International Journal of Population Data Science

Journal Website: www.ijpds.org

Estimating Incident Dementia Cases Not Recorded in Passive Surveillance Data in Ontario: A Capture-Recapture Analysis

Aaron Jones[1,2,*], Yasaman Dolatshahi[1,2], James Lee[2], Margaret Carley[3], Andrew Costa[1,2], Etuini Ma'u[4], and Gary Cheung[4]

Submission History

Submitted:	01/06/2026
Accepted:	01/10/2026
Published:	23/11/2026

[1] Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
[2] ICES, Toronto, Ontario, Canada
[3] Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Ontario, Canada
[4] Department of Psychological Medicine, Faculty of Medical and Health Science, School of Medicine, The University of Auckland, Auckland, New Zealand

Abstract

Introduction
Dementia surveillance systems in Canada rely on routinely-collected health administrative data, in which cases are identified only when individuals access care, receive a diagnosis, and have that diagnosis recorded. Such passive surveillance systems may underestimate the population burden of dementia.

Objectives
To estimate the incidence of dementia cases among older adults in Ontario that are unrecorded in health administrative data over 10 years using capture-recapture methods.

Methods
We conducted a descriptive, population-based retrospective cohort study of Ontario residents aged 65 years and older who on July 1, 2014 had no previously recorded dementia. Incident dementia cases between July 1, 2014 and June 30, 2024 were identified using linked physician billing, hospital discharge, and prescription drug databases. Log-linear capture-recapture models with hierarchical interaction structures estimated total number of incident dementia cases, including cases estimated to be unrecorded in all sources. We calculated observed and model estimated cumulative incidence and the estimated proportion of cases not recorded, overall and by demographic subgroup.

Results
Among 2,032,383 older adults, 273,139 met dementia ascertainment criteria in at least one administrative data source over 10 years, corresponding to an observed cumulative incidence of 13.4% and an incidence rate of 166 per 10,000 person-years. Capture-recapture modelling estimated 472,271 incident cases (95% CI 462,509-482,526), which yields an estimated cumulative incidence of 23.2% (22.8%, 23.7%). Overall, 42.2% (40.9-43.4%) of incident dementia cases were estimated to be unrecorded in the three administrative data sources. The proportion unrecorded was broadly similar across sex, rurality, and neighbourhood income. The unrecorded proportion was 41.6% after incorporating home care and long-term care assessments as a fourth ascertainment source and 54.8% in a secondary three-year analysis.

Conclusions
More than 40% of incident dementia cases among older adults in Ontario may go unrecorded in routine administrative data used for passive surveillance over a 10-year period. Surveillance relying on these administrative sources may underestimate dementia incidence in Ontario, highlighting the need for more comprehensive data approaches for health system planning, policy, and evaluation.

Keywords
dementia, linked population data, public health surveillance, epidemiology

Highlights
- We used linked population-based administrative health data to estimate recorded and unrecorded incident dementia in Ontario.
- Capture-recapture modelling estimated that 42% of incident dementia cases were unrecorded in the three routinely used administrative data sources between 2014 and 2024.
- The proportion unrecorded was broadly similar across sex, rurality, and neighbourhood income quintile.
- Findings suggest that passive dementia surveillance systems may substantially underestimate population needs for care, support, and health system planning.

*Corresponding Author:
Email Address: jonesa13@mcmaster.cap (Aaron Jonese)

https://doi.org/10.23889/ijpds.v11i1.3780
November 2026 © The Authors. Open Access under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/deed.en)

Introduction

Dementia is a clinical syndrome marked by progressive cognitive decline that interferes with daily functioning [1]. The umbrella term encompasses multiple underlying conditions including Alzheimer's disease, vascular dementia, frontotemporal dementia, and other causes [2]. Common symptoms include impairments in memory, language, judgement, mood, and decision-making, which often lead to increasing dependence in everyday activities [2]. Dementia is a major and growing global health challenge [3]. In Canada, over 770,000 individuals are living with dementia as of 2025 and it is estimated that the prevalence will increase to more than 1.7 million by 2050 [4]. The societal and economic burden of dementia is also considerable. The annual cost of dementia in Canada is projected to increase from $40 billion in 2020 to $110 billion in 2050 in large part due to the immense contributions of unpaid care partners [5].

A key challenge in dementia care is that diagnoses are frequently delayed or missed [6]. Missed diagnoses undermine opportunities for timely treatment and care, pose safety risks, contribute to underestimation of disease burden, and hinder health system planning [7, 8]. Systematic reviews have suggested that upwards of 60% of individuals living with dementia in community settings worldwide may be undiagnosed [9]. Missed or delayed diagnoses have been linked to numerous patient and provider factors, including limited dementia education, negative attitudes towards dementia, communication challenges, and barriers to diagnostic imaging services [6]. Individuals living with dementia may not be able to access care, may not seek care, or may not be recognised within the system. Evidence suggests that undiagnosed dementia may disproportionately affect men, younger older adults, and people from disadvantaged socioeconomic backgrounds [6, 9, 10]. Unequal detection of dementia may therefore exacerbate inequalities in dementia care and outcomes and hinder efforts to improve health equity.

In Canada, population-level dementia surveillance is conducted with passive systems such as the Canadian Chronic Disease Surveillance System (CCDSS), which uses linked health administrative databases to identify likely dementia cases based on healthcare encounters [11]. Identification within administrative data relies on a series of steps: an individual must access care, receive a clinical diagnosis, and have that diagnosis recorded in the data. A breakdown at any stage may result in missed identification. Capture-recapture methods have been employed in the literature to estimate the number of cases of dementia that are unrecorded in administrative data sources [12–14]. Adapted from ecological sciences, capture-recapture methods analyse the overlaps across multiple methods of case ascertainment to estimate the total number of cases, including those not recorded in any source [15]. Dementia is amenable to capture-recapture because it may be recorded through several distinct but overlapping healthcare pathways and is not defined from a single diagnostic marker. Recent studies from England, New Zealand, and Australia applied capture-recapture methods to dementia and found that a significant proportion of dementia cases were absent from routine health records, with under-recording varying by socioeconomic strata [16–18]. Although capture-recapture methods have been applied in several international contexts, they have not previously been used to estimate dementia incidence in Canada. The objective of this study was to apply capture-recapture methods to linked, person-level administrative health data from Ontario to estimate the incidence of recorded and unrecorded dementia cases over a 10-year period.

Methods

Study Design and Setting

We conducted a descriptive, population-based retrospective cohort study in Ontario, Canada. Ontario has a universal, publicly funded health care system that provides medically necessary hospital and physician services to eligible residents, with prescription drug coverage for adults aged 65 years and older through the Ontario Drug Benefit program. In Ontario, primary care is the first point of contact for the health system. Primary care providers may refer patients suspected of dementia to specialists for confirmation. Individuals with dementia are frequent users of the acute care system, and dementia may be first recognised in hospital among individuals who had not previously sought or been able to access care in the community.

Population

We identified a cohort of individuals who, on the index date of July 1, 2014, were aged 65 years and older, were eligible for public health insurance, and had not previously been identified as having dementia in health administrative data. The cohort was closed with no new entries permitted over time. A secondary cohort with an index date of July 1, 2021 was identified in the same manner. These dates were chosen as the most recent 10- and 3-year periods for which all necessary data were available.

