# Evaluating the use of health administrative data for population surveillance of homelessness: a validation study

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2019-030221 |
| Article Type: | Research |
| Date Submitted by the Author: | 05-Mar-2019 |
| Complete List of Authors: | Richard, Lucie; ICES Western<br>Hwang, Stephen W. ; St Michaels Hospital, Toronto<br>Forchuk, Cheryl; Western University<br>Nisenbaum, Rosane; St. Michael's Hospital, Centre for Research on Inner City Health; University of Toronto, Dalla Lana School of Public Health<br>Clemens, Kristin; Western University<br>Wiens, Kathryn; University of Toronto<br>Booth, Richard; Western University<br>Azimaee, Mahmoud; Institute for Clinical Evaluative Sciences<br>Shariff, Salimah; ICES Western |
| Keywords: | EPIDEMIOLOGY, HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PUBLIC HEALTH |
| | |

# Evaluating the use of health administrative data for population surveillance of homelessness: a validation study

**Richard, Lucie**, MA, ICES Western
**Hwang, Stephen W.**, MD, MPH, Centre for Urban Health Solutions, Li Ka Shing Knowledge Institute, St Michael's Hospital; Dalla Lana School of Public Health, University of Toronto
**Forchuk, Cheryl**, PhD, Western University
**Nisenbaum, Rosane**, PhD, Centre for Urban Health Solutions, Applied Health Research Centre, Li Ka Shing Knowledge Institute, St Michael's Hospital; Dalla Lana School of Public Health, University of Toronto
**Clemens, Kristin**, MD, MSc, Western University
**Wiens, Kathryn**, MSc, University of Toronto
**Booth, Richard**, PhD, Western University
**Azimaee, Mahmoud**, ICES
**Shariff, Salimah Z.**, PhD, ICES Western

Corresponding author: Lucie Richard lucie.richard@ices.on.ca

Funding Statement: This study was supported by an Ontario Trillium Foundation – Local Poverty Reduction Fund grant (held by Drs. Forchuk, Booth and Shariff), an Academic Medical Organization of Southwestern Ontario Academic Health Science Centre Alternate Funding Innovation Fund grant (held by Drs. Clemens and Shariff) and two Canadian Institutes of Health Research grants, CIHR MOP-86765 and HOA-80066 (held by Drs. Hwang and Nisenbaum).

Declaration of author(s) competing interests: Dr. Clemens received a Diabetes Canada Junior Investigator Award in 2017 funded by Astra Zeneca. Dr. Clemens has attended conferences sponsored by Merck Inc.

Author contributions: LR conceived of the study, participated in the study coordination, study design, acquisition of data and interpretation of results, performed the analysis and drafted the manuscript. SZS conceived of the study, participated in the study design, interpretation of study results and provided feedback on the manuscript. HW, RN and RB participated in the study design, acquisition of data, and interpretation of study results and provided feedback on the manuscript. CF, KC and KW contributed to the study design, interpretation of study results and provided feedback on the manuscript. All authors read and approved the final manuscript.

Word count: 2,681

Keywords: Homelessness; validation; health services research

## Abstract

**Objectives:** To validate case ascertainment algorithms for identifying individuals experiencing homelessness in health administrative databases; and to estimate homelessness prevalence trends in Ontario, Canada between 2007 and 2016.

**Design:** A population-based retrospective validation study

**Setting:** Ontario, Canada, from 2007 to 2014.

**Participants**: Our reference standard was the known housing status of a longitudinal cohort of housed (n=137,200) and homeless or vulnerably housed (n=686) individuals. Two reference standard definitions of homelessness were adopted: the housing episode and the annual housing experience (any homelessness within a calendar year).

**Main outcome measures**: Sensitivity, specificity, positive and negative predictive value and positive likelihood ratios of 30 case ascertainment algorithms for detecting homelessness using up to eight health services databases.

**Results**: Sensitivity estimates ranged from 10.8% to 28.9% (housing episode definition) and 18.5 to 35.6% (annual housing experience definition). Specificities exceeded 99% and positive likelihood ratios were high using both definitions. The most optimal algorithm estimates that 59,974 (95% CI: 55,231 to 65,208) Ontarians (0.53% of the adult population) experienced homelessness in 2016, a 67.3% increase from 2007.

**Conclusions**: In Ontario, case ascertainment algorithms for identifying homelessness had low sensitivity but very high specificity and positive likelihood ratio. The use of health administrative databases may offer opportunities to track individuals experiencing homelessness over time and inform efforts to improve housing and health status in this vulnerable population.

## Article Summary

Strengths and limitations

- This study validated health administrative codes used in Canadian health databases against a longitudinally collected, representative sample of individuals with known housing status;
- Health administrative data for certain subgroups without Ontario health coverage (e.g. First Nations on reserves, individuals newly arrived to Ontario) was unavailable;
- Our general population sample was assumed housed for the entirety of their observation period. It is possible despite our screening efforts that certain individuals experienced homelessness episodes during their participation in this study.

# Introduction

Individuals experiencing homelessness commonly face physical and mental health challenges, increased morbidity, mortality and health care usage (1, 2). However, surveillance of this population has proven challenging (3-7), with most efforts to date primarily focused on enumerating the homeless at a given point in time (8). While such ecological measures are valuable for service planning, they have been criticized as inaccurate and unrepresentative. Further, these measures do not permit follow up over time or the evaluation of targeted strategies (9, 10). In nations like Canada where standardized health administrative databases are used, such as for hospital services (11), and where financial barriers to healthcare are minimized through provision of universal healthcare, it is possible to measure and track individuals experiencing homelessness at the population level. However, such data are prone to errors in misclassification (12); validation studies are thus necessary to evaluate the accuracy of case ascertainment algorithms (13-15).

The aims of this study were to (a) develop and validate case ascertainment algorithms to identify individuals experiencing homelessness in health administrative databases in Ontario, Canada; and (b) estimate annual population-prevalence of homelessness in Ontario over a 10-year period using the best performing algorithm.

# Methods

<u>Study design and participants</u>

We validated 30 case ascertainment algorithms to detect homelessness using up to eight health administrative databases in Ontario, Canada's most populous province. All databases were linked using unique encoded identifiers and analyzed at ICES (16). This study was approved by the St Michael's Hospital Research Ethics Board, and follows STARD guidelines for reporting diagnostic accuracy studies.

Patient and public involvement

Due to the coded nature of ICES data, this research was done without patient involvement. Patients were not involved in the development of the research question, invited to comment on the study design, consulted to interpret the results, and were not invited to contribute to the writing or editing of this document for readability or accuracy.

Data availability

While data sharing agreements prohibit ICES from making the dataset publicly available, access to the data may be granted to those who meet pre-specified criteria for confidential access, available at www.ices.on.ca/DAS. The full dataset creation plan and underlying analytic code are available from the authors upon request, understanding that the computer programs rely upon coding templates or macros that are unique to ICES and are therefore either inaccessible or may require modification.

Participants

Our validation cohort included adults (18 years or older) eligible for Ontario health coverage who participated in the HHiT study (the "HHiT sample"), which prospectively followed a representative sample of homeless or vulnerably housed adults in Toronto and Ottawa, Ontario (17). Participant data were organized into consecutive self-reported housing episodes, ranging from an earliest date of January 31, 2007 to a latest date of March 14, 2014. Due to the low prevalence (<5%) of exclusively housed individuals in this cohort, an additional group of adults presumed housed (the "general population sample") was randomly selected from the ICES Registered Persons Database (RPDB), which includes all individuals eligible for Ontario health coverage. A similar approach was used in previous validation studies (18, 19). To ensure our general population sample had a high likelihood of being housed, we deemed individuals eligible if they were not part of the HHiT study, resided in Toronto or Ottawa throughout the study period and did not reside in a postal code associated with shelter services. We randomly selected 200 individuals for each HHiT participant to approximate the nearest available Canadian homelessness prevalence estimate (20).

Reference standard

The period over which housing status is assessed substantially impacts any analysis of agreement between the reference standard and case ascertainment algorithms. Thus, we *a priori* selected two reference standard definitions (units of analysis) based on their expected utility: a) the housing episode and b) the annual housing experience. Within the HHiT cohort, housing episodes were categorized as *housed* or *homeless* based on pre-established criteria (21). The general population sample was assumed housed for their observation period. For the annual housing experience definition, individuals were categorized as homeless if a homeless episode occurred during the calendar year.

Case Ascertainment Algorithms and Data Sources

Homeless indicators were identified by searching the ICES data dictionary (22) for data elements indicative of housing status (search terms included: 'homeless', 'shelter', 'housing', 'residence', 'transient')(Supplement Table 2). We assessed housing status indicators (Supplement Table 1) present in: the Discharge Abstract Database (DAD), the National Ambulatory Care Reporting System emergency (NACRS), the Ontario Mental Health Reporting System (OMHRS), the Home Care Database (HCD), the Resident Assessment Instrument Contact Assessment Database (RAICA), the National Rehabilitation Reporting System (NRS) and the Canadian Organ Replacement Registry (CORR). The first three sources report hospital encounters and are tracked by the

Canadian Institute for Health Information (CIHI)(11); for brevity these are hereafter referred to as "CIHI databases".

Postal codes are also often recorded in the above records; therefore, we additionally assessed postal codes where present and in the ICES PSTLYEAR database (which provides a yearly postal code for individuals with Ontario health coverage) against Toronto and Ottawa-based postal codes identifying shelter services or hospitals (which are sometimes erroneously coded instead of shelters)(23). Postal codes which included residential addresses, as determined through a Geographic Information System, were not used to avoid misclassifying housed individuals as homeless.

We tested 30 case ascertainment algorithms which varied by: 1) databases included (all vs. CIHI only); 2) inclusion or exclusion of postal code indicators (none, in health service databases or in PSTLYEAR) and 3) extension of time intervals (ranging 0 days to ±180 days) before and after the reference period. The practice of extending time intervals is known to enhance the sensitivity of case ascertainment algorithms (24, 25). Reference periods without overlapping healthcare encounters were coded as test negative by default.

Other data sources used to describe the cohort (described in Supplement Table 2) included the ICES RPDB, Ontario Health Insurance Physicians (OHIP) claims database, the Immigration, Refugee and Citizenship Canada (IRCC) Permanent Residents database, and several ICES-derived population-surveillance datasets including: the Chronic Obstructive Pulmonary Disease (COPD)(26), Ontario Diabetes Dataset (ODD)(27), Congestive Heart Failure (CHF)(28) and Ontario HIV (29) derived cohorts.

## Statistical analysis

We provided cohort demographics, comorbidities and recent health services usage (variables defined in Supplement Table 3). Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and positive likelihood ratio (LR+) were calculated for all algorithms (formulae in Supplement Table 4). Confidence intervals (95% CIs) were calculated using the Wilson score method (30). For each reference standard, we selected an optimal algorithm that maximized validation statistics while considering scalability (i.e. applicability outside Ontario).

We then applied the optimal annual housing experience algorithm to identify Ontarians experiencing homelessness in each of the 2007 to 2016 calendar years, further describing those identified during 2016. Finally, we estimated population-prevalence of homelessness between 2007 and 2016, correcting for sensitivity. Prevalence rates were calculated by dividing estimated population prevalence by the total adult Ontario population for each year. A Poisson regression model was used to estimate the annual change in prevalence over time.

