M, Ellegård LM, Gerdtham UG, Nilsson T. Social assistance and mental health: evidence from longitudinal administrative data on pharmaceutical consumption. *Applied Economics*. 2020:52(20):2165–2177.

Abbreviations

MCCSS:	Ministry of Children, Community and Social Services
SA:	Social Assistance
OW:	Ontario Works
ODSP:	Ontario Disability Support Program
SAMS:	Social Assistance Management System
SDMT:	Service Delivery Model Technology
IKN:	ICES Key Number

Appendix

Appendix Table 1: Deterministic and probabilistic linkage by social assistance program

Socio-demographic and program-specific characteristics	Ontario works		Ontario disability Support program		Standardized difference
	N	%	N	%	
Overall	1,884,863	68.9%	851,490	31.1%	
Linkage result					
Linked	1,630,743	86.5%	774,372	90.9%	0.14
Unlinked	254,120	13.5%	77,118	9.1%	
Member role					
Applicant	1,024,361	54.3%	558,639	65.6%	0.23
Spouse	132,293	7.0%	92,044	10.8%	0.13
Dependent adult	73,392	3.9%	69,964	8.2%	0.18
Dependent child	654,817	34.7%	130,843	15.4%	0.46
Sex					
Male	923,035	49.0%	415,147	48.8%	0
Female	894,797	47.5%	397,410	46.7%	0.02
Unknown	67,031	3.6%	38,933	4.6%	0.05
Age					
Mean (SD)	25.40 ± 16.89		41.25 ± 20.12		0.08
Median (IQR)	24 (11-37)		45 (22-59)		0.81
Migrant status					
N/A (Canadian-born and long-term residents)	617,605	32.8%	202,229	23.8%	0.20
All other (immigrants and refugees)	1,267,258	67.2%	649,261	76.2%	
Rural dwelling					
Yes	136,671	7.3%	101,316	11.9%	0.16
No	1,730,522	91.8%	745,003	87.5%	0.14
Missing	17,670	0.9%	5,171	0.6%	0.04
Family composition					
Single without children	660,950	35.1%	438,413	51.5%	0.34
Single with children	763,301	40.5%	143,589	16.9%	0.54
Couples without children	67,648	3.6%	113,789	13.4%	0.36
Couples with children	392,964	20.8%	155,699	18.3%	0.06
Accommodation status					
Homeless	20,908	1.1%	3,644	0.4%	0.08
Not homeless	1,863,955	98.9%	847,846	99.6%	
Data entry system					
In SDMT only: January 2003 – October 2014	1,202,568	63.8%	312,892	36.7%	0.56
In SAMS only: November 2014 – December 2016	228,720	12.1%	63,295	7.4%	0.16
In both system	453,575	24.1%	475,303	55.8%	0.69
Number of months on social assistance					
Mean (SD)	30.35 ± 34.19		84.78 ± 55.61		0.77
Median (IQR)	17 (6-41)		79 (33-136)		1.19

Legend: N/A – not applicable; SDMT – Service Delivery Model Technology; SAMS – Social Assistance Management System; SD – standard deviation; IQR – interquartile range.

Appendix Table 2a: Deterministic and probabilistic linkage by year (2003–2009)