Data Sources

We used multiple, linked, population-based administrative databases in Ontario. Datasets were linked using unique encoded identifiers and analysed at ICES, formerly the Institute for Clinical Evaluative Sciences. ICES is an independent, non-profit research institute and prescribed entity under Ontario health information privacy law. This legal status permits ICES to collect and analyse health care and demographic data, without individual consent, for health system evaluation, monitoring, planning, and improvement. We accessed physician billings using the Ontario Health Insurance Plan (OHIP) database, hospital admissions in the Discharge Abstract Database (DAD), and medication dispensations in the Ontario Drug Benefit (ODB) database. Clinical assessments for publicly funded home care clients and long-term care residents were extracted from the interRAI Home Care and Continuing Care Reporting System databases (RAI). Population counts, demographics, and health care eligibility information were extracted from the Registered Persons Database. More information on these databases can be found in Supplementary Appendix 1.

Dementia Case Ascertainment

We identified dementia cases during the follow-up period using a validated case ascertainment algorithm that aligns with the CCDSS [19, 20]. In the primary analysis, individuals were identified as having dementia if they met at least one of the following criteria: (1) three physician claims (OHIP) with a dementia diagnostic code over the course of two years, with at least 30 days between individual claims; (2) one hospital admission (DAD) with a dementia diagnosis code; or (3) one dispensation of a dementia-specific medication (ODB). Identification in any one of the three databases (OHIP, DAD, or ODB) was sufficient to meet the case definition. For individuals identified in OHIP, the earliest of the three eligible physician claims was taken as the ascertainment date. For individuals identified in multiple databases, the earliest ascertainment date across all sources was taken as the incident date. All qualifying records that occurred during the observation period were used in the capture-recapture analysis. The lookback to exclude individuals with prevalent dementia on the index date extended to 1996, which was the earliest that all data needed for ascertainment was available.

As a secondary analysis, we incorporated a fourth source of case ascertainment using the interRAI HC and Resident Assessment Instrument-Minimum Data Set 2.0 assessments (RAI) [21]. These assessments are regularly administered to individuals receiving publicly funded home care in the community and residents of long-term care homes. The RAI includes validated items on chronic conditions, including dementia. A complete list of all diagnostic codes, Anatomical Therapeutic Chemical codes, and clinical assessment items can be found in Supplementary Appendix 2. Ascertainment of dementia in administrative data relies on documentation in health care records and may not align with the patient experience of receiving a clinical diagnosis. Accordingly, "incident dementia" in this study refers to first administrative ascertainment during follow-up among individuals without previously recorded dementia and does not necessarily correspond to the clinical onset of dementia.

Covariates

We extracted data on the following covariates at baseline: age in years, sex (male or female), rurality, and neighbourhood income quintile. Rural communities were defined as those with a population size of less than 10,000 [22]. Neighbourhood income quintile was assigned using the Statistics Canada's postal code conversion file, which links residential postal codes to census dissemination areas and calculates area-level household income adjusted for household size. Dissemination areas were ranked within census metropolitan areas, census agglomerations, or other provincial regions and divided into quintiles from lowest to highest income [23].

Observed Dementia Incidence

For observed incidence calculation, individuals were followed from July 1, 2014 until the earliest date of dementia identification in administrative data, death, or the end of follow-up on June 30, 2024. We expressed observed dementia incidence as cumulative incidence (# number of administratively ascertained cases / total individuals in the cohort) and incidence rate (# number of cases / 10,000 person-years at risk) over the 10-year observation period. All incidence rates were crude, not standardised for age or sex. We stratified dementia incidence measures by sex, age (<75, 75+) at baseline, rurality, and neighbourhood income quintile. Similarly we followed the secondary cohort from July 1, 2021 until June 30, 2024, calculating the same measures.

Capture-Recapture Analysis

We used capture-recapture methods to estimate the total number of incident cases over the observation window, including cases observed in the administrative data and cases estimated to be unrecorded in all administrative sources. Individuals identified as incident dementia cases during follow-up were classified by which case ascertainment sources they appeared in during the observation period. For the primary analysis, individuals could appear in one, two, or three sources, creating a set of seven unique observable source membership patterns. We fit a series of log-linear Poisson models to estimate the total number of incident cases. These models use the observed pattern of overlap between sources to estimate the number of individuals expected to be missed by all sources while accounting for dependencies between data sources through varying numbers of interaction terms [24]. Let $h = (h_1, \ldots, h_k)$ represent a source membership pattern and n_h its observed frequency. Assuming $n_h \sim Poisson(\mu_h)$, the model equations are given by:

$$\log(\mu_{\mathrm{h}}) = \beta_0 + \sum_{j=1}^{3} \beta_j h_j + \sum_{j<k} \beta_{jk} h_j h_k$$

The main effects $\beta_j h_j$ represent differences in how likely cases were to be recorded in each source and were included in all models. The interaction terms $\beta_{jk} h_j h_k$ account for non-independence in the likelihood of being recorded in multiple sources. We fit all possible models under a hierarchical interaction structure, meaning that inclusion of a higher-order interaction required inclusion of all corresponding lower-order terms. For the primary analysis this resulted in eight possible models. We ranked the models by Bayesian Information Criterion (BIC) and selected the model with the lowest BIC as the primary model. From this model we estimated the unobserved cell $\hat{n}_{0\ldots0} = \exp\left(\hat{\beta}_0\right)$ with 95% multinomial profile likelihood confidence intervals, which is the estimated number of cases unrecorded in all administrative sources. We added this estimate to the observed case count to produce an estimated total count of incident cases and calculated the proportion of total cases estimated to be unrecorded. We further present an estimated cumulative incidence which uses the estimated total count of cases while retaining the same denominator as the observed incidence. We repeated the capture-recapture analysis within strata of sex, age, rurality, and neighbourhood income quintile. These stratified analyses were conducted using the same interaction structure that was selected by the main analysis. All analyses were performed using the Rcapture package with the closedpMs.t and closedpCI.t functions [25].

In a secondary analysis we included the interRAI assessments as a fourth ascertainment source, which yielded

Table 1: Baseline characteristics of individuals aged 65 years and older without previously recorded dementia Ontario on July 1, 2014

Characteristics	Total n=2,032,383 n(%)	DAD n=66,968 n(%)	ODB n=103,723 n(%)	OHIP n=233,398 n(%)
Age yrs, median (Q1, Q3)	73 (68, 79)	77 (72, 82)	76 (72, 81)	78 (73, 84)
Sex, F	1,097,824 (54)	37,489 (56)	60,667 (58)	140,340 (60)
Rural	259,374 (13)	7,894 (12)	12,384 (12)	27,032 (12)
Neighbourhood income				
Q1 (lowest)	414,335 (20)	15,621 (23)	21,212 (20)	53,257 (23)
Q2	432,219 (21)	15,014 (22)	22,570 (22)	52,144 (22)
Q3	403,394 (20)	13,137 (20)	20,330 (20)	44,724 (19)
Q4	373,483 (18)	11,356 (17)	19,060 (18)	40,316 (17)
Q5 (highest)	401,899 (20)	11,670 (17)	20,309 (20)	42,331 (18)

DAD: Identified during 10-year observation period in the Discharge Abstract Database.
ODB: Identified during 10-year observation period in the Ontario Drug Benefit Database.
OHIP: Identified during 10-year observation period in the Ontario Health Insurance Plan Database.

113 possible models. We also employed Bayesian model averaging for both the three-source and four-source analysis to help account for uncertainty in model selection. Under this approach, incidence estimates from multiple models are combined, weighted by their relative likelihood as derived from BIC [26]. Finally, we also examined a three-year cohort to determine sensitivity of our results to the length of the observation window.