All analyses were conducted using SAS, version 9.4 (31).

# Results

Cohort

We identified 686 eligible HHiT participants (6,948 housing episodes, 3,443 of which were homeless) and randomly selected a further 137,200 individuals from the RPDB (137,200 housing episodes) to generate a total cohort of 137,886 individuals contributing 144,148 housing episodes (Figure 1). HHiT participants experienced homelessness for, on average, 40.4% of their overall participation period, with a median homeless episode of 75 days (Interquartile range [IQR]: 29 to 181 days)(Table 1). We found substantial differences between the HHiT and general population samples, with HHiT participants being younger, more likely male, less likely to have recently immigrated and having more chronic health conditions and recent healthcare use.

Validation Results

Algorithm sensitivities when identifying a homeless housing episode (among 144,148 total episodes) ranged from 10.8% to 28.9%, with specificities exceeding 99% (Table 2). Extending time intervals or including postal code indicators in health services databases increased sensitivity, while marginally decreasing specificity. The use of all databases, as opposed to only CIHI databases, resulted in negligible gains in sensitivity. Positive likelihood ratios were all in excess of 10, indicating a substantial increase in the likelihood of homelessness following a positive test (32). Based on these findings, we chose *any CIHI database indicator +/- 45 days* as the optimal algorithm based on its scalability and optimized sensitivity, specificity and positive predictive values. More false-positives (n=595) using this algorithm came from the HHiT sample (n=397, or 66.7% of false positives) than the general population sample (n=238) (Supplement Table 5A). Absence of a healthcare encounter during the reference period accounted for 64.5% (n= 1,825) of false negatives.

Algorithm sensitivities when identifying homeless annual housing experiences (n=491,213 total calendar years) ranged from 18.5% to 35.6%, with specificities at 99.9% (Table 2). Positive likelihood ratios were all in excess of 200, indicating a very substantial increase in the probability of homelessness following a positive test (32). Sensitivity increased without impacting specificity when time windows were extended or when postal code indicators during healthcare encounters or in PSTLYEAR were included. The use of all databases, as opposed to solely CIHI databases, resulted in negligible gains in sensitivity.

The algorithm that maximized validation statistics was *any CIHI database indicator +/- 15 days or a PSTLYEAR postal code*. Most false-positives (n=365) using this algorithm were sourced from the general population sample (n=250; 68.5% of false positives overall)(Supplement Table 5B). Absence of a health encounter within the reference period accounted for 62.7% (or 997) of false negatives. However, because this algorithm requires a comprehensive database of postal codes uniquely identifying shelters or hospitals to be scaled, we deemed this algorithm suboptimal and therefore opted to use *any CIHI database indicator +/- 15 days* for generating provincial estimates.

Estimates of homelessness

Applying the optimal annual housing experience algorithm, we identified 11,731 Ontarians experiencing homelessness during 2016 (Table 3). Flagged individuals were predominantly male (70%) and between the ages of 25 to 65. One in ten were recent immigrants, about one third resided in Metropolitan Toronto, and a large proportion recently received mental or substance use-related health care (25.7% for psychotic disorders; 54.8% for non-psychotic disorders and 41.9% for substance use disorders). Over 10 years, we identified a total of 54,873 adults who experienced homelessness, of which 18,217 (33.2%) were detected in more than one year (Supplement Table 5C).

Correcting for sensitivity, we estimate a total 2016 homeless population of 59,974 (95% CI: 55,231 to 65,208) Ontarians, or 0.53% of the adult Ontario population (Figure 2). Between 2007 and 2016, the number and rate of individuals experiencing homelessness increased by 67.3% and 48.1%, respectively, with an annual percentage increase of 4.4% in the estimated rate of homelessness (95% CI: 4.2% - 4.7%).

## Discussion

We validated health administrative database algorithms for homelessness against the known housing status of individuals in a longitudinally collected, representative sample at risk for homelessness and a random sample of housed individuals in Ontario, Canada. We tested our algorithms' ability to identify individuals during an experience of homelessness and during a year in which homelessness occurred, as either definition could be used for different purposes (research and surveillance, respectively). In both cases, algorithms exhibited low sensitivity but excellent specificities and positive likelihood ratios. The low sensitivity of the algorithms can be partially explained by the large proportion of reference periods without a healthcare encounter, which increased false-negatives by default. This reaffirms the consensus that homelessness is ephemeral for many individuals and difficult to capture (1, 3, 5).

Our population prevalence estimates suggest substantial increases in homelessness between 2007 and 2016, both in absolute and relative terms. No Ontario-specific statistics exist against which to directly compare our estimate; however, if we assume Ontario's "share" of Canadian homelessness as recently reported (33) reflects its overall share of the Canadian population (38.3% in 2016)(34), approximately 90,000 homeless individuals would be attributable to Ontario in 2016, compared to our 2016 estimate of approximately 60,000. However, individuals identified as homeless in our algorithm share similar demographics with individuals in that report: approximately 25% in both sources are ages 50 and older; 16-19% are youth; and roughly 30% are women (33). Furthermore, one in three individuals were identified in multiple years, similar to the proportion of individuals using shelters in multiple years reported recently (35).

This is the first study to validate health administrative data algorithms against a reference standard with the intended purpose of population-surveillance. Most prior work (36-43) identified homelessness using homeless indicators or shelter addresses given during healthcare encounters, assuming these data represented true housing status. Recently, Vickery et al. validated addresses indicative of homelessness during healthcare encounters against self-reported housing status in a sample of Medicaid recipients, finding sensitivities between 30% and 76% and specificities between 79% and 97% (44). However, this study required the use of location- and time- specific shelter address registries, making the methodology challenging to scale or generalize. Moreover, this study's results refer to the population using healthcare (rather than the population overall) and assumed self-reported housing status did not vary over the nearly four year study period.

We readily acknowledge some limitations to this validation. First, we could only validate homelessness among individuals eligible for Ontario healthcare coverage, which although near-complete (>99%) does not include recent arrivals to Ontario, First Nations on reserves, Inuit, certain refugee claimant groups, inmates in federal penitentiaries, eligible veterans and serving members of the Canadian Forces. Since veterans and First Nations, Metis and Inuit individuals are believed to be over-represented among the homeless (20, 33, 35, 45), our algorithms almost certainly underestimate homelessness in these populations, which may account for the gap between our population estimate and the estimate loosely calculated from the *State of Homelessness in Canada 2016* (33). However, this gap is the result of linkage through Ontario-specific identifiers rather than an inherent limitation of the indicators: future pan-Canadian homelessness surveillance and research can include these populations by accessing these indicators through CIHI.

Second, we were forced to assume our general population sample was housed during the entirety of their assigned housing period. It is possible despite our screening efforts that some individuals experienced homelessness during their participation in this study. Upon review of the false positives, we identified 238 individuals from the general population sample (0.17% of that sample) who may have thus been misclassified. However, given the low prevalence of homelessness the impact of such individuals should be negligible to our overall findings.

Despite the recent Canadian federal government commitment of $2.2 billion over 10 years to tackle homelessness (46), current costs associated with enumeration (47) and program evaluation are high, necessarily reducing funding for program implementation. Overall, our algorithms present, despite their low sensitivity, important potential cost-savings opportunities as a homelessness enumeration and surveillance tool. Moreover, these algorithms can track individuals over time and be used to evaluate efforts to improve housing and health status, similar to applications from other previous validation work for population surveillance (48-52).

## Acknowledgements

This study was conducted at the ICES Western Site, which receives core funding from the Academic Medical Organization of Southwestern Ontario (AMOSO), the Schulich School of Medicine and Dentistry, Western University, and the Lawson Health Research Institute. ICES is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care. This project was funded by the Academic Health Science Centre Alternate Funding Innovation Fund, the Ontario Trillium Foundation – Local Poverty Reduction Fund, and two Canadian Institutes of Health Research grants. The opinions, results, and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by ICES or the funding agencies is intended or should be inferred. Parts of this material are based on data and/or information compiled and provided by CIHI. However, the analyses, conclusions, opinions and statements expressed in the material are those of the author(s), and not necessarily those of CIHI. Information in the Immigration, Refugees and Citizenship Canada dataset were provided by Immigration, Refugees and Citizenship Canada. We gratefully acknowledge the HHiT team for the use of their data, and Emily Kovacs for compiling postal codes identifying shelter services in Toronto and Ottawa.

## References

1. Fazel S, Geddes JR, Kushel M. The health of homeless people in high-income countries: descriptive epidemiology, health consequences, and clinical and policy recommendations. The Lancet. 2014 Oct 25;384(9953):1529-40.

2. Hwang SW, Chambers C, Chiu S, Katic M, Kiss A, Redelmeier DA, Levinson W. A comprehensive assessment of health care utilization among homeless adults under a system of universal health insurance. American Journal of Public Health. 2013 Dec;103(S2):S294-301.

3. Fitzgerald ST, Shelley MC, Dail PW. Research on homelessness: Sources and implications of uncertainty. American Behavioral Scientist. 2001 Sep;45(1):121-48.

4. Stefancic A, Schaefer-McDaniel NJ, Davis AC, Tsemberis S. Maximizing follow-up of adults with histories of homelessness and psychiatric disabilities. Evaluation and Program Planning. 2004 Nov 1;27(4):433-42.

5. Booth S. Researching health and homelessness: methodological challenges for researchers working with a vulnerable, hard to reach, transient population. Australian Journal of Primary Health. 1999;5(3):76-81.

6. Laflamme F. Survey of homeless in Canada: Street component feasibility study. Ottawa, ON: Statistics Canada Statistical Consultation Group; 2001.

7. Canadian Definition of Homelessness. Canadian Observatory on Homelessness. www.homelesshub.ca/CHRNhomelessdefinition. Published 2012. Accessed January 2, 2019.

8. Employment and Social Development Canada. Everyone counts: A Guide to Point-in-Time Counts in Canada – 2nd Edition. http://www12.esdc.gc.ca/sgpe-pmps/servlet/sgpp-pmpspub?lang=eng&curjsp=p.5bd.2t.1.3ls@eng.jsp&curactn=dwnld&pid=55638&did=4956. Published 2017. Accessed January 3, 2019.

9. Burnes D, DiLeo D, eds. Ending homelessness: why we haven't, how we can. Boulder, Colorado: Lynne Rienner; 2016.

10. Gabbard WJ, Snyder CS, Lin MB, Chadha JH, May JD, Jaggers J. Methodological issues in enumerating homeless individuals. Journal of Social Distress and the Homeless. 2007 May 1;16(2):90-103.

11. CIHI. Canadian Institute for Health Information. https://www.cihi.ca/en. Updated 2018. Accessed January 3, 2019.

12. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. Journal of clinical epidemiology. 2011 Aug 1;64(8):821-9.

13. Carrara G, Scirè CA, Zambon A, Cimmino MA, Cerra C, Caprioli M, et al. A validation study of a new classification algorithm to identify rheumatoid arthritis using administrative health databases: Case–control and cohort diagnostic accuracy studies. Results from the RECord linkage On Rheumatic Diseases study of the Italian Society for Rheumatology. BMJ open. 2015 Jan 1;5(1):e006029.