Socio-demographic and program-specific characteristics	2003		2004		2005		2006		2007		2008		2009	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Overall	134,275		120,031		112,699		107,550		101,566		98,168		105,125	
Linkage result														
Linked	98,654	73.5	93,205	77.7	89,698	79.6	89,283	83.0	86,115	84.8	83,157	84.7	89,940	85.6
Unlinked	35,621	26.5	26,826	22.3	23,001	20.4	18,267	17.0	15,451	15.2	15,011	15.3	15,185	14.4
Program														
OW	108,230	80.6	93,564	77.9	84,723	75.2	84,772	78.8	81,795	80.5	78,307	79.8	84,856	80.7
ODSP	26,045	19.4	26,467	22.1	27,976	24.8	22,778	21.2	19,771	19.5	19,861	20.2	20,269	19.3
Member role														
Applicant	59,389	44.2	58,111	48.4	55,992	49.7	56,176	52.2	55,248	54.4	53,546	54.5	59,742	56.8
Spouse	11,517	8.6	10,954	9.1	10,090	9.0	9,815	9.1	8,906	8.8	8,607	8.8	9,893	9.4
Dependent adult	9,429	7.0	8,075	6.7	7,496	6.7	6,475	6.0	6,011	5.9	5,631	5.7	5,161	4.9
Dependent child	53,940	40.2	42,891	35.7	39,121	34.7	35,084	32.6	31,401	30.9	30,384	31.0	30,329	28.9
Sex														
Male	52,259	38.9	51,702	43.1	49,785	44.2	50,150	46.6	48,794	48.0	47,532	48.4	52,587	50.0
Female	51,459	38.3	50,457	42.0	48,886	43.4	49,592	46.1	47,721	47.0	46,462	47.3	48,585	46.2
Unknown	30,557	22.8	17,872	14.9	14,028	12.4	7,808	7.3	5,051	5.0	4,174	4.3	3,953	3.8
Age														
Mean (SD)	25.84 ± 19.22		27.59 ± 19.62		28.26 ± 19.79		28.85 ± 19.85		29.21 ± 19.69		29.34 ± 19.77		29.95 ± 19.66	
Median (IQR)	21 (10-38)		23 (12-41)		24 (12-42)		25 (13-43)		26 (13-43)		26 (13-44)		27 (15-44)	
Migrant status														
N/A (Canadian-born and long-term residents)	45,690	34.0	43,930	36.6	39,595	35.1	39,844	37.0	36,868	36.3	34,471	35.1	34,365	32.7
All other (immigrants and refugees)	88,585	66.0	76,101	63.4	73,104	64.9	67,706	63.0	64,698	63.7	63,697	64.9	70,760	67.3
Rural dwelling														
Yes	11,287	8.4	9,564	8.0	9,698	8.6	9,005	8.4	8,879	8.7	8,604	8.8	9,624	9.2
No	122,212	91.0	109,817	91.5	102,338	90.8	97,767	90.9	91,905	90.5	88,775	90.4	94,598	90.0
Missing	776	0.6	650	0.5	663	0.6	778	0.7	782	0.8	789	0.8	903	0.9
Family composition														
Single without children	36,927	27.5	36,941	30.8	36,183	32.1	36,313	33.8	36,562	36.0	35,535	36.2	40,885	38.9
Single with children	54,074	40.3	44,226	36.8	41,154	36.5	39,115	36.4	36,663	36.1	35,264	35.9	34,226	32.6
Couples without children	8,791	6.5	8,391	7.0	8,021	7.1	7,897	7.3	6,871	6.8	6,669	6.8	7,808	7.4
Couples with children	34,483	25.7	30,473	25.4	27,341	24.3	24,225	22.5	21,470	21.1	20,700	21.1	22,206	21.1
Accommodation status														
Homeless	447	0.3	483	0.4	518	0.5	554	0.5	541	0.5	535	0.5	603	0.6
Not homeless	133,828	99.7	119,548	99.6	112,181	99.5	106,996	99.5	101,025	99.5	97,633	99.5	104,522	99.4
Data entry system														
In SDMT only: January 2003 – October 2014	134,275	100	120,031	100	112,699	100	107,550	100	101,566	100	98,168	100	105,125	100
In SAMS only: November 2014 – December 2016	0	0	0	0	0	0	0	0	0	0	0	0	0	0
In both system	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Number of months on social assistance														
Mean (SD)	6.86 ± 13.10		12.96 ± 11.67		17.75 ± 14.05		20.52 ± 16.15		23.03 ± 19.32		25.82 ± 22.94		25.65 ± 26.01	
Median (IQR)	5 (2-8)		13 (5-18)		16 (6-28)		16 (6-37)		17 (6-40)		17 (6-43)		14 (5-40)	

Legend: OW – Ontario Works; ODSP – Ontario Disability Support Program; N/A – not applicable; SDMT – Service Delivery Model Technology; SAMS – Social Assistance Management System; SD – standard deviation; IQR – interquartile range.

Appendix Table 2b: Deterministic and probabilistic linkage by year (2010–2016)