Ethics

ICES is a prescribed entity under Ontario's Personal Health Information Protection Act (PHIPA). Section 45 of PHIPA authorises ICES to collect personal health information, without consent, for the purpose of analysis or compiling statistical information with respect to the management, evaluation, monitoring, allocation of resources, or planning of all or part of the health system. This study used only data collected and held by ICES under section 45 of PHIPA. Under this legal framework, projects that use only ICES data collected under section 45 are exempt from research ethics board review. The use of data in this project was authorised under section 45 of PHIPA and approved by the ICES Privacy and Legal Office.

Results

We identified 2,032,383 individuals without previously recorded dementia aged 65 years or older living in Ontario on July 1, 2014 (Table 1). The median age of the population was 73 years, and 54% were female. Thirteen percent of individuals lived in rural areas. Between July 1, 2014 and June 30, 2024, we identified 273,139 incident dementia cases in the administrative data (Figure 1). Among these observed cases, 66,968 (24.5%) were present in DAD, 103,723 (38.0%) in ODB, and 233,398 (85.5%) in OHIP. Only 19,111 (7.0%) cases were identified in all three sources. Differences in observed characteristics between individuals identified in the DAD, ODB, and OHIP databases were minimal (Table 1).

We observed 273,139 incident dementia cases across the 10-year follow-up, yielding a cumulative incidence of 13.4%. There were 16,468,481 person-years spent at risk with an average follow-up time of 8.1 years, corresponding to an observed incidence rate of 166 per 10,000 person-years(PY) (Table 2). Observed cumulative incidence was substantially higher in the 75+ group compared to <75 (21.8% vs. 7.6%), higher in females than males (14.8% vs. 11.9%), and slightly higher in urban than rural residents (13.6% vs. 12.5%). The observed incidence increased with neighbourhood socioeconomic disadvantage, with a cumulative incidence of 15.1% in the lowest income quintile and 12.2% in the highest income quintile.

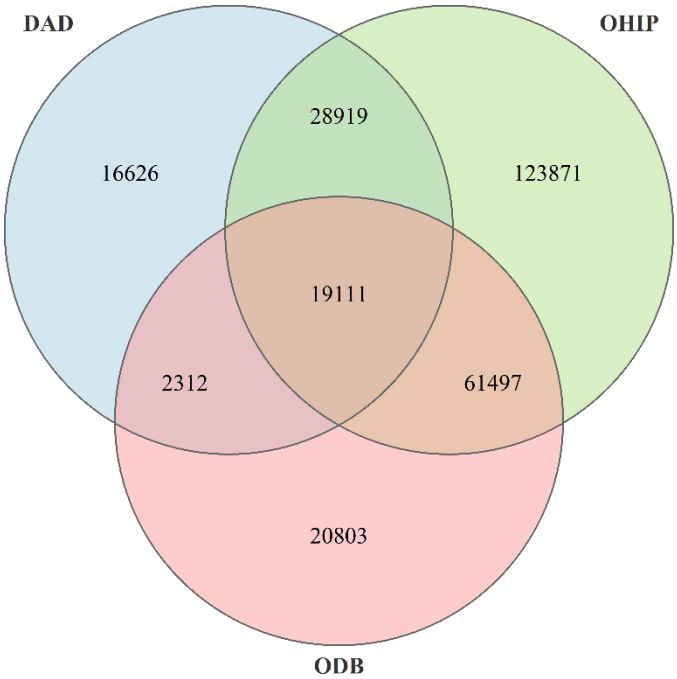
Capture-Recapture Estimates

Using OHIP, DAD, and ODB as three sources of ascertainment, the log-linear model with the lowest BIC was the saturated model with all possible two-way interactions between the three sources (Supplementary Appendix 3). The selected model estimated that there were a total of 472,271 dementia cases (95% CI 462,509, 482,526) compared to the 273,139 observed cases. This corresponds to an estimated cumulative incidence of 23.2% (22.8%, 23.7%) and an unrecorded proportion of 42.2% (40.9%, 43.4%) (Table 3). The estimated proportion unrecorded was slightly higher in males than females (43.8% vs 41.0%) and in urban vs rural residents (42.3% vs 41.6%). Estimated unrecorded dementia was somewhat less common among individuals younger than 75 years compared to those aged 75 years and older (39.3% vs 43.2%). There was no clear gradient in the proportion unrecorded across neighbourhood income quintiles. The highest estimated proportion unrecorded was 44.8% (42.0%, 47.7%) in the third income quintile and the lowest estimated proportion unrecorded was 39.6% (36.7%, 42.5%) in the highest income quintile.

Secondary Analyses

Adding interRAI assessments increased the observed number of incident cases to 308,412, which corresponds to cumulative incidence of 15.2%. The fourth source log-linear model with

Figure 1: Number of incident dementia cases identified within each combination of databases between July 1, 2014 and June 30, 2024



DAD: Discharge Abstract Database.
ODB: Ontario Drug Benefit.
OHIP: Ontario Health Insurance Plan.

Table 2: Observed incidence of dementia in fixed cohort of older adults in Ontario, 2014-2024

Strata	# of individuals at index	# of identified dementia cases	Cumulative incidence (%)	# of person-years of follow-up	Incidence rate per 10,000 person-years
All	2,032,383	273,139	13.4%	16,468,481	166
Female	1,097,824	162,321	14.8%	8,955,768	181
Male	934,559	110,818	11.9%	7,512,713	148
Rural	259,374	32,326	12.5%	2,068,040	156
Urban	1,765,625	240,170	13.6%	14,336,290	168
Age <75	1,194,097	90,681	7.6%	10,718,313	85
Age 75+	838,286	182,458	21.8%	5,750,054	317
Income Q1	414,335	62,423	15.1%	3,208,363	195
Income Q2	432,219	60,924	14.1%	3,453,471	176
Income Q3	403,394	52,668	13.1%	3,280,671	161
Income Q4	373,483	47,258	12.7%	3,082,488	153
Income Q5	401,899	49,140	12.2%	3,374,179	146

the lowest BIC was the model with all possible three-way interactions except the interaction between ODB, OHIP, and RAI (Supplementary Appendix 4). The selected model estimated a total number of cases of 527,716, which corresponds to an estimated cumulative incidence of 26.0% (25.6, 26.4%) and an unrecorded proportion of 41.6% (Table 4).

The secondary three-year cohort with an index date of July 1, 2021 identified 2,612,115 older adults in Ontario without recorded dementia. Using the three ascertainment sources, 96,290 incident dementia cases were identified, yielding a cumulative incidence of 3.7% (Supplementary Appendix 5). The log-linear model with the lowest BIC was again the saturated model with all possible two-way interactions between the three sources. The selected model estimated a total of 213,253 dementia cases (95% CI 204,711, 222,459), yielding an estimated cumulative incidence of 8.2% across the 3-year period and an estimated proportion unrecorded of 54.8%.