14. Butt DA, Tu K, Young J, Green D, Wang M, Ivers N, et al. A validation study of administrative data algorithms to identify patients with Parkinsonism with prevalence and incidence trends. Neuroepidemiology. 2014;43(1):28-37.

15. Widdifield J, Bernatsky S, Paterson JM, Tu K, Ng R, Thorne JC, et al. Accuracy of Canadian health administrative databases in identifying patients with rheumatoid arthritis: a validation study using the medical records of rheumatologists. Arthritis care & research. 2013 Oct 1;65(10):1582-91.

16. ICES. www.ices.on.ca. Updated 2019. Accessed January 2, 2019.

17. Hwang SW, Aubry T, Palepu A, Farrell S, Nisenbaum R, Hubley AM, et al. The health and housing in transition study: a longitudinal study of the health of homeless and vulnerably housed adults in three Canadian cities. International journal of public health. 2011 Dec 1;56(6):609-23.

18. Chan J, Mack DR, Manuel DG, Mojaverian N, de Nanassy J, Benchimol EI. Validation of an algorithm to identify children with biopsy-proven celiac disease from within health administrative data: An assessment of health services utilization patterns in Ontario, Canada. PloS one. 2017 Jun 29;12(6):e0180338.

19. Ducharme R, Benchimol EI, Deeks SL, Hawken S, Fergusson DA, Wilson K. Validation of diagnostic codes for intussusception and quantification of childhood intussusception incidence in Ontario, Canada: a population-based study. The Journal of pediatrics. 2013 Oct 1;163(4):1073-9.

20. Gaetz S, Donaldson J, Richter T, Gulliver T. The State of Homelessness in Canada 2013. Toronto: Canadian Homelessness Research Network Press.

21. Harris M, Gadermann A, Norena M, et al. Residential moves and its association with substance use, healthcare needs, and acute care use among homeless and vulnerably housed persons in Canada. International Journal of Public Health. 2018;1-11. doi: 10.1007/s00038-018-1167-6.

22. ICES. Data Dictionary. https://datadictionary.ices.on.ca/Applications/DataDictionary/Default.aspx. Updated 2019. Accessed January 3, 2019.

23. Canadian Institute for Health Information. Recording Postal Codes in the DAD and NACRS. https://www.cihi.ca/en/recording-postal-codes-in-dad-and-nacrs_en.pdf. Published N.D. Accessed December 27, 2018.

24. Robinson JR, Young TK, Roos LL, Gelskey DE. Estimating the burden of disease: comparing administrative data and self-reports. Medical care. 1997 Sep 1:932-47.

25. Hoogenveen R, Westert G, Dijkgraaf M, Schellevis F, de Bakker D. Disease prevalence estimations based on contact registrations in general practice. Statistics in medicine. 2002 Aug 15;21(15):2271-85.

26. Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying individuals with physician diagnosed COPD in health administrative databases. COPD. 2009;6(5):388-394.

27. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: Determination of prevalence and incidence of diabetes using a validated administrative data algorithm. Diabetes Care. 2002;25(3):512-516.

28. Schultz SE, Rothwell DM, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: A validation study using primary care patient records. Chronic Dis Inj Can. 2013;33(3):160-166.

29. Antoniou T, Zagorski B, Loutfy MR, Strike C, Glazier RH. Validation of case-finding algorithms derived from administrative data for identifying adults living with human immunodeficiency virus infection. Plos One. 2011;6(6):e21748.

30. Wilson EB. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association. 1927;22(158):209-212.

31. SAS Enterprise Guide. V9.4. Cary, NC: SAS Institute Inc.

32. McGee, S. Simplifying likelihood ratios. Journal of general internal medicine. 2002; 17(8): 647-650.

33. Gaetz S, Dej E, Richter T, Redman M. The State of Homelessness in Canada 2016. Toronto: Canadian Observatory on Homelessness Press.

34. Statistics Canada. Ontario [Province] and Canada [Country] (table) Census Profile. 2016 Census. Statistics Canada Catalogue no. 98-316-X2016001. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E. Released November 29, 2017.

35. Gaetz S, Gulliver T, Richter T. The state of homelessness in Canada 2014. Toronto: Canadian Observatory on Homelessness Press.

36. Peterson R, Gundlapalli AV, Metraux S, et al. Identifying homelessness among veterans using VA administrative data: Opportunities to expand detection criteria. PloS one. 2015;10(7):e0132664.

37. Tsai J, Doran KM, Rosenheck RA. When health insurance is not a factor: National comparison of homeless and nonhomeless US veterans who use Veterans Affairs Emergency Departments. Am J Public Health. 2013;103(S2):S225-S231.

38. Bharel M, Lin WC, Zhang J, O'Connell E, Taube R, Clark RE. Health care utilization patterns of homeless individuals in Boston: Preparing for Medicaid expansion under the Affordable Care Act. Am J Public Health. 2013;103(S2):S311-S317.

39. Zech J, Husk G, Moore T, Kuperman GJ, Shapiro JS. Identifying homelessness using health information exchange data. Journal of the American Medical Informatics Association. 2015;22(3): 682-687.

40. Romero-Ortuño R, O'Riordan D, Silke B. Profiling the medical admissions of the homeless. Acute Med. 2012;11:197-204.

41. Gundlapalli AV, Carter ME, Palmer M et al. Using Natural Language Processing on the Free Text of Clinical Documents to Screen for Evidence of Homelessness Among US Veterans. AMIA Annual Symposium Proceedings. 2013: 537–546.

42. Erickson J, Abbott K, Susienka L. Automatic address validation and health record review to identify homeless Social Security disability applicants. Journal of biomedical informatics. 2018;2: 41-46.

43. Government of Alberta. Total Inpatient Costs by Homeless Status. Health Trends Alberta. 2018.

44. Vickery KD, Shippee ND, Bodurtha P, et al. Identifying homeless Medicaid enrollees using enrollment addresses. Health services research. 2018;53(3):1992-2004.

45. Patrick C. Aboriginal Homelessness in Canada: A Literature Review. http://homelesshub.ca/sites/default/files/AboriginalLiteratureReview.pdf. Published 2014. Accessed January 3, 2019.

46. Government of Canada. Reaching Home: Canada's Homelessness Strategy. https://www.canada.ca/en/employment-social-development/news/2018/11/reaching-home-canadas-homelessness-strategy.html. Published 2018. Accessed January 3, 2019.

47. City of Toronto. Backgrounder: Results of the 2018 Street Needs Assessment. https://www.toronto.ca/home/media-room/backgrounders-other-resources/backgrounder-results-of-the-2018-street-needs-assessment/ Published 2018. Accessed December 3, 2018.

48. Public Health Agency of Canada. The Canadian Chronic Disease Surveillance System – An Overview. https://www.canada.ca/en/public-health/services/publications/canadian-chronic-disease-surveillance-system-factsheet.html. Published 2018. Accessed January 3, 2019.

49. Government of Canada. Publications – Diseases and Conditions. https://www.canada.ca/en/services/health/publications/diseases-conditions.html. Published 2018. Accessed January 3, 2019.

50. Lix L, Ayles J, Bartholomew S, et al. The Canadian chronic disease surveillance system: A model for collaborative surveillance. International Journal of Population Data Science. 2018;3(3). doi: https://doi.org/10.23889/ijpds.v3i3.433

51. El Emam K, Samet S, Hu J, et al. A protocol for the secure linking of registries for HPV surveillance. PLoS One. 2012;7(7):e39915.

52. Wilson K, Hawken S, Kwong JC, Deeks SL, Crowcroft NS, Manuel D. Vaccine and immunization surveillance in Ontario (VISION): Using linked health administrative databases to monitor vaccine safety. Vaccine. 2018;30(43):6115-6120

**Table 1 – Cohort characteristics at the start of a randomly selected housing episode, by source**

| Characteristic | Validation Participants (N=137,886) | HHiT sample Study (N=686) | General population sample (N=137,200) | P-value |
|---|---|---|---|---|
| Mean % (SD) of period spent homeless | n/a | 40.4% (29.4%) | n/a | n/a |
| Median days (IQR) of homelessness episode | n/a | 75 (29 – 181) | n/a | n/a |
| Age, mean (SD) | 46.1 (18.0) | 43.5 (10.6) | 46.1 (18.0) | < 0.001 |
| Female, n (%) | 70,535 (51.2%) | 208 (30.3%) | 70,327 (51.3%) | < 0.001 |
| Located in Ottawa, n (%) | 104,059 (75.5%) | 357 (52%) | 103,702 (75.6%) | < 0.001 |
| Located in Toronto, n (%) | 33,827 (24.5%) | 329 (48%) | 33,498 (24.4%) | < 0.001 |
| Recent immigrant, n (%) | 32,657 (23.7%) | 45 (6.6%) | 32,612 (23.8%) | < 0.001 |
| Date of immigration, n (%) | | | | |
| < 1 year | 1,152 (0.8%) | <=5 | NR | < 0.001 |
| 1 to 3 years | 2,381 (1.7%) | <=5 | NR | |
| 4-10 years | 9,606 (7.0%) | 9 (1.3%) | 9,597 (7.0%) | |
| Over 10 years | 19,518 (14.2%) | 33 (4.8%) | 19,485 (14.2%) | |
| Refugee status, n (%) | 5,907 (4.3%) | 18 (2.6%) | 5,889 (4.3%) | < 0.001 |
| Congestive heart failure, n (%) | 2,186 (1.6%) | 6 (0.9%) | 2,180 (1.6%) | 0.14 |
| Chronic obstructive pulmonary disease, n (%) | 6,627 (4.8%) | 91 (13.3%) | 6,536 (4.8%) | < 0.001 |
| Diabetes, n (%) | 11,332 (8.2%) | 67 (9.8%) | 11,265 (8.2%) | 0.14 |
| HIV, n (%) | 402 (0.3%) | 30 (4.4%) | 372 (0.3%) | < 0.001 |
| Chronic kidney disease[1], n (%) | 2,431 (1.8%) | 20 (2.9%) | 2,411 (1.8%) | 0.02 |
| Chronic liver disease[1], n (%) | 2,939 (2.1%) | 87 (12.7%) | 2,852 (2.1%) | < 0.001 |
| Mental health related care[2], n (%) | | | | |
| Psychotic disorders | 928 (0.7%) | 76 (11.1%) | 852 (0.6%) | < 0.001 |
| Non-psychotic disorders | 15,128 (11.0%) | 248 (36.2%) | 14,880 (10.8%) | < 0.001 |
| Substance use disorders | 1,640 (1.2%) | 204 (29.7%) | 1,436 (1.0%) | < 0.001 |
| Charlson comorbidity index, n (%) | | | | |
| 0 | 7,866 (5.7%) | 86 (12.5%) | 7,780 (5.7%) | < 0.001 |
| 1 | 1,589 (1.2%) | 25 (3.6%) | 1,564 (1.1%) | |
| 2+ | 2,476 (1.8%) | 25 (3.6%) | 2,451 (1.8%) | |
| No Hospitalizations | 125,955 (91.3%) | 550 (80.2%) | 125,405 (91.4%) | |
| Primary care visits[2], mean (SD) | 13.0 (17.5) | 21.1 (31.7) | 12.9 (17.4) | < 0.001 |
| Emergency department visits[2], mean (SD) | 1.6 (1.7) | 3.9 (5.1) | 1.6 (1.5) | < 0.001 |
| Hospitalizations[2], mean (SD) | 1.3 (0.9) | 1.7 (1.4) | 1.3 (0.9) | < 0.001 |