Socio-demographic and program-specific characteristics	2010		2011		2012		2013		2014		2015		2016	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Overall	123,608		129,726		150,501		145,925		202,361		134,770		1,070,048*	
Linkage result														
Linked	105,594	85.4	111,669	86.1	127,556	84.8	125,155	85.8	175,030	86.5	123,881	91.9	1,006,178	94.0
Unlinked	18,014	14.6	18,057	13.9	22,945	15.2	20,770	14.2	27,331	13.5	10,889	8.1	63,870	6.0
Program														
OW	102,759	83.1	107,447	82.8	123,293	81.9	117,745	80.7	146,965	72.6	110,992	82.4	559,415	52.3
ODSP	20,849	16.9	22,279	17.2	27,208	18.1	28,180	19.3	55,396	27.4	23,778	17.6	510,633	47.7
Member role														
Applicant	71,170	57.6	74,776	57.6	84,615	56.2	81,941	56.2	106,391	52.6	78,285	58.1	687,618	64.3
Spouse	11,490	9.3	11,467	8.8	12,929	8.6	12,446	8.5	15,613	7.7	10,752	8.0	79,858	7.5
Dependent adult	6,098	4.9	6,875	5.3	8,159	5.4	8,936	6.1	15,821	7.8	8,222	6.1	40,967	3.8
Dependent child	34,850	28.2	36,608	28.2	44,798	29.8	42,602	29.2	64,536	31.9	37,511	27.8	261,605	24.4
Sex														
Male	62,710	50.7	65,335	50.4	74,109	49.2	72,659	49.8	101,803	50.3	68,936	51.2	539,821	50.4
Female	56,737	45.9	60,431	46.6	69,144	45.9	68,100	46.7	98,592	48.7	65,832	48.8	530,209	49.6
Unknown	4,161	3.4	3,960	3.1	7,248	4.8%	5,166	3.5%	1,966	1.0%	<=5	0	18	0
Age														
Mean (SD)	29.66 ± 19.26		29.62 ± 19.40		29.08 ± 19.38		29.38 ± 19.53		27.81 ± 18.80		29.27 ± 18.81		32.88 ± 19.09	
Median (IQR)	27 (15-44)		27 (15-44)		26 (14-43)		26 (14-44)		24 (14-40)		26 (15-43)		32 (18-50)	
Migrant status														
N/A (Canadian-born and long-term residents)	42,464	34.4	44,032	33.9	50,266	33.4	47,902	32.8	55,163	27.3	37,335	27.7	267,909	25.0
All other (immigrants and refugees)	81,144	65.6	85,694	66.1	100,235	66.6	98,023	67.2	147,198	72.7	97,435	72.3	802,139	75.0
Rural dwelling														
Yes	10,685	8.6	11,053	8.5	12,702	8.4	12,027	8.2	17,719	8.8	12,170	9.0	94,970	8.9
No	111,944	90.6	117,623	90.7	136,493	90.7	132,624	90.9	182,831	90.3	121,199	89.9	965,399	90.2
Missing	979	0.8	1,050	0.8	1,306	0.9	1,274	0.9	1,811	0.9	1,401	1.0	9,679	0.9
Family composition														
Single without children	49,158	39.8	51,857	40.0	58,692	39.0	57,001	39.1	74,045	36.6	54,903	40.7	494,361	46.2
Single with children	39,183	31.7	42,107	32.5	49,944	33.2	48,838	33.5	73,686	36.4	46,305	34.4	322,105	30.1
Couples without children	8,592	7.0	8,745	6.7	9,838	6.5	9,564	6.6	12,141	6.0	7,790	5.8	70,319	6.6
Couples with children	26,675	21.6	27,017	20.8	32,027	21.3	30,522	20.9	42,489	21.0	25,772	19.1	183,263	17.1
Accommodation status														
Homeless	773	0.6	911	0.7	1,105	0.7	1,361	0.9	1,911	0.9	1,778	1.3	13,032	1.2
Not homeless	122,835	99.4	128,815	99.3	149,396	99.3	144,564	99.1	200,450	99.1	132,992	98.7	1,057,016	98.8
Data entry system														
In SDMT only: January 2003 – October 2014	123,608	100	129,726	100	150,501	100	145,925	100	186,286	92.1	0	0	0	0
In SAMS only: November 2014 – December 2016	0	0	0	0	0	0	0	0	2,113	1.0	32,328	24.0	257,574	24.1
In both system	0	0	0	0	0	0	0	0	13,962	6.9	102,442	76.0	812,474	75.9
Number of months on social assistance														
Mean (SD)	27.30 ± 27.92		31.24 ± 31.00		34.72 ± 34.22		38.65 ± 37.55		46.94 ± 42.44		43.87 ± 42.21		76.10 ± 56.52	
Median (IQR)	16 (6-39)		19 (7-45)		22 (8-49)		25 (9-56)		32 (12-74)		27 (12-63)		65 (26-125)	

Legend: OW – Ontario Works; ODSP – Ontario Disability Support Program; N/A – not applicable; SDMT – Service Delivery Model Technology; SAMS – Social Assistance Management System; SD – standard deviation; IQR – interquartile range.
Note: * This value includes all long-term clients of social assistance in Ontario up until 2016.