Bayesian model averaging had a negligible impact on our results. In the primary analysis, the BIC of the selected model was 170.8 compared to the next lowest of 884.9, indicating substantially better model fit. Given the large difference in BIC, the model averaging procedure gave >99.99% of the weight to the selected model. In the four-source analysis, the selected model had a BIC of 348.8 compared to the next best of 357.4. The next best fitting model had a slightly higher estimated

Table 3: Observed and estimated incident dementia cases, cumulative incidence, and unrecorded proportion using three administrative data sources (OHIP, DAD, ODB), July 1, 2014 to June 30, 2024

Strata	Observed cases	Estimated		
		Total cases (95% CI)	Cumulative Incidence (95% CI)	Proportion unrecorded (95% CI)
All	273,139	472,271 (462,509, 482,526)	23.2 (22.8, 23.7)	42.2 (40.9, 43.4)
Female	162,321	275,271 (267,796, 283,268)	25.1 (24.4, 25.8)	41.0 (39.4, 42.7)
Male	110,818	197,092 (190,923, 203,724)	21.1 (20.4, 21.8)	43.8 (42.0, 45.6)
Rural	32,326	55,384 (52,461, 58,710)	21.4 (20.2, 22.6)	41.6 (38.4, 44.9)
Urban	240,170	416,289 (406,984, 426,104)	23.6 (23.1, 24.1)	42.3 (41.0, 43.6)
Age <75	90,681	149,417 (144,910, 154,280)	12.5 (12.1, 12.9)	39.3 (37.4, 41.2)
Age 75+	182,458	321,287 (312,535, 330,625)	38.3 (37.3, 39.4)	43.2 (41.6, 44.8)
Income Q1	62,423	109,717 (104,985, 114,963)	26.5 (25.3, 27.7)	43.1 (40.5, 45.7)
Income Q2	60,924	105,974 (101,389, 111,068)	24.5 (23.5, 25.7)	42.5 (39.9, 45.1)
Income Q3	52,668	95,415 (90,773, 100,620)	23.7 (22.5, 24.9)	44.8 (42.0, 47.7)
Income Q4	47,258	78,512 (75,023, 82,423)	21.0 (20.1, 22.1)	39.8 (37.0, 42.7)
Income Q5	49,140	81,307 (77,668, 85,395)	20.2 (19.3, 21.2)	39.6 (36.7, 42.5)

Q1 = lowest quintile, Q5 = highest quintile.

Table 4: Observed and estimated incident dementia cases, cumulative incidence, and unrecorded proportion using four administrative data sources (OHIP, DAD, ODB and RAI), July 1, 2014 to June 30, 2024

Strata	Observed cases	Estimated		
		Total cases (95% CI)	Cumulative Incidence (95% CI)	Proportion unrecorded (95% CI)
All	308,412	527,716 (519,561, 536,152)	26.0 (25.6, 26.4)	41.6% (40.6%, 42.5%)
Female	182,927	304,498 (298,626, 310,632)	27.7 (27.2, 28.3)	39.9% (38.7%, 41.1%)
Male	125,485	222,884 (217,288, 228,787)	23.8 (23.3, 24.5)	43.7% (42.2%, 45.2%)
Rural	37,897	65,501 (62,766, 68,497)	25.3 (24.2, 26.4)	42.1% (39.6%, 44.7%)
Urban	269,785	461,603 (453,921, 469,571)	26.1 (25.7, 26.6)	41.6% (40.6%, 42.5%)
Age <75	98,177	179,538 (174,251, 185,161)	15.0 (14.6, 15.5)	45.3% (43.7%, 47.0%)
Age 75+	210,235	350,379 (343,905, 357,127)	41.8 (41.0, 42.6)	40.0% (38.9%, 41.1%)
Income Q1	71,392	124,611 (120,482, 129,046)	30.1 (29.1, 31.1)	42.7% (40.7%, 44.7%)
Income Q2	68,934	118,329 (114,483, 122,464)	27.4 (26.5, 28.3)	41.7% (39.8%, 43.7%)
Income Q3	59,612	101,485 (98,053, 105,188)	25.2 (24.3, 26.1)	41.3% (39.2%, 43.3%)
Income Q4	53,113	93,647 (90,136, 97,456)	25.1 (24.1, 26.1)	43.3% (41.1%, 45.5%)
Income Q5	54,536	89,805 (86,772, 93,090)	22.3 (21.6, 23.2)	39.3% (37.2%, 41.4%)

Q1 = lowest; Q5 = highest.

unrecorded proportion of 44.0%, and with Bayesian model averaging the estimated proportion unrecorded was unchanged at 41.6%.

Discussion

In this population-based study of adults aged 65 years and older in Ontario, 13.4% of the cohort was ascertained as having dementia in the three administrative databases commonly used for passive surveillance over 10 years. Capture-recapture modelling estimated a total cumulative incidence of 23.2%, suggesting that 42.2% of incident cases were unrecorded in all three sources. The estimated proportion unrecorded was broadly similar across sex, rurality, and neighbourhood income and changed little after incorporating home care and long-term care assessments (41.6%). The estimate increased to 54.8% in the three-year analysis, indicating sensitivity to the length of the observation window.

Our findings are broadly consistent with prior capture-recapture studies conducted in England, New Zealand, and Australia, all of which reported substantial under-ascertainment of dementia in routine data sources [16–18]. Our estimate of unrecorded incident dementia was somewhat higher than those reported in other international studies (29%–35%); however, direct comparisons should be made cautiously because of differences in health systems, data sources, follow-up time, and analytic approaches. At the same time, our estimates remain lower than both Canadian and international cohort-based estimates of missing cases [9, 27], which rely on fundamentally different methods of recruitment and ascertainment. Unlike some prior studies we did not detect large differences across the subpopulations we examined. This may be due in part to our lack of individual-level data on

social determinants of health. Also, this analysis is based solely on overlap patterns in ascertainment over a multi-year time frame and thus the experience of receiving, or not receiving, a dementia diagnosis could still vary widely while producing the same overlap patterns.

Our capture-recapture estimates depend on several assumptions, including model form, population closure, and homogeneity of ascertainment probability. As is common practice in the literature, we used an information criterion approach to select the best fitting interaction structure, however this does not guarantee that the model form is correct. We used Bayesian model averaging to examine sensitivity to model selection, which had a negligible impact on our results. Population closure can only be approximated in this study as we must measure incidence over a multi-year period. Individuals who were ascertained near the end of follow-up would have less opportunity to appear in multiple sources, which may reduce overlap and increase estimates of unrecorded dementia. While we took a 10-year approach in the primary analysis to provide time for individuals to appear in multiple sources, we also examined a three-year window. The meaningful differences (42% vs. 55%) indicate that the findings are sensitive to the observation window and should be interpreted accordingly. Finally, there may be heterogeneity in ascertainment probability due to underlying severity of dementia (i.e. mild cases may be more likely to be missed) or other clinical or social factors. The RAI dataset, for example, would typically only include individuals with meaningful functional impairment, meaning that it may be more likely to capture severe dementia cases and miss mild cases. Unaccounted for heterogeneity in capture-recapture analysis typically results in an underestimation of the total number of cases. While we stratified by several sociodemographic factors, there is still likely residual heterogeneity from factors that we did not measure. Ultimately our results should be interpreted as modelling estimates of administratively ascertained dementia and not as precise counts of missed clinical diagnoses.

Our finding that more than 40% of incident dementia cases may be missed by passive surveillance has important implications for public health and health system planning. Incidence measures the flow of new cases into the population and is crucial for monitoring changes over time, identifying population differences, projecting future care needs, and evaluating public and population health interventions. Our study suggests that surveillance based only on recorded cases may substantially underestimate the population burden of dementia and, consequently, the demand for primary and specialist care, home and community services, long-term care, and care-partner support. Incomplete ascertainment may result in under-resourcing that negatively affects the continuum of dementia care with downstream consequences for patients, care partners, and health systems [28, 29].