1. Within past 3 years; 2. Occurring in the past year. Cells representing <=5 individuals are suppressed to protect participant privacy. Individual immigration status defined based on presence of a landing date in the Immigration, Refugees and Citizenship Canada Permanent Resident Database from 1985 to 2018. NR = Not reportable, due to associated small cell suppression; NS=Not significant; HIV=Human immunodeficiency virus

**Table 2 – Accuracy of case ascertainment algorithms in identifying individuals experiencing homelessness**

***Reference Standard Definition: Housing Episode (n = 144,148 overall, with 3,443 homeless episodes)***

| Algorithm Definition | TP | FP | FN | TN | Sensitivity (%) (95% CI) | Specificity (%) (95% CI) | PPV (%) (95% CI) | NPV (%) (95% CI) | LR+ |
|---|---|---|---|---|---|---|---|---|---|
| 1 indicator +/- 0 days | 372 | 528 | 3,071 | 140,177 | 10.8 (9.8 - 11.9) | 99.6 (99.6 - 99.7) | 41.3 (38.2 - 44.6) | 97.9 (97.8 - 97.9) | 28.8 |
| 1 indicator +/- 15 days | 482 | 591 | 2,961 | 140,114 | 14.0 (12.9 - 15.2) | 99.6 (99.5 - 99.6) | 44.9 (42.0 - 47.9) | 97.9 (97.9 – 98.0) | 33.3 |
| 1 indicator +/- 45 days | 619 | 665 | 2,824 | 140,040 | 18.0 (16.7 - 19.3) | 99.5 (99.5 - 99.6) | 48.2 (45.5 - 50.9) | 98.0 (98.0 - 98.1) | 38.0 |
| 1 indicator +/- 90 days | 718 | 765 | 2,725 | 139,940 | 20.9 (19.5 - 22.2) | 99.5 (99.4 - 99.5) | 48.4 (45.9 – 51.0) | 98.1 (98.0 - 98.2) | 38.4 |
| 1 indicator +/- 180 days | 861 | 897 | 2,582 | 139,808 | 25.0 (23.6 - 26.5) | 99.4 (99.3 - 99.4) | 49.0 (46.6 - 51.3) | 98.2 (98.1 - 98.3) | 39.2 |
| 1 indicator OR postal code +/- 0 days | 450 | 679 | 2,993 | 140,026 | 13.1 (12.0 - 14.2) | 99.5 (99.5 - 99.6) | 39.9 (37.0 - 42.7) | 97.9 (97.8 – 98.0) | 27.1 |
| 1 indicator OR postal code +/- 15 days | 572 | 758 | 2,871 | 139,947 | 16.6 (15.4 - 17.9) | 99.5 (99.4 - 99.5) | 43.0 (40.4 - 45.7) | 98.0 (97.9 - 98.1) | 30.8 |
| 1 indicator OR postal code +/- 45 days | 714 | 845 | 2,729 | 139,860 | 20.7 (19.4 - 22.1) | 99.4 (99.4 - 99.4) | 45.8 (43.3 - 48.3) | 98.1 (98.0 - 98.2) | 34.5 |
| 1 indicator OR postal code +/- 90 days | 824 | 967 | 2,619 | 139,738 | 23.9 (22.5 - 25.4) | 99.3 (99.3 - 99.4) | 46.0 (43.7 - 48.3) | 98.2 (98.1 - 98.2) | 34.8 |
| 1 indicator OR postal code +/- 180 days | 994 | 1,135 | 2,449 | 139,570 | 28.9 (27.4 - 30.4) | 99.2 (99.1 - 99.2) | 46.7 (44.6 - 48.8) | 98.3 (98.2 - 98.3) | 35.8 |
| 1 CIHI indicator +/- 0 days | 368 | 466 | 3,075 | 140,239 | 10.7 (9.7 - 11.8) | 99.7 (99.6 - 99.7) | 44.1 (40.8 - 47.5) | 97.9 (97.8 - 97.9) | 36.9 |
| 1 CIHI indicator +/- 15 days | 477 | 528 | 2,966 | 140,177 | 13.9 (12.7 – 15.0) | 99.6 (99.6 - 99.7) | 47.5 (44.4 - 50.6) | 97.9 (97.9 – 98.0) | 39.6 |
| **1 CIHI indicator +/- 45 days** | **613** | **595** | **2,830** | **140,110** | **17.8 (16.6 - 19.1)** | **99.6 (99.5 - 99.6)** | **50.7 (47.9 - 53.6)** | **98.0 (97.9 - 98.1)** | **42.0** |
| 1 CIHI indicator +/- 90 days | 710 | 693 | 2,733 | 140,012 | 20.6 (19.3 – 22.0) | 99.5 (99.5 - 99.5) | 50.6 (48.0 - 53.2) | 98.1 (98.0 - 98.2) | 41.7 |
| 1 CIHI indicator +/- 180 days | 852 | 822 | 2,591 | 139,883 | 24.8 (23.3 - 26.2) | 99.4 (99.4 - 99.5) | 50.9 (48.5 - 53.3) | 98.2 (98.1 - 98.3) | 41.8 |
| 1 CIHI indicator OR postal code +/- 0 days | 444 | 575 | 2999 | 140130 | 12.9 (11.8 - 14.1) | 99.6 (99.6 - 99.6) | 43.6 (40.6 - 46.6) | 97.9 (97.8 – 98.0) | 32.3 |
| 1 CIHI indicator OR postal code +/- 15 days | 566 | 652 | 2877 | 140,053 | 16.4 (15.2 - 17.7) | 99.5 (99.5 - 99.6) | 46.5 (43.7 - 49.3) | 98.0 (97.9 - 98.1) | 36.9 |
| 1 CIHI indicator OR postal code +/- 45 days | 707 | 734 | 2736 | 139,971 | 20.5 (19.2 - 21.9) | 99.5 (99.4 - 99.5) | 49.1 (46.5 - 51.6) | 98.1 (98.0 - 98.2) | 42.1 |
| 1 CIHI indicator OR postal code +/- 90 days | 817 | 852 | 2626 | 139,853 | 23.7 (22.3 - 25.2) | 99.4 (99.4 - 99.4) | 49.0 (46.6 - 51.3) | 98.2 (98.1 - 98.2) | 41.9 |
| 1 CIHI indicator OR postal code +/- 180 days | 985 | 1017 | 2458 | 139,688 | 28.6 (27.1 - 30.1) | 99.3 (99.2 - 99.3) | 49.2 (47.0 - 51.4) | 98.3 (98.2 - 98.3) | 42.4 |

***Reference Standard Definition: Annual Housing Experience (n = 491,213 calendar years overall, with 2,290 homeless years)***

| Algorithm Definition | TP | FP | FN | TN | Sensitivity (%) (95% CI) | Specificity (%) (95% CI) | PPV (%) (95% CI) | NPV (%) (95% CI) | LR+ |
|---|---|---|---|---|---|---|---|---|---|
| 1 indicator +/- 0 days | 429 | 334 | 1,861 | 488,589 | 18.7 (17.2 - 20.4) | 99.9 (99.9 - 99.9) | 56.2 (52.7 - 59.7) | 99.6 (99.6 - 99.6) | 274.2 |
| 1 indicator +/- 15 days | 454 | 352 | 1,836 | 488,571 | 19.8 (18.2 - 21.5) | 99.9 (99.9 - 99.9) | 56.3 (52.9 - 59.7) | 99.6 (99.6 - 99.6) | 275.4 |
| 1 indicator +/- 45 days | 487 | 406 | 1,803 | 488,517 | 21.3 (19.6 – 23.0) | 99.9 (99.9 - 99.9) | 54.5 (51.3 - 57.8) | 99.6 (99.6 - 99.6) | 256.1 |
| 1 indicator +/- 90 days | 529 | 472 | 1,761 | 488,451 | 23.1 (21.4 - 24.9) | 99.9 (99.9 - 99.9) | 52.8 (49.8 - 55.9) | 99.6 (99.6 - 99.7) | 239.3 |
| 1 indicator +/- 180 days | 590 | 588 | 1,700 | 488,335 | 25.8 (24.0 - 27.6) | 99.9 (99.9 - 99.9) | 50.1 (47.2 - 52.9) | 99.7 (99.6 - 99.7) | 214.2 |
| 1 indicator OR postal code +/- 0 days | 512 | 433 | 1,778 | 488,490 | 22.4 (20.7 - 24.1) | 99.9 (99.9 - 99.9) | 54.2 (51.0 - 57.3) | 99.6 (99.6 - 99.7) | 252.5 |
| 1 indicator OR postal code +/- 15 days | 543 | 458 | 1,747 | 488,465 | 23.7 (22.0 - 25.5) | 99.9 (99.9 - 99.9) | 54.2 (51.1 - 57.3) | 99.6 (99.6 - 99.7) | 253.1 |
| 1 indicator OR postal code +/- 45 days | 581 | 525 | 1,709 | 488,398 | 25.4 (23.6 - 27.2) | 99.9 (99.9 - 99.9) | 52.5 (49.6 - 55.5) | 99.7 (99.6 - 99.7) | 236.3 |
| 1 indicator OR postal code +/- 90 days | 629 | 610 | 1,661 | 488,313 | 27.5 (25.7 - 29.3) | 99.9 (99.9 - 99.9) | 50.8 (48.0 - 53.5) | 99.7 (99.6 - 99.7) | 220.2 |
| 1 indicator OR postal code +/- 180 days | 707 | 754 | 1,583 | 488,169 | 30.9 (29.0 - 32.8) | 99.9 (99.8 - 99.9) | 48.4 (45.8 – 51.0) | 99.7 (99.7 - 99.7) | 200.2 |
| 1 indicator +/- 0 days OR PSTLYEAR postal code | 588 | 356 | 1,702 | 488,567 | 25.7 (23.9 - 27.5) | 99.9 (99.9 - 99.9) | 62.3 (59.2 - 65.3) | 99.7 (99.6 - 99.7) | 352.6 |
| 1 indicator +/- 15 days OR PSTLYEAR postal code | 706 | 402 | 1,584 | 488,521 | 30.8 (29.0 - 32.8) | 99.9 (99.9 - 99.9) | 63.7 (60.8 - 66.5) | 99.7 (99.7 - 99.7) | 375.0 |
| 1 indicator +/- 45 days OR PSTLYEAR postal code | 734 | 452 | 1,556 | 488,471 | 32.1 (30.2 – 34.0) | 99.9 (99.9 - 99.9) | 61.9 (59.1 - 64.6) | 99.7 (99.7 - 99.7) | 346.7 |
| 1 indicator +/- 90 days OR PSTLYEAR postal code | 766 | 518 | 1,524 | 488,405 | 33.4 (31.5 - 35.4) | 99.9 (99.9 - 99.9) | 59.7 (56.9 - 62.3) | 99.7 (99.7 - 99.7) | 315.7 |
| 1 indicator +/- 180 days OR PSTLYEAR postal code | 816 | 633 | 1,474 | 488,290 | 35.6 (33.7 - 37.6) | 99.9 (99.9 - 99.9) | 56.3 (53.7 - 58.8) | 99.7 (99.7 - 99.7) | 275.2 |
| 1 CIHI indicator +/- 0 days | 423 | 300 | 1,867 | 488,623 | 18.5 (16.9 - 20.1) | 99.9 (99.9 - 99.9) | 58.5 (54.9 – 62.0) | 99.6 (99.6 - 99.6) | 301.0 |
| **1 CIHI indicator +/- 15 days** | **448** | **315** | **1,842** | **488,608** | **19.6 (18.0 - 21.2)** | **99.9 (99.9 - 99.9)** | **58.7 (55.2 - 62.2)** | **99.6 (99.6 - 99.6)** | **303.6** |
| 1 CIHI indicator +/- 45 days | 480 | 358 | 1,810 | 488,565 | 21.0 (19.3 - 22.7) | 99.9 (99.9 - 99.9) | 57.3 (53.9 - 60.6) | 99.6 (99.6 - 99.6) | 286.3 |
| 1 CIHI indicator +/- 90 days | 521 | 405 | 1,769 | 488,518 | 22.8 (21.1 - 24.5) | 99.9 (99.9 - 99.9) | 56.3 (53.0 - 59.4) | 99.6 (99.6 - 99.7) | 274.7 |
| 1 CIHI indicator +/- 180 days | 581 | 519 | 1,709 | 488,404 | 25.4 (23.6 - 27.2) | 99.9 (99.9 - 99.9) | 52.8 (49.9 - 55.8) | 99.7 (99.6 - 99.7) | 239.0 |
| 1 CIHI indicator OR postal code +/- 0 days | 508 | 370 | 1,782 | 488,553 | 22.2 (20.5 - 23.9) | 99.9 (99.9 - 99.9) | 57.9 (54.6 - 61.1) | 99.6 (99.6 - 99.7) | 293.1 |