These results point to the need for enhanced surveillance strategies. More complete surveillance may require moving beyond passive administrative data toward integrated approaches that combine multiple sources of person-level information. One potential strategy is a dedicated dementia registry that can link administrative data with primary care electronic medical records, cognitive screening data, and memory clinic diagnostic data [30]. Improved surveillance would provide a stronger basis for forecasting, resource allocation, policy evaluation, and monitoring progress toward dementia strategies. However, better data infrastructure does not by itself improve clinical recognition or access to care and should complement efforts to address limited dementia awareness, stigma, and barriers to care [31].

Limitations

Our study has several limitations beyond the approximate nature of the estimates produced. First, as baseline dementia status was determined from prior administrative records, some individuals may have had clinically present but administratively unrecorded dementia at the start of follow-up. Additionally, we are unable to reliably identify severity of dementia or subtypes of dementia in our data. Finally, we lacked data on race and ethnic background, which may be relevant to patterns of dementia under-recording.

Conclusion

We estimated that over 40% of incident dementia cases in Ontario between 2014 and 2024 were likely not captured in routine surveillance systems. These findings add to growing international evidence that dementia is substantially under-ascertained in administrative data and highlight the need for more complete surveillance approaches to support dementia policy, planning, and care delivery.

Acknowledgements

This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health (MOH) and the Ministry of Long-Term Care (MLTC). This document used data adapted from the Statistics Canada Postal CodeOM Conversion File, which is based on data licensed from Canada Post Corporation, and/or data adapted from the Ontario Ministry of Health Postal Code Conversion File, which contains data copied under licence from ©Canada Post Corporation and Statistics Canada. Parts of this material are based on data and/or information compiled and provided by MOH, CIHI, and IQVIA Solutions Canada Inc. 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 is intended or should be inferred. We thank IQVIA Solutions Canada Inc. for use of their Drug Information File.

Funding Statement

This study received funding from the Alzheimer Society of Canada. The funder had no role in study design, data analysis, interpretation, manuscript preparation, or the decision to submit the manuscript for publication.

Conflict of Interests Statement

None declared

Ethics Statement

ICES is a prescribed entity under Ontario's Personal Health Information Protection Act (PHIPA). Section 45 of PHIPA authorises ICES to collect personal health information, without consent, for the purpose of analysis or compiling statistical information with respect to the management, evaluation, monitoring, allocation of resources, or planning of all or part of the health system. This study used only data collected and held by ICES under section 45 of PHIPA. Under this legal framework, projects that use only ICES data collected under section 45 are exempt from research ethics board review. The use of data in this project was authorised under section 45 of PHIPA and approved by the ICES Privacy and Legal Office.

AI Disclosure Statement

The authors used GPT-5.6 SOL, OpenAI, for copyediting and consistency checks. The output was reviewed and verified by the authors.

Data Access Statement

The dataset from this study is held securely in coded form at ICES. While legal data sharing agreements between ICES and data providers (e.g., healthcare organizations and government) prohibit ICES from making the dataset publicly available, access may be granted to those who meet pre-specified criteria for confidential access, available at www.ices.on.ca/DAS (email: das@ices.on.ca). The full dataset creation plan and underlying analytic code are available from the authors upon request.

References

1. Emmady PD, Schoo C, Tadi P. Major Neurocognitive Disorder (Dementia). In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2026.

2. Public Health Agency of Canada. Dementia: Overview. 2014. https://www.canada.ca/en/public-health/services/diseases/dementia.html. Accessed 28 Oct 2025.

3. Dementia: Health at a Glance 2025. OECD. 2025. https://www.oecd.org/en/publications/health-at-a-glance-2025_8f9e3f98-en/full-report/dementia_8b14a2ea.html. Accessed 22 Apr 2026.

4. Alzheimer Society of Canada. Navigating the Path Forward for Dementia in Canada: The Landmark Study Report #1. Toronto, Canada; 2022.

5. Canadian Centre for Economic Analysis. Dementia in Canada: Economic Burden 2020 to 2050. 2023.

6. Bradford A, Kunik ME, Schulz P, Williams SP, Singh H. Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. Alzheimer Dis Assoc Disord. 2009;23:306–14. https://doi.org/10.1097/WAD.0b013e3181a6bebc

7. Gauthier S, Webster C, Servaes S, Morais JA, Rosa-Neto P. World Alzheimer Report 2022: Life after diagnosis: Navigating treatment, care and support. London, England; 2022.

8. Amjad H, Roth DL, Samus QM, Yasar S, Wolff JL. Potentially Unsafe Activities and Living Conditions of Older Adults with Dementia. J Am Geriatr Soc. 2016;64:1223–32. https://doi.org/10.1111/jgs.14164

9. Lang L, Clifford A, Wei L, Zhang D, Leung D, Augustine G, et al. Prevalence and determinants of unrecorded dementia in the community: a systematic literature review and a meta-analysis. BMJ Open. 2017;7:e011146. https://doi.org/10.1136/bmjopen-2016-011146

10. Klee M, Leist AK, Veldsman M, Ranson JM, Llewellyn DJ. Socioeconomic Deprivation, Genetic Risk, and Incident Dementia. Am J Prev Med. 2023;64:621–30. https://doi.org/10.1016/j.amepre.2023.01.012

11. Public Health Agency of Canada. Canadian Chronic Disease Surveillance System (CCDSS) — Canada.ca. 2024. https://health-infobase.canada.ca/ccdss/data-tool/Index. Accessed 16 Oct 2024.

12. Ma'u E, Cullum S, Mukadam N, Davis D, Rivera-Rodriguez C, Cheung G. Estimating the Prevalence of Dementia in New Zealand Using Capture-recapture Analysis on Routinely Collected Health Data. Int J Geriatr Psychiatry. 2024;39:e6131. https://doi.org/10.1002/gps.6131

13. Sanderson M, Benjamin JT, Lane MJ, Cornman CB, Davis DR. Application of capture-recapture methodology to estimate the prevalence of dementia in South Carolina. Ann Epidemiol. 2003;13:518–24. https://doi.org/10.1016/S1047-2797(03)00036-X

14. Welberry HJ, Brodaty H, Hsu B, Barbieri S, Jorm LR. Measuring dementia incidence within a cohort of 267,153 older Australians using routinely collected linked administrative data. Sci Rep. 2020;10:8781. https://doi.org/10.1038/s41598-020-65273-w

15. Stephen C. Capture-recapture methods in epidemiological studies. Infect Control Hosp Epidemiol. 1996;17:262–6. https://doi.org/10.1086/647290

16. Ma'u E, Cullum S, Mukadam N, Davis D, Rivera-Rodriguez C, Cheung G. Estimating the incidence of dementia in New Zealand: a cohort study applying capture-recapture modelling to routinely collected linked health datasets. Lancet Reg Health - West Pac. 2024;52:101216. https://doi.org/10.1016/j.lanwpc.2024.101216

17. Mukadam N, Marston L, Flanagan K, Ma'u E, Cheung G, Böhning D. Estimating undiagnosed dementia in England using capture recapture techniques. BMC Geriatr. 2025;25:1. https://doi.org/10.1186/s12877-024-05591-0

18. Waller M, Mishra GD, Dobson AJ. Estimating the prevalence of dementia using multiple linked administrative health records and capture–recapture methodology. Emerg Themes Epidemiol. 2017;14:3. https://doi.org/10.1186/s12982-017-0057-3

19. Jaakkimainen RL, Bronskill SE, Tierney MC, Herrmann N, Green D, Young J, et al. Identification of Physician-Diagnosed Alzheimer's Disease and Related Dementias in Population-Based Administrative Data: A Validation Study Using Family Physicians' Electronic Medical Records. J Alzheimers Dis JAD. 2016;54:337–49. https://doi.org/10.3233/JAD-160105

20. Public Health Agency of Canada. Public Health Infobase, Canadian Chronic Disease Surveillance System (CCDSS) Disease-Specific Case Definitions. 2019.

21. Hirdes JP, Ljunggren G, Morris JN, Frijters DH, Finne Soveri H, Gray L, et al. Reliability of the interRAI suite of assessment instruments: a 12-country study of an integrated health information system. BMC Health Serv Res. 2008;8:277. https://doi.org/10.1186/1472-6963-8-277

22. Statistics Canada. Postal Code Conversion File Plus (PCCF+) Version 8A, Reference Guide. 2022.

23. Canadian Institute for Health Information. *Measuring Health Inequalities: A Toolkit*. 2018. Accessed August 30, 2026. https://www.cihi.ca/en/measuring-health-inequalities-a-toolkit.

24. Chao A, Tsay PK, Lin S, Shau W, Chao D. The applications of capture-recapture models to epidemiological data. Stat Med. 2001;20:3123–57. https://doi.org/10.1002/sim.996

25. Rivest L-P, Baillargeon S. Rcapture: Loglinear Models for Capture-recapture Experiments. 2022.

26. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian Model Averaging: A Tutorial. Stat Sci. 1999;14:382–401.

27. Sternberg SA, Wolfson C, Baumgarten M. Unrecorded Dementia in Community-Dwelling Older People: The Canadian Study of Health and Aging. J Am Geriatr Soc. 2000;48:1430–4. https://doi.org/10.1111/j.1532-5415.2000.tb02633.x

28. Jacobsohn GC, Hollander M, Beck AP, Gilmore-Bykovskyi A, Werner N, Shah MN. Factors Influencing Emergency Care by Persons With Dementia: Stakeholder Perceptions and Unmet Needs. J Am Geriatr Soc. 2019;67:711–8. https://doi.org/10.1111/jgs.15737

29. Herron RV, Rosenberg MW. "Not there yet": Examining community support from the perspective of people with dementia and their partners in care. Soc Sci Med 1982. 2017;173:81–7. https://doi.org/10.1016/j.socscimed.2016.11.041

30. Preparing for the future of dementia care in Ontario. Alzheimer Society of Ontario. 2025. https://admin.alzheimer.ca/on/en/preparing_for_the_future. Accessed 28 Oct 2025.

31. Public Health Agency of Canada. A Dementia Strategy for Canada: Together We Achieve - 2024 Annual Report. 2024. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/dementia-strategy-annual-report-parliament-2024.html. Accessed 21 July 2025.

Abbreviations

BIC:	Bayesian Information Criterion
CI:	Confidence Interval
DAD:	Discharge Abstract Database
ODB:	Ontario Drug Benefit Database
OHIP:	Ontario Health Insurance Plan Database
PY:	person-year
RAI:	Resident Assessment Instrument

Supplementary Appendix 1: Summary of data sources used in study

Database	Description and population coverage
Discharge Abstract Database (DAD)	National hospital administrative database maintained by the Canadian Institute for Health Information (CIHI). The DAD captures demographic, administrative and clinical information for discharges from acute inpatient hospitals in Ontario, including discharges following death or transfer. Reporting of acute inpatient care is mandatory in Ontario. During the study period, diagnoses were coded using **ICD-10-CA** and interventions using the Canadian Classification of Health Interventions. **Metadata and data quality:** https://www.cihi.ca/en/discharge-abstract-database-dad-metadata
Ontario Drug Benefit (ODB) Claims Database	Contains records of prescription medications dispensed and reimbursed through Ontario's publicly funded drug programs. Coverage includes all Ontario residents aged $\geq$65 years as well as eligible younger residents, including recipients of Ontario Works or the Ontario Disability Support Program, residents of long-term care homes, recipients of publicly funded home and community care, and people enrolled in the Trillium Drug Program. **Metadata and data quality:** https://data.ontario.ca/dataset/ontario-drug-benefit-odb-monthly-data https://pubmed.ncbi.nlm.nih.gov/12879144/
Ontario Health Insurance Plan (OHIP) Claims Database	Contains claims submitted for physician services covered by Ontario's universal health insurance program, including inpatient and outpatient physician services. It contains traditional fee-for-service claims and "shadow billings" submitted by physicians remunerated through many non-fee-for-service models. **Metadata and data quality** https://datadictionary.ices.on.ca/Applications/DataDictionary/Library.aspx?Library=OHIP https://pubmed.ncbi.nlm.nih.gov/27567819/
interRAI assessment data — Home Care and Long-Term Care *(RAI-HC and RAI-MDS 2.0/CCRS during the study period)*	Standardised comprehensive clinical assessments used for people receiving publicly funded home care or living in long-term care. RAI-HC contains multidimensional assessments of home care clients, while the legacy RAI-MDS 2.0 data submitted to CIHI's Continuing Care Reporting System (CCRS) contain corresponding information for long-term care and continuing care residents. Assessments include clinical, cognitive, functional, behavioural, social and service-use information. **Metadata and data quality:** https://www.cihi.ca/en/continuing-care-reporting-system-ccrs-metadata https://www.cihi.ca/en/home-care-reporting-system-hcrs-metadata https://www.cihi.ca/sites/default/files/document/hcrs-data-users-guide-2022-2023-en.pdf https://www.cihi.ca/sites/default/files/document/long-term-care-data-2022-2023-guide-en.pdf
Registered Persons Database (RPDB)	Population registry containing information on people who have been issued an Ontario health card. It contains demographic information and periods of eligibility for publicly funded provincial health insurance. Current descriptions identify date of birth, sex, residential information, dates of health insurance eligibility and, for deceased individuals, date of death. **Metadata and data quality:** https://data.ontario.ca/dataset/registered-persons-database-rpdb

Supplementary Appendix 2: Diagnostic codes and medications used in case ascertainment of dementia

Codes used in case ascertainment of dementia

ICD-9: Description
046.1x Creutzfeldt-Jakob disease
290.0x Senile dementia, uncomplicated
290.1x Presenile dementia
290.2x Senile dementia with delusional or depressive features
290.3x Senile dementia with delirium
290.4x Vascular dementia
294.1x Dementia in other diseases
294.2x Dementia, unspecified
331.0x Alzheimer's disease
331.1x Frontotemporal dementia
331.5x Idiopathic normal pressure hydrocephalus

ICD-10: Description
F00.x Dementia in Alzheimer's disease
F01.x Vascular dementia
F02.x Dementia in other diseases classified elsewhere
F03.x Unspecified dementia
G30.x Alzheimer's disease

OHIP Diagnosis Codes
290 Dementias
331 Other cerebral degenerations

ATC code: Drug name
N06DX01 – Memantine
N06DA03 – Rivastigmine
N06DA04 – Galantamine
N06DA02 – Donepezil
N06DA01 – Tacrine

interRAI Assessment Items: Description
J1G (RAI-HC): Alzheimer's
J1H (RAI-HC): Dementia other than Alzheimer's
I1C (interRAI HC): Alzheimer's disease
I1D (interRAI HC): Dementia other than Alzheimer's disease
I1R (CCRS-LTC): Alzheimer's disease
I1V (CCRS-LTC): Dementia other than Alzheimer's disease

Supplementary Appendix 3: List of all log-linear models fit for 10-year, three ascertainment source analysis