| Algorithm Definition | TP | FP | FN | TN | Sensitivity (%) (95% CI) | Specificity (%) (95% CI) | PPV (%) (95% CI) | NPV (%) (95% CI) | LR+ |
|---|---|---|---|---|---|---|---|---|---|
| 1 CIHI indicator OR postal code +/- 15 days | 539 | 390 | 1,751 | 488,533 | 23.5 (21.8 - 25.3) | 99.9 (99.9 - 99.9) | 58.0 (54.8 - 61.2) | 99.6 (99.6 - 99.7) | 295.1 |
| 1 CIHI indicator OR postal code +/- 45 days | 576 | 442 | 1,714 | 488,481 | 25.2 (23.4 – 27.0) | 99.9 (99.9 - 99.9) | 56.6 (53.5 - 59.6) | 99.7 (99.6 - 99.7) | 278.2 |
| 1 CIHI indicator OR postal code +/- 90 days | 622 | 502 | 1,668 | 488,421 | 27.2 (25.4 – 29.0) | 99.9 (99.9 - 99.9) | 55.3 (52.4 - 58.2) | 99.7 (99.6 - 99.7) | 264.5 |
| 1 CIHI indicator OR postal code +/- 180 days | 699 | 634 | 1,591 | 488,289 | 30.5 (28.7 - 32.4) | 99.9 (99.9 - 99.9) | 52.4 (49.8 - 55.1) | 99.7 (99.7 - 99.7) | 235.4 |
| 1 CIHI indicator +/- 0 days OR PSTLYEAR postal code | 583 | 322 | 1,707 | 488,601 | 25.5 (23.7 - 27.3) | 99.9 (99.9 - 99.9) | 64.4 (61.2 - 67.5) | 99.7 (99.6 - 99.7) | 386.6 |
| 1 CIHI indicator +/- 15 days OR PSTLYEAR postal code | 701 | 365 | 1,589 | 488,558 | 30.6 (28.8 - 32.5) | 99.9 (99.9 - 99.9) | 65.8 (62.9 - 68.5) | 99.7 (99.7 - 99.7) | 410.0 |
| 1 CIHI indicator +/- 45 days OR PSTLYEAR postal code | 728 | 404 | 1,562 | 488,519 | 31.8 (29.9 - 33.7) | 99.9 (99.9 - 99.9) | 64.3 (61.5 – 67.0) | 99.7 (99.7 - 99.7) | 384.7 |
| 1 CIHI indicator +/- 90 days OR PSTLYEAR postal code | 760 | 451 | 1,530 | 488,472 | 33.2 (31.3 - 35.1) | 99.9 (99.9 - 99.9) | 62.8 (60.0 - 65.4) | 99.7 (99.7 - 99.7) | 359.8 |
| 1 CIHI indicator +/- 180 days OR PSTLYEAR postal code | 809 | 564 | 1,481 | 488,359 | 35.3 (33.4 - 37.3) | 99.9 (99.9 - 99.9) | 58.9 (56.3 - 61.5) | 99.7 (99.7 - 99.7) | 306.2 |

Bold lines indicate optimal case algorithm definitions. TP = True Positive (flagged as homeless and truly homeless); FP = False Positive (flagged as homeless but not truly homeless); FN = False Negative (flagged as housed but truly homeless); TN = True Negative (flagged as housed and truly housed); PPV = Positive Predictive Value; NPV = Negative Predictive Value; LR+ = Positive Likelihood Ratio; CIHI=Discharge Abstract Database, National Ambulatory Care Reporting System or Ontario Mental Health Reporting System; PSTLYEAR = ICES PSTLYEAR postal code, indicating the best estimate of an individual’s postal code for the year using ICES databases.

**Table 3 – Characteristics of individuals identified as homeless in 2016 using the optimal annual housing experience algorithm (Any CIHI indicator +/- 15 days)**

| | Individuals identified as homeless in 2016 (N = 11,731) |
|---|---|
| Age group, in years, N (%) | |
| 18 to 24 | 1,901 (16.2%) |
| 25 to 34 | 3,498 (29.8%) |
| 35 to 50 | 3,246 (27.7%) |
| 51 to 65 | 2,352 (20.1%) |
| Over 65 | 734 (6.3%) |
| Female sex, N (%) | 3,497 (29.8%) |
| City of residence in 2016, N (%) | |
| Toronto | 4,299 (36.7%) |
| Ottawa | 684 (5.8%) |
| In a rural area, N (%) | 667 (5.7%) |
| Recent immigrant, N (%) | 1,172 (10.0%) |
| Immigrated as refugee, N (%) | 366 (3.2%) |
| Charlson comorbidity index, N (%) | |
| 0 | 1,825 (15.6%) |
| 1 | 550 (4.7%) |
| 2+ | 465 (4.0%) |
| No hospitalizations | 8,891 (75.8%) |
| Comorbidities, N (%) | |
| Congestive heart failure | 222 (1.9%) |
| Chronic obstructive pulmonary disease | 1,258 (10.7%) |
| Diabetes | 1,233 (10.5%) |
| Chronic kidney disease[1] | 588 (5.0%) |
| Chronic liver disease[1] | 1,244 (10.6%) |
| HIV positive | 202 (1.7%) |
| Primary care visits[2], mean (SD) | 33.0 (43.6) |
| Emergency department visits[2], mean (SD) | 5.5 (9.2) |
| Admissions to hospital[2], mean (SD) | 1.9 (1.7) |
| Mental health related care[2], N (%) | |
| Psychotic disorders | 3,014 (25.7%) |
| Non-psychotic disorders | 6,433 (54.8%) |
| Substance use disorders | 4,917 (41.9%) |

1. Within the past 3 years; 2. Occurring in the past year; TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative; PPV = Positive Predictive Value; NPV = Negative Predictive Value; LR+ = Positive Likelihood Ratio; CIHI= Discharge Abstract Database, National Ambulatory Care Reporting System or Ontario Mental Health Reporting System

Figure 1. Cohort Build

Figure 2. Estimated number of individuals and population-prevalence (per 100 adults) experiencing homelessness in Ontario from 2007 to 2016 using the optimal annual housing experience case ascertainment algorithm (any CIHI indicator +/-15 days), with 95% confidence intervals, correcting for sensitivity. Annual Percentage Change with confidence interval was calculated using a Poisson regression

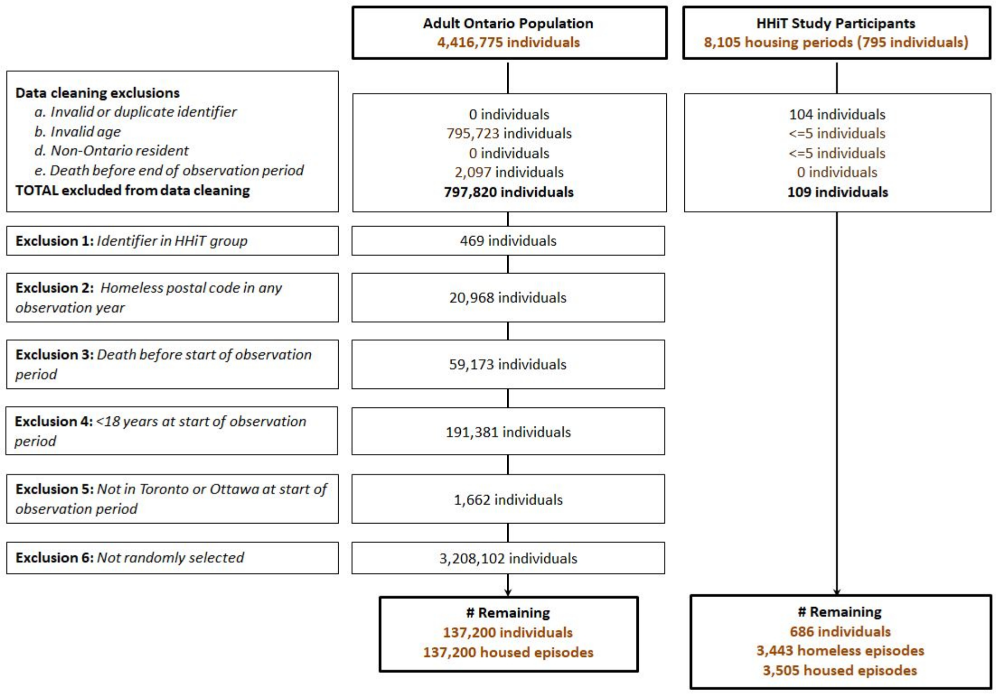



Figure 1 - Cohort build flow diagram

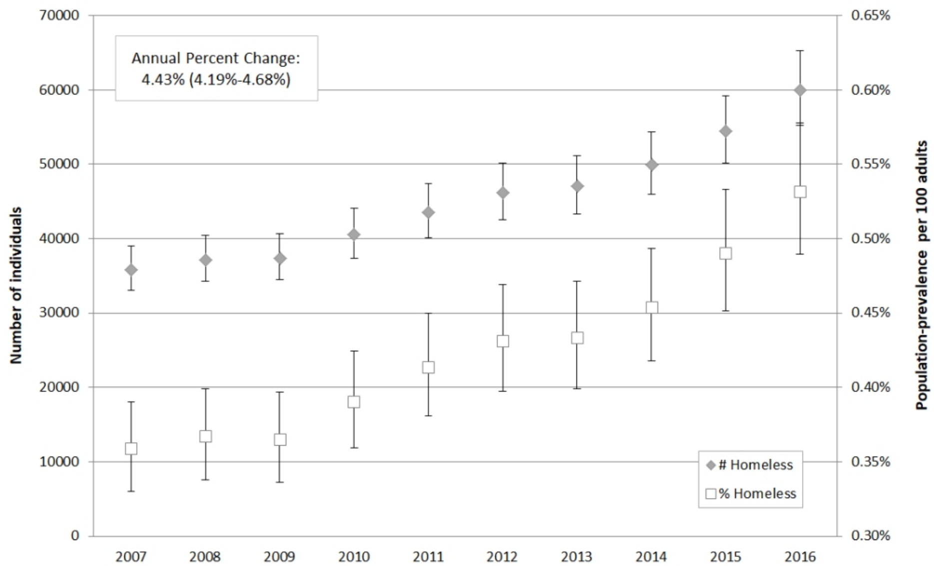



Figure 2 - Estimated population prevalence of homelessness in Ontario 2007-2016

## Supplement Table 1 – Data Elements Indicative of Homelessness, Supportive Housing or Shelter Use

| Database | Variable Name | Indicator Value | Description |
|---|---|---|---|
| **DAD** | HOMELESS | "Y" | Homelessness indicator |
| | INSTTYPE | "SH" | Institution Type = Supportive Housing |
| | DX10CODE1 to DX10CODE25 | "Z590" or "Z591" | ICD-10 diagnosis codes for "Homelessness" and "Inadequate housing" |
| | CMGDIAG | "Z590" or "Z591" | ICD-10 diagnosis codes for "Homelessness" and "Inadequate housing" |
| | PSTLCODE | "XX" | Used to indicate transient/homeless patients |
| **NACRS** | DX10CODE1 to DX10CODE10 | "Z590" or "Z591" | ICD-10 diagnosis codes for "Homelessness" and "Inadequate housing" |
| | RESTYPE | "3" or "4" | Residence Type = "Homeless" or "Shelter" |
| | PSTLCODE | "XX" | Used to indicate transient/homeless patients |
| **OMHRS** | PREDX10CODE to PREDX10CODE11 | "Z590" or "Z591" | ICD-10 diagnosis codes for "Homelessness" and "Inadequate housing" |
| | POSTDX10CODE1 to POSTDX10CODE24 | "Z590" or "Z591" | ICD-10 diagnosis codes for "Homelessness" and "Inadequate housing" |
| | PRIOR_RESIDENCE | "6" | Prior residential status = "Homeless (with or without shelter)" |
| | USUAL_RESIDENCE | "8" | Usual residential status = "Homeless (with or without shelter)" |
| | ADMITFROM | "8" | Admitted from = "Homeless (with or without shelter)" |
| | DISCHLIVING | "8" | Living arrangement at discharge = "Homeless (with or without shelter)" |
| | P5_Retired_2009 | "6" | (Variable retired in 2009) Living arrangement = "Homeless (with or without shelter)" |
| | PSTLCODE | "XX" | Used to indicate transient/homeless patients |
| **HCD** | DXCODE | "V600" or "V601" | ICD-9 diagnosis codes for "Lack of housing" or "Inadequate housing" |
| | REQUEST_PROGRAM | "6" | Program Requested = "Supportive Housing" |
| | RESIDENCE_TYPE | "1604", "2200" or "3400" | Residence Type = "Other Supportive Living Unit", "Hostel/Shelter" or "No fixed address" |
| **RAICA** | B4 | "8" | Expected residential/living status during service provision = "Homeless (with / without shelter)" |
| **NRS** | ALIVESET | "6" | Admission living setting = "Shelter" |
| | FLIVESET | "6" | Follow-up living setting = "Shelter" |

| Database | Variable Name | Indicator Value | Description |
|---|---|---|---|
| | PRIM_DISCH_WAIT_REASON | “1.1” | Primary Discharge Wait Reason = “Assisted Living/Supportive Housing” |
| | SECND_DISCH_WAIT_REASON | “1.1” | Secondary Discharge Wait Reason = “Assisted Living/Supportive Housing” |
| **CORR** | PROVINCE_CODE | “XX” | “Transient/Homeless” |
| | HEALTH_CARD_PROVINCE_CODE | “XX” | “Transient/Homeless” |

ICD=International Classification of Diseases

## Supplement Table 2: Databases Used

| Name | Data Source | Description |
|---|---|---|
| **Canadian Institute for Health Information Discharge Abstract Database (DAD)** | Canadian Institute for Health Information (CIHI) | The DAD contains administrative, clinical (diagnoses and procedures/interventions), demographic, and administrative information for all admissions to acute care hospitals in Ontario. At ICES, consecutive DAD records are linked together to form 'episodes of care' among the hospitals to which patients have been transferred after their initial admission |
| **Ontario Mental Health Reporting System (OMHRS)** | Canadian Institute for Health Information (CIHI) | The OMHRS contains administrative, clinical (diagnoses and procedures), demographic, and administrative information for all admissions to adult designated inpatient mental health beds. This includes beds in general hospitals, provincial psychiatric facilities, and specialty psychiatric facilities. Clinical assessment data is ascertained using the Resident Assessment Instrument for Mental Health (RAI-MH), but different amounts of information are collected using this instrument depending on the length of stay in the mental health bed. Multiple assessments may occur during the length of a mental health admission. |
| **National Ambulatory Care Reporting System (NACRS)** | Canadian Institute for Health Information (CIHI) | The NACRS contains administrative, clinical (diagnoses and procedures), demographic, and administrative information for all patient visits made to hospital- and community-based ambulatory care centres (emergency departments, day surgery units, hemodialysis units, and cancer care clinics) in Ontario. At ICES, NACRS records are linked with other data sources (DAD, Ontario Mental Health Reporting System [OMHRS]) to identify transitions to other care settings, such as inpatient acute care or psychiatric care. |
| **Home Care Database (HCD)** | Ontario Association of Community Care Access Centres | The Home Care Database contains administrative data about the patients, episodes, and services who receive home care through CCACs. The data included here is extracted from the CCAC administrative data system (CHRIS). |
| **Resident Assessment Instrument Contact Assessment Database (RAICA)** | Ontario Association of Community Care Access Centres | The interRAIContact Assessment (interRAICA) is a short screening assessment completed for adults at the time of intake to CCAC service (i.e. home care and / or palliative care) from community or hospital (including ED). It was designed to support decision-making about the urgency for immediate service provision, record essential clinical information on persons who would not be receiving comprehensive assessment at a later stage, and provide the minimum clinical information to enable short-term services to be put in place before completion of a full RAI assessment (ie. RAI-HC) |

| Name | Data Source | Description |
|---|---|---|
| **National Rehabilitation Reporting System (NRS)** | Ministry of Health and Long-Term Care | The National Rehabilitation Reporting System (NRS) contains client data collected from participating adult inpatient rehabilitation facilities and programs across Canada. Data elements include socio-demographic information, administrative data, patient health characteristics, activities and participation and interventions. |
| **Canadian Organ Replacement Registry (CORR)** | Canadian Institute for Health Information (CIHI) | The Ontario portion of the Canadian Organ Replacement Register (CORR) records activity and outcomes of vital organ transplantation and renal dialysis activities. |
| **ICES-derived PSTLYEAR database** | ICES; Ministry of Health and Long-Term Care | The ICES-derived PSTLYEAR database contains the best known postal code for persons in the OHIP Registered Persons Database on July 1st of each year starting from year 1991. Postal codes supplied by the Ministry of Health and Long-Term Care are enriched with information in CIHI and other ICES-housed datasets to take advantage of the postal code information recorded each time an individual accesses certain healthcare services. |
| **OHIP Registered Persons Database** | Ministry of Health and Long-Term Care | The OHIP RPDB provides basic demographic information (age, sex, location of residence, date of birth, and date of death for deceased individuals) for those issued an Ontario health insurance number. The RPDB also indicates the time periods for which an individual was eligible to receive publicly funded health insurance benefits and provides the best known postal code for each registrant on July 1st of each year. |
| **Ontario Health Insurance Plan (OHIP)** | Ministry of Health and Long-Term Care | The OHIP claims database contains information on inpatient and outpatient services provided to Ontario residents eligible for the province's publicly funded health insurance system by fee-for-service health care practitioners (primarily physicians) and "shadow billings" for those paid through non-fee-for-service payment plans. Billing codes on the claims (OHIP fee codes) identify the care provider, their area of specialization and the type and location of service. OHIP billing claims also contain a 3-digit diagnosis code - the main reason for the service - captured using a modified version of the ICD, 8th revision coding system. |
| **Immigration, Refugees, and Citizenship Canada's Permanent Resident database (IRCC)** | Immigration, Refugees and Citizenship Canada | The Ontario portion of the IRCC Permanent Resident Database includes immigration application records for people who initially applied to land in Ontario since 1985. The dataset contains permanent residents' demographic information such as country of citizenship, level of education, mother tongue, and landing date. New immigrants who are currently residing in Ontario but originally landed in another province are not captured in this dataset. |

| Name | Data Source | Description |
|---|---|---|
| **Ontario COPD Database (COPD)** | Canadian Institute for Health Information (CIHI) | The Ontario COPD Database is created using two separate algorithms applied to inpatient hospitalization (DAD), same day surgery (SDS) records, and physician billing claims (OHIP) data to determine the diagnosis date for incident cases of chronic obstructive pulmonary disease in Ontario.<br><br>In an algorithm which maximizes sensitivity, the definition for COPD is any physician billing claim with a diagnosis for COPD (OHIP diagnosis codes: 491, 492, 496) or any inpatient hospitalization or same day surgery record with a diagnosis for COPD (ICD-9 diagnosis codes: 491, 492, 496; ICD-10 diagnosis codes: J41- J44; in any diagnostic code space). When using expert panel review of primary care charts as the reference standard, this definition has been shown to have the following performance characteristics: Sensitivity (85.0%), Specificity (78.4%), Positive Predictive Value (57.5%), and Negative Predictive Value (93.8%).(7)<br><br>In an algorithm which maximizes specificity, the definition for COPD is ≥3 physician billing claims with a diagnosis for COPD (OHIP diagnosis codes: 491, 492, 496) or ≥1 inpatient hospitalization or same day surgery record with a diagnosis for COPD (ICD-9 diagnosis codes: 491, 492, 496; ICD-10 diagnosis codes: J41, J42, J43, J44; in any diagnostic code space) in a two-year period. When using expert panel review of primary care charts as the reference standard, this definition has been shown to have the following performance characteristics: Sensitivity (57.5%), Specificity (95.4%), Positive Predictive Value (81.3%), and Negative Predictive Value (86.7%).(1) |
| **Ontario Diabetes Database (ODD)** | Canadian Institute for Health Information (CIHI) | The ODD is created using algorithms applied to inpatient hospitalization (DAD) records, same day surgery (SDS) records, and physician billing claims (OHIP) data to determine the diagnosis date for incident cases of diabetes in Ontario. For adults aged 19 years and greater, the definition for diabetes is 2 physician billing claims with a diagnosis for diabetes (OHIP diagnosis code: 250) or 1 inpatient hospitalization or same day surgery record with a diagnosis for diabetes (ICD-9 diagnosis code: 250; ICD-10 diagnosis codes: E10, E11, E13, E14; in any diagnostic code space) within a 2 year period. Physician claims and hospitalizations with a diagnosis of diabetes occurring within 120 prior to and 180 days after a gestational hospitalization record were excluded. When using primary care chart abstraction as the reference standard, this definition has been shown to have the following performance characteristics: Sensitivity (86.1%), Specificity (97.1%), Positive Predictive Value (79.8%), and Negative Predictive Value (98.1%).(2) |