All Ontario Without RAI (8 Models)	abundance	stderr	df	AIC	BIC
[DADODB,DADOHIP,ODBOHIP]	472271	5104.18	0	97.1567	170.781
[DADOHIP,ODBOHIP]	422737	3500.17	1	821.767	884.873
[DADODB,ODBOHIP]	344354	769.789	1	2426.71	2489.81
[ODBOHIP,DAD]	341229	679.361	2	2541.03	2593.62
[DADODB,DADOHIP]	315042	411.135	1	5424.6	5487.71
[DADOHIP,ODB]	315642	402.234	2	5454.04	5506.63
[DADODB,OHIP]	318085	361.288	2	5592.6	5645.19
[DAD,ODB,OHIP]	318390	354.017	3	5604.66	5646.74

Supplementary Appendix 4: List of all log-linear models fit for 10-year, four ascertainment source analysis

All Ontario With RAI (113 Models)	abundance	stderr	df	AIC	BIC
[DADODBOHIP,DADODBRAI,DADOHIPRAI]	527716	4231.75	1	199.776	348.725
[DADODBOHIP,DADODBRAI,DADOHIPRAI,ODBOHIPRAI]	550754	13044.4	0	197.821	357.409
[DADODBOHIP,DADODBRAI,ODBOHIPRAI]	499582	4247.67	1	219.269	368.217
[DADODBRAI,ODBOHIPRAI,DADOHIP]	492675	3718.73	2	233.255	371.564
[DADODBOHIP,DADODBRAI,OHIPRAI]	512689	3096.06	2	233.804	372.113
[DADODBRAI,DADOHIPRAI,ODBOHIP]	517246	3655.34	2	235.702	374.012
[DADODBOHIP,DADOHIPRAI,ODBRAI]	521486	3992.86	2	237.18	375.49
[DADODBRAI,DADOHIPRAI,ODBOHIPRAI]	501920	7683.49	1	233.186	382.135
[DADODBOHIP,ODBOHIPRAI,DADRAI]	491757	3789.6	2	243.826	382.135
[DADODBRAI,DADOHIP,ODBOHIP,OHIPRAI]	506919	2794.87	3	257.626	385.296
[DADODBOHIP,DADOHIPRAI,ODBOHIPRAI]	514280	9834.03	1	238.568	387.516
[ODBOHIPRAI,DADODB,DADOHIP,DADRAI]	484237	3200.62	3	260.459	388.129
[DADODBOHIP,DADRAI,ODBRAI,OHIPRAI]	507004	2877.48	3	270.892	398.563
[DADOHIPRAI,ODBOHIPRAI,DADODB]	482555	6101.34	2	262.356	400.665
[DADOHIPRAI,DADODB,ODBOHIP,ODBRAI]	509673	3356.29	3	281.419	409.089
[DADODB,DADOHIP,DADRAI,ODBOHIP,ODBRAI,OHIPRAI]	499845	2490.59	4	302.682	419.713
[ODBOHIPRAI,DADOHIP,DADRAI]	480587	3128.92	4	387.362	504.393
[DADOHIPRAI,ODBOHIPRAI]	463998	5206.85	3	376.954	504.624
[DADOHIP,DADRAI,ODBOHIP,ODBRAI,OHIPRAI]	498475	2483.97	5	445.913	552.305
[DADOHIPRAI,ODBOHIP,ODBRAI]	505510	3279.22	4	435.351	552.383
[DADODBRAI,ODBOHIPRAI]	391635	1247.71	3	3911.67	4039.34
[ODBOHIPRAI,DADODB,DADRAI]	398190	1259.7	4	4088.91	4205.94
[ODBOHIPRAI,DADRAI]	393030	1149.64	5	4379.09	4485.49
[DADODBRAI,ODBOHIP,OHIPRAI]	427229	1418.13	4	5403.74	5520.77
[DADODB,DADRAI,ODBOHIP,ODBRAI,OHIPRAI]	429211	1357.85	5	5418.22	5524.61
[DADRAI,ODBOHIP,ODBRAI,OHIPRAI]	422811	1256.7	6	5932.94	6028.69
[DADODBOHIP,ODBOHIPRAI]	377378	1217.22	3	8493.71	8621.38
[ODBOHIPRAI,DADODB,DADOHIP]	391880	1349.97	4	9040.61	9157.64
[ODBOHIPRAI,DADOHIP]	386119	1222.71	5	9413.37	9519.76
[DADODBOHIP,DADOHIPRAI]	395559	1443.02	3	10172.8	10300.4
[DADODBOHIP,ODBRAI,OHIPRAI]	425820	1598.11	4	10338.8	10455.8
[DADODB,DADOHIP,ODBOHIP,ODBRAI,OHIPRAI]	432049	1554.97	5	10430.9	10537.3
[DADOHIPRAI,DADODB,ODBOHIP]	407761	1503.28	4	10477.7	10594.7
[DADODBOHIP,DADRAI,OHIPRAI]	422401	1556.41	4	10866.8	10983.9
[DADOHIPRAI,ODBOHIP]	400944	1366.21	5	10955.1	11061.5
[DADODB,DADOHIP,DADRAI,ODBOHIP,OHIPRAI]	429013	1523.18	5	10976	11082.4
[DADOHIP,ODBOHIP,ODBRAI,OHIPRAI]	425304	1453.09	6	11062	11157.8
[ODBOHIPRAI,DADODB]	359087	612.049	5	11439.8	11546.1
[DADOHIP,DADRAI,ODBOHIP,OHIPRAI]	422457	1422.25	6	11589.4	11685.2
[ODBOHIPRAI,DAD]	355866	549.297	6	11785.3	11881.1
[DADODB,DADRAI,ODBOHIP,OHIPRAI]	399336	1028.66	6	13683.1	13778.8
[DADRAI,ODBOHIP,OHIPRAI]	390602	883.1	7	14606.8	14691.9
[DADODBOHIP,DADODBRAI]	354701	421.324	3	14666.7	14794.4
[DADODBRAI,DADOHIPRAI]	357804	603.797	3	14769	14896.7
[DADODBOHIP,DADRAI,ODBRAI]	356795	431.231	4	15420.5	15537.5
[DADOHIPRAI,DADODB,ODBRAI]	361807	633.753	4	15434	15551.1
[DADODB,ODBOHIP,ODBRAI,OHIPRAI]	380648	703.362	6	15481.8	15577.5
[DADOHIPRAI,ODBRAI]	360838	613.001	5	15508.9	15615.3
[DADODBRAI,DADOHIP,ODBOHIP]	358046	434.772	4	15940.7	16057.8
[ODBOHIP,ODBRAI,OHIPRAI,DAD]	374868	616.907	7	16295.2	16380.3
[DADOHIP,DADRAI,ODBOHIP,ODBRAI]	360078	443.494	6	16621.9	16717.7
[DADODB,DADOHIP,DADRAI,ODBOHIP,ODBRAI]	360097	443.49	5	16620.9	16727.3
[DADODBRAI,ODBOHIP]	352643	365.337	5	17124.6	17231
[DADODBRAI,DADOHIP,OHIPRAI]	372953	689.153	4	17358	17475
[DADODB,DADOHIP,DADRAI,ODBRAI,OHIPRAI]	376470	705.376	5	17761.7	17868.1
[DADOHIP,DADRAI,ODBRAI,OHIPRAI]	374973	680.429	6	17918.2	18014
[DADODB,DADRAI,ODBOHIP,ODBRAI]	354295	371.719	6	17932.4	18028.1