| Name | Data Source | Description |
| --- | --- | --- |
| **Ontario CHF Database (CHF)** | Canadian Institute for Health Information (CIHI) | The Ontario CHF Database is created using a definition of ≥2 physician billing claims with a diagnosis of congestive heart failure (OHIP diagnosis code: 428) and/or ≥1 inpatient hospitalization or same day surgery record with a diagnosis of congestive heart failure (ICD-9 diagnosis code: 428; ICD-10 diagnosis code: I50; in the primary diagnostic code space) in a two-year period applied to hospitalization (DAD), same day surgery (SDS), and physician billing claims (OHIP) data to determine the diagnosis date for incident cases of CHF in Ontario.<br><br>When using electronic medical record data abstraction as the reference standard, the above definition has been demonstrated to have the following performance characteristics: Sensitivity (84.8%), Specificity (97.0%), and Positive Predictive Value (55.3%).(3) |
| **Ontario HIV Database (HIV)** | Canadian Institute for Health Information (CIHI) | The Ontario HIV Database contains all Ontario HIV positive patients identified since 1992. HIV positive patients are defined as persons having received at least 3 physician claims with OHIP diagnosis code 042, 043, or 044 within 3 years. The diagnosis date is the first of these claims, unless a prior OHIP record with the above diagnosis codes or a hospitalization having an ICD-10 diagnosis code of B20, B21, B22, B23, or B24 occurs earlier.<br><br>This definition has been shown to have high sensitivity (96.2%) and specificity (99.6%)(4) |

# Supplement Table 3: Variable Definitions

| Variable | Data Source | Definition Description |
|---|---|---|
| **Age** | RPDB | Age of the individual at the index date |
| **Sex** | RPDB | Sex of the individual |
| **Rural status** | RPDB | Resides in a rural area as defined as a settlement of <10 000 individuals |
| **Location (city)** | RPDB | City in which the individual is believed to reside as of July 1st of the index year, based on their census division information |
| **Recent immigrant** | IRCC | Presence of a landing date in the Immigration, Refugees and Citizenship Canada Permanent Database indicates immigration to Ontario between 1985 to 2018 |
| **Date of immigration** | IRCC | Time, in years, since immigration to Ontario from outside Canada occurred |
| **Refugee status** | IRCC | Class of immigration status = Refugee |
| **Congestive heart failure** | CHF | Presence in the database indicates the individual has a history of congestive heart failure[1] |
| **Chronic obstructive pulmonary disease** | COPD | Presence in the database indicates the individual has a history of COPD[2] |
| **Diabetes** | ODD | Presence in the database indicates the individual has a history of diabetes[3] |
| **HIV status** | HIV | Presence in the database indicates the individual is HIV positive[4]. |
| **Chronic kidney disease** | DAD, NACRS, OHIP | 1 hospitalization or 3 ED visit or physician claims in 1 year within 3 years of the index date with any of the following eligible codes:<br>ICD-10: E102, E112, E132, E142, I12, I13, N00, N01, N02, N03, N04, N05, N06, N07, N08, N1, N20, N21, N22, N23<br>OHIP dx: 403, 585 |
| **Chronic liver disease** | DAD, NACRS, OHIP | 1 hospitalization, ED visit or physician claim within 3 years of the index date with any of the following eligible codes:<br>ICD-10: B16, B17, B18, B19, B942, E830, E831, I85, K70, K713, K714, K715, K717, K721, K729, K73, K74, K753, K754, K758, K759, K76, K77, R160, R162, R17, R18, Z225<br>OHIP dx: 070, 571, 573<br>OHIP fee: Z551, Z554 |
| **Psychosis related mental health care** | DAD, NACRS, OMHRS, OHIP | 1 hospitalization, ED visit or physician claim within 1 year of the index date with any of the following eligible codes:<br>ICD-10: F20, F22, F23, F24, F25, F28, F29<br>DSM-IV: 295, 297, 298<br>OHIP dx: 295, 297, 298 |

| Variable | Data Source | Definition Description |
|---|---|---|
| **Non-psychotic disorders related mental health care** | DAD,<br>NACRS,<br>OMHRS,<br>OHIP | 1 hospitalization, ED visit or physician claim within 1 year of the index date with any of the following eligible codes:<br>ICD-10: F30, F31, F32, F33, F34, F38, F39, F40, F41, F42, F43, F48, F60, F93<br>DSM-IV: 296, 300, 301<br>OHIP dx: 296, 300, 301, 309, 311 |
| **Substance use related mental health care** | DAD,<br>NACRS,<br>OMHRS,<br>OHIP | 1 hospitalization, ED visit or physician claim within 1 year of the index date with any of the following eligible codes:<br>ICD-10: F10, F11, F12, F13, F14, F15, F16, F17, F18, F19, F55<br>DSM-IV: 291, 292, 303, 304, 305<br>OHIP dx: 291, 292, 303, 304, 305 |
| **Outpatient visits** | OHIP | Number of physician visits within 1 year prior to the index date, defined as one visit per day per physician |
| **Emergency department visits** | NACRS | Number of ED visits within 1 year prior to the index date |
| **Hospitalizations** | DAD | Number of admissions to acute care hospitals within 1 year prior to the index date. |

(1) Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying individuals with physcian diagnosed COPD in health administrative databases. COPD 2009; 6(5):388-394.

(2) Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. Diabetes Care 2002; 25(3):512-516.

(3) Schultz SE, Rothwell DM, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. Chronic Dis Inj Can 2013; 33(3):160-166.

(4) Tony Antoniou, Brandon Zagorski, Mona R. Loutfy, Carol Strike, Richard H. Glazier. Validation of Case-Finding Algorithms Derived from Administrative Data for Identifying Adults Living with Human Immunodeficiency Virus Infection. Plos One. 2011;6(6):e21748. Epub 2011 Jun 30.

# Supplement Table 4 – Validation Statistic Formulae

The following diagnostic tests were used to assess the validity of each case ascertainment algorithm.

| Validation Statistic | Formula |
|---|---|
| **Sensitivity** | TP / (TP + FN) |
| **Specificity** | TN / (FP + TN) |
| **Positive Predictive Value** | TP / (TP + FP) |
| **Negative Predictive Value** | TN / (FN + TN) |
| **Positive Likelihood Ratio** | Sensitivity / (1 - Specificity) |

TP=True positive (truly experiencing homelessness and flagged as homeless by the case ascertainment algorithm)
FP=False positive (truly housed but flagged as homeless by the case ascertainment algorithm)
FN=False negative (truly experiencing homelessness but not flagged as homeless by the case ascertainment algorithm)
TN=True negative (truly housed and flagged as housed by the case ascertainment algorithm)

# Supplement Table 5 – Additional Tables

## Table 5A – Characteristics of true positives, false positives and false negatives using the optimal housing episode algorithm

| | True Positives (N=613) | False Positives (N=595) | False Negatives (N=2,830) |
|---|---|---|---|
| **Episodes without encounters, n (% of group)** | 0 (0%) | 0 (0%) | 1,825 (64.5%) |
| **Cohort source = HHiT study, n (% of group)** | 613 (100%) | 397 (66.7%) | 2,830 (100%) |

Optimal housing episode algorithm = 1 CIHI indicator +/-45 days of the housing episode start and end dates

## Table 5B – Characteristics of true positives, false positives and false negatives using the (non-scalable) optimal annual housing experience algorithm

| | True Positives (N=701) | False Positives (N=365) | False Negatives (N=1,589) |
|---|---|---|---|
| **Episodes without encounters, n (% of group)** | 0 (0%) | 0 (0%) | 997 (62.7%) |
| **Cohort source = HHiT study, n (% of group)** | 701 (100%) | 115 (31.5%) | 2,830 (100%) |

Optimal annual housing experience algorithm = 1 CIHI indicator +/-15 days of the calendar year start and end dates or one postal code from PSTLYEAR

**Table 5C – Number of adult Ontarians identified as experiencing homelessness by the optimal annual housing experience algorithm between 2007 and 2016**

| Year | # identified (95% CI) | Adult ON Population | Unadjusted Rate (95% CI) |
|---|---|---|---|
| **2007** | 7,012 (6,850-7,178) | 9,995,143 | 0.07% (0.069% - 0.072%) |
| **2008** | 7,271 (7,106-7,440) | 10,125,078 | 0.072% (0.07% - 0.073%) |
| **2009** | 7,318 (7,152-7,488) | 10,250,718 | 0.071% (0.07% - 0.073%) |
| **2010** | 7,934 (7,761-8,110) | 10,393,961 | 0.076% (0.075% - 0.078%) |
| **2011** | 8,521 (8,342-8,704) | 10,529,817 | 0.081% (0.079% - 0.083%) |
| **2012** | 9,028 (8,844-9,216) | 10,699,090 | 0.084% (0.083% - 0.086%) |
| **2013** | 9,202 (9,016-9,392) | 10,859,071 | 0.085% (0.083% - 0.086%) |
| **2014** | 9,769 (9,577-9,965) | 11,001,544 | 0.089% (0.087% - 0.091%) |
| **2015** | 10,658 (10,458-10,862) | 11,117,135 | 0.096% (0.094% - 0.098%) |
| **2016** | 11,731 (11,521-11,945) | 11,287,810 | 0.104% (0.102% - 0.106%) |
| **Total individuals identified over 10 years** | | | **54,873** |
| **Individuals present in > 1 year estimate** | | | **18,217 (33.2% of total)** |

Adult ON Population derived from Ontario inter-censal population estimates.
Optimal annual housing experience algorithm = 1 CIHI indicator +/-15 days of the calendar year start and end dates.
Confidence intervals calculated using the Wilson score method.

# Reporting checklist for diagnostic test accuracy study.

Based on the STARD guidelines.

## Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the STARD reporting guidelines, and cite them as:

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, LijmerJG Moher D, Rennie D, de Vet HCW, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF, For the STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies.

| | | Reporting Item | Page Number |
|---|---|---|---|
| | #1 | Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC) | 1 |
| | #2 | Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts) | 2 |
| | #3 | Scientific and clinical background, including the intended use and clinical role of the index test | 4 |
| | #4 | Study objectives and hypotheses | 4 |
| Study design | #5 | Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study) | 4 |
| Participants | #6 | Eligibility criteria | 5 |
| | #7 | On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry) | 5 |
| | #8 | Where and when potentially eligible participants were identified (setting, location and dates) | 5 |
| | #9 | Whether participants formed a consecutive, random or convenience series | 5 |
| Test | #10a | Index test, in sufficient detail to allow replication | 5 |

| | | | |
|---|---|---|---|
| methods | | | |
| | #10b | Reference standard, in sufficient detail to allow replication | 6 |
| | #11 | Rationale for choosing the reference standard (if alternatives exist) | 6 |
| | #12a | Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory | See note 1 |
| | #12b | Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory | 7 |
| | #13a | Whether clinical information and reference standard results were available to the performers / readers of the index test | See note 2 |
| | #13b | Whether clinical information and index test results were available to the assessors of the reference standard | See note 3 |
| Analysis | #14 | Methods for estimating or comparing measures of diagnostic accuracy | 7 |
| | #15 | How indeterminate index test or reference standard results were handled | See note 4 |
| | #16 | How missing data on the index test and reference standard were handled | 7 |
| | #17 | Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory | 7 |
| | #18 | Intended sample size and how it was determined | 5 |
| Participants | #19 | Flow of participants, using a diagram | Figure 1 |
| | #20 | Baseline demographic and clinical characteristics of participants | 8 |
| | #21a | Distribution of severity of disease in those with the target condition | 8 |
| | #21b | Distribution of alternative diagnoses in those without the target condition | See note 5 |
| | #22 | Time interval and any clinical interventions between index test and reference standard | See note 6 |
| Test results | #23 | Cross tabulation of the index test results (or their distribution) by the results of the reference standard | 8-9, Table 2 |
| | #24 | Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals) | 8-9, Table 2 |
| | #25 | Any adverse events from performing the index test or the reference standard | See note 7 |
| | #26 | Study limitations, including sources of potential bias, statistical uncertainty, and generalisability | 11-12 |

| | | |
|---|---|---|
| #27 | Implications for practice, including the intended use and clinical role of the index test | 10-12 |
| #28 | Registration number and name of registry | See note 8 |
| #29 | Where the full study protocol can be accessed | See note 9 |
| #30 | Sources of funding and other support; role of funders | See note 10 |

## Author notes

1. n/a - variables are binary
2. n/a - index test uses administrative data, i.e. there were no index test performers
3. n/a - index test uses administrative data. i.e. by definition the index test was not available to those assessing the reference standard
4. n/a - no indeterminate results were possible
5. n/a - those without target definition were assumed housed by default, as described in the Methods
6. n/a - no clinical interventions are relevant and time intervals were included in case algorithm definitions, as described in the Methods
7. n/a - not a clinical test
8. n/a - not registered
9. n/a - full protocol described in-text
10. 1 (title page), 11 (acknowledgements)

BMJ Open

# A validation study of health administrative data algorithms to identify individuals experiencing homelessness and estimate population prevalence of homelessness in Ontario, Canada

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2019-030221.R1 |
| Article Type: | Original research |
| Date Submitted by the Author: | 01-Aug-2019 |
| Complete List of Authors: | Richard, Lucie; Institute for Clinical Evaluative Sciences, N/A<br>Hwang, Stephen W. ; St Michaels Hospital, Toronto<br>Forchuk, Cheryl; Western University<br>Nisenbaum, Rosane; St. Michael's Hospital, Centre for Research on Inner City Health; University of Toronto, Dalla Lana School of Public Health<br>Clemens, Kristin; Western University<br>Wiens, Kathryn; University of Toronto<br>Booth, Richard; Western University<br>Azimaee, Mahmoud; Institute for Clinical Evaluative Sciences<br>Shariff, Salimah; Institute for Clinical Evaluative Sciences, N/A |
| <b>Primary Subject Heading</b>: | Epidemiology |
| Secondary Subject Heading: | Health policy, Health services research, Public health, Research methods |
| Keywords: | EPIDEMIOLOGY, HEALTH SERVICES ADMINISTRATION & MANAGEMENT, PUBLIC HEALTH |
| | |

# A validation study of health administrative data algorithms to identify individuals experiencing homelessness and estimate population prevalence of homelessness in Ontario, Canada

**Richard, Lucie**, MA, ICES Western
**Hwang, Stephen W.**, MD, MPH, Centre for Urban Health Solutions, Li Ka Shing Knowledge Institute, St Michael's Hospital; Dalla Lana School of Public Health, University of Toronto
**Forchuk, Cheryl**, PhD, Western University
**Nisenbaum, Rosane**, PhD, Centre for Urban Health Solutions, Applied Health Research Centre, Li Ka Shing Knowledge Institute, St Michael's Hospital; Dalla Lana School of Public Health, University of Toronto
**Clemens, Kristin**, MD, MSc, Western University
**Wiens, Kathryn**, MSc, University of Toronto
**Booth, Richard**, PhD, Western University
**Azimaee, Mahmoud**, ICES
**Shariff, Salimah Z.**, PhD, ICES Western

Corresponding author: Lucie Richard lucie.richard@ices.on.ca

Funding Statement: This study was supported by an Ontario Trillium Foundation – Local Poverty Reduction Fund grant (held by Drs. Forchuk, Booth and Shariff), an Academic Medical Organization of Southwestern Ontario Academic Health Science Centre Alternate Funding Innovation Fund grant (held by Drs. Clemens and Shariff) and two Canadian Institutes of Health Research grants, CIHR MOP-86765 and HOA-80066 (held by Drs. Hwang and Nisenbaum).

Declaration of author(s) competing interests: Dr. Clemens reports grants from Astra-Zeneca, and attendance to a Merck Inc. conference, occurring outside the submitted work. All other authors have no competing interests to declare.

Author contributions: LR conceived of the study, participated in the study coordination, study design, acquisition of data and interpretation of results, performed the analysis and drafted the manuscript. SZS conceived of the study, participated in the study design, interpretation of study results and provided feedback on the manuscript. HW, RN and RB participated in the study design, acquisition of data, and interpretation of study results and provided feedback on the manuscript. CF, KC, MA and KW contributed to the study design, interpretation of study results and provided feedback on the manuscript. All authors read and approved the final manuscript.

Word count: 3,670

Keywords: Homelessness; validation; health services research

# Abstract

**Objectives:** To validate case ascertainment algorithms for identifying individuals experiencing homelessness in health administrative databases between 2007 and 2014; and to estimate homelessness prevalence trends in Ontario, Canada between 2007 and 2016.

**Design:** A population-based retrospective validation study

**Setting:** Ontario, Canada, from 2007 to 2014 (validation) and 2007 to 2016 (estimation).

**Participants**: Our reference standard was the known housing status of a longitudinal cohort of housed (n=137,200) and homeless or vulnerably housed (n=686) individuals. Two reference standard definitions of homelessness were adopted: the housing episode and the annual housing experience (any homelessness within a calendar year).

**Main outcome measures**: Sensitivity, specificity, positive and negative predictive value and positive likelihood ratios of 30 case ascertainment algorithms for detecting homelessness using up to eight health services databases.

**Results**: Sensitivity estimates ranged from 10.8% to 28.9% (housing episode definition) and 18.5 to 35.6% (annual housing experience definition). Specificities exceeded 99% and positive likelihood ratios were high using both definitions. The most optimal algorithm estimates that 59,974 (95% CI: 55,231 to 65,208) Ontarians (0.53% of the adult population) experienced homelessness in 2016, a 67.3% increase from 2007.

**Conclusions**: In Ontario, case ascertainment algorithms for identifying homelessness had low sensitivity but very high specificity and positive likelihood ratio. The use of health administrative databases may offer opportunities to track individuals experiencing homelessness over time and inform efforts to improve housing and health status in this vulnerable population.

# Article Summary

Strengths and limitations

- This study validated health administrative codes used in Canadian health databases against a longitudinally collected, representative sample of individuals with known housing status;
- Health administrative data for certain subgroups without Ontario health coverage (e.g. First Nations on reserves, individuals newly arrived to Ontario) was unavailable;
- Our general population sample was assumed housed for the entirety of their observation period. It is possible despite our screening efforts that certain individuals experienced homelessness episodes during their participation in this study.

## Introduction

Individuals experiencing homelessness commonly face physical and mental health challenges, increased morbidity, mortality and health care usage (1, 2). However, surveillance of this population has proven challenging (3-8), with most efforts to date primarily focused on enumerating homeless people at a given point in time (8-9). In Canada, the most recent such effort estimates 235,000 individuals, or 0.67% of the population, experienced homelessness in 2016 (10). While such ecological measures are of some value for service planning, they have been criticized as inaccurate and unrepresentative. Cross-sectional counts taken at select dates may not reflect the homeless population year-round (3-5, 8), are likely to miss certain types of vulnerably housed individuals (for instance, those temporarily or transitionally housed)(3-5, 8), and are resource and time consuming (11-12). Further, these measures do not permit follow up over time or the evaluation of targeted strategies (13, 14), including Canada's recently announced National Housing Strategy (15).

In the absence of concerted surveillance, nations like Canada that provide government-funded universal health care may offer an alternate avenue to measure and track individuals experiencing homelessness. In particular, several administrative databases such as those for hospital services are standardized nation-wide, allowing for population-level tracking of health and health care delivery of Canadians (16). Health administrative data are already widely used in Canada for population surveillance of health conditions such as diabetes, asthma and ischemic heart disease (17-21), permitting counts of the population at any point in time as well as tracking changes in group demographics, health status, health care trajectories and gaps in care (22-24). Currently, the utility of these data in tracking social determinants of health, such as homelessness, are less well understood. Moreover, although health administrative data provide a convenient and low cost option for population surveillance, they are prone to errors in misclassification (25). Validation studies are thus necessary to evaluate the accuracy of case ascertainment algorithms (26-28).

The aims of this study were to (a) develop and validate case ascertainment algorithms to identify individuals experiencing homelessness in health administrative databases in Ontario, Canada; and (b) estimate annual population-prevalence of homelessness in Ontario over a 10-year period using the best performing algorithm.

# Methods

Study design and participants

We validated 30 case ascertainment algorithms to detect homelessness using up to eight health administrative databases in Ontario, Canada's most populous province. All databases were linked using unique encoded identifiers and analyzed at ICES (formerly known as the Institute for Clinical Evaluative Sciences) (29), a not-for-profit research institute. ICES is a prescribed entity under section 45 of Ontario's Personal Health Information Protection Act, which authorizes ICES to collect personal health information, without consent, for the purpose of analysis or compiling statistical information with respect to the management of, evaluation or monitoring of, the allocation of resources to or planning for all or part of the health system. This study was approved by the St Michael's Hospital Research Ethics Board, and follows STARD guidelines for reporting diagnostic accuracy studies.

Patient and public involvement

Due to the coded nature of ICES data, this research was conducted without patient involvement. Patients were not involved in the development of the research question, invited to comment on the study design, consulted to interpret the results, and were not invited to contribute to the writing or editing of this document for readability or accuracy.

Data availability

While data sharing agreements prohibit ICES from making the dataset publicly available, access to the data may be granted to those who meet pre-specified criteria for confidential access, available at www.ices.on.ca/DAS. The full dataset creation plan and underlying analytic code detailing all analysis procedures are available from the authors upon request, understanding that computer programs rely upon coding templates or macros unique to ICES, which may be either inaccessible or require modification.

Participants

Our validation cohort included adults (18 years or older) eligible for Ontario health coverage who participated in the Health and Housing in Transition study (the "HHiT sample")(30). The HHiT study was conducted between 2009 and 2014 in three Canadian cities (Toronto, Ottawa and Vancouver) and aimed to