Continued

Supplementary Appendix 4: Continued

All Ontario With RAI (113 Models)	abundance	stderr	df	AIC	BIC
[DADRAI,ODBOHIP,ODBRAI]	354017	367.625	7	17974.7	18059.8
[DADODBRAI,OHIPRAI]	364212	555.028	5	18598.1	18704.4
[DADODBOHIP,OHIPRAI]	381931	920.595	5	18735.4	18841.8
[DADODB,DADRAI,ODBRAI,OHIPRAI]	367077	569.152	6	19149.1	19244.8
[DADRAI,ODBRAI,OHIPRAI]	366204	555.004	7	19252	19337.1
[DADODB,DADOHIP,ODBOHIP,OHIPRAI]	389475	962.065	6	19263.3	19359
[DADOHIP,ODBOHIP,OHIPRAI]	386399	913.574	7	19629.1	19714.2
[DADODBOHIP,ODBRAI]	348952	355.142	5	20536.2	20642.6
[DADODBOHIP,DADRAI]	345874	323.267	5	21724.8	21831.2
[DADOHIPRAI,DADODB]	343332	378.14	5	21761.2	21867.6
[DADOHIPRAI,ODB]	342504	362.387	6	21878	21973.7
[DADODB,DADOHIP,ODBOHIP,ODBRAI]	351746	365.331	6	22049.2	22145
[DADODB,ODBOHIP,OHIPRAI]	366157	567.118	7	22068.2	22153.3
[DADOHIP,ODBOHIP,ODBRAI]	351323	360.786	7	22161	22246.1
[ODBOHIP,OHIPRAI,DAD]	362926	516.125	8	22560.1	22634.5
[DADODB,ODBOHIP,ODBRAI]	348200	320.903	7	22892.2	22977.3
[ODBOHIP,ODBRAI,DAD]	347466	311.388	8	23074.9	23149.4
[DADODB,DADOHIP,DADRAI,ODBOHIP]	348404	332.026	6	23382.4	23478.1
[DADODB,DADOHIP,ODBRAI,OHIPRAI]	364891	592.784	6	23409.7	23505.4
[DADOHIP,DADRAI,ODBOHIP]	348070	328.456	7	23473.5	23558.6
[DADOHIP,ODBRAI,OHIPRAI]	361977	544.524	7	23826.3	23911.4
[DADODB,DADRAI,ODBOHIP]	345652	297.585	7	24055.6	24140.8
[DADODBRAI,DADOHIP]	342672	286.51	5	24064.9	24171.3
[DADRAI,ODBOHIP]	345073	290.096	8	24203.6	24278.1
[DADODBRAI,OHIP]	340746	259.909	6	24517.4	24613.2
[DADODB,ODBRAI,OHIPRAI]	354457	428.534	7	25046.1	25131.2
[DADOHIP,DADRAI,ODBRAI]	343872	289.855	7	25161.3	25246.4
[DADODB,DADOHIP,DADRAI,ODBRAI]	343968	291.859	6	25152	25247.8
[ODBRAI,OHIPRAI,DAD]	352935	405.434	8	25323.2	25397.6
[DADRAI,ODBRAI,OHIP]	341799	262.561	8	25673.4	25747.9
[DADODB,DADRAI,ODBRAI,OHIP]	341862	264.01	7	25668.5	25753.7
[DADODB,DADOHIP,DADRAI,OHIPRAI]	352780	416.255	6	25962.5	26058.2
[DADODBOHIP,RAI]	340917	279.209	6	26005.9	26101.7
[DADOHIP,DADRAI,OHIPRAI,ODB]	351401	395.27	7	26212	26297.1
[DADODB,DADRAI,OHIPRAI]	349922	376.88	7	26494.5	26579.6
[DADRAI,OHIPRAI,ODB]	348869	361.26	8	26700.7	26775.1
[DADODB,DADOHIP,ODBOHIP,RAI]	343103	286.508	7	27923.8	28008.9
[DADOHIP,ODBOHIP,RAI]	342880	284.023	8	27984.4	28058.9
[DADODB,ODBOHIP,RAI]	341340	263.416	8	28351.6	28426.1
[ODBOHIP,DAD,RAI]	340962	258.356	9	28448	28511.8
[DADODB,DADOHIP,ODBRAI]	340172	261.893	7	29193.3	29278.4
[DADOHIP,ODBRAI]	339820	256.883	8	29277.2	29351.7
[DADODB,ODBRAI,OHIP]	338718	241.406	8	29535.7	29610.2
[ODBRAI,DAD,OHIP]	338443	237.59	9	29605.4	29669.2
[DADODB,DADOHIP,DADRAI]	338599	247.822	7	29973	30058.2
[DADOHIP,DADRAI,ODB]	338311	243.653	8	30041.6	30116
[DADODB,DADRAI,OHIP]	337398	230.434	8	30252	30326.4
[DADRAI,ODB,OHIP]	337167	227.168	9	30309.7	30373.5
[DADODB,DADOHIP,OHIPRAI]	346717	365.175	7	30491.4	30576.5
[DADOHIP,OHIPRAI,ODB]	345828	350.777	8	30651.7	30726.2
[DADODB,OHIPRAI]	343327	312.929	8	31124.7	31199.2
[OHIPRAI,DAD,ODB]	342748	304.188	9	31243	31306.8
[DADODB,DADOHIP,RAI]	335663	225.236	8	33587.6	33662.1
[DADOHIP,ODB,RAI]	335468	222.218	9	33631.1	33694.9
[DADODB,OHIP,RAI]	334845	212.21	9	33762	33825.8
[DAD,ODB,OHIP,RAI]	334683	209.744	10	33799.4	33852.6

Supplementary Appendix 5: Capture-recapture results for three-year, three ascertainment source analysis

Strata	Observed			
	Cases	Cumulative incidence (%)	Person-years of follow-up	Incidence per 10,000 person-years
All	96,290	3.7%	7,419,648	130
Female	55,778	4.0%	3,983,180	140
Male	40,512	3.4%	3,436,469	118
Rural	10,921	3.3%	923,066	118
Urban	85,068	3.7%	6,464,559	132
Income Q1 (lowest)	24,453	4.6%	1,471,738	166
Income Q2	21,215	3.8%	1,541,020	138
Income Q3	17,872	3.4%	1,451,445	123
Income Q4	16,199	3.4%	1,413,460	115
Income Q5 (highest)	16,222	3.1%	1,507,930	108

Strata	Estimated		
	Cases (95% CI)	Cumulative incidence% (95% CI)	Proportion unrecorded (95% CI)
All	213,253 (204,711, 222,459)	8.2 (7.8, 8.5)	54.8 (53.0, 56.7)
Female	124,034 (117,266, 131,546)	8.8 (8.3, 9.3)	55.0 (52.4, 57.6)
Male	89,752 (84,586, 95,507)	7.5 (7.0, 8.0)	54.9 (52.1, 57.6)
Rural	23,301 (21,035, 26,043)	7.0 (6.3, 7.8)	53.1 (48.1, 58.1)
Urban	189,982 (181,701, 198,969)	8.4 (8.0, 8.8)	55.2 (53.2, 57.2)
Income Q1 (lowest)	54,793 (50,526, 59,740)	10.3 (9.5, 11.2)	55.4 (51.6, 59.1)
Income Q2	43,402 (40,184, 47,142)	7.8 (7.2, 8.5)	51.1 (47.2, 55.0)
Income Q3	44,977 (40,557, 50,274)	8.7 (7.8, 9.7)	60.3 (55.9, 64.5)
Income Q4	38,555 (34,730, 43,176)	8.0 (7.2, 9.0)	58.0 (53.4, 62.5)
Income Q5 (highest)	32,847 (29,974, 36,307)	6.4 (5.8, 7.0)	50.6 (45.9, 55.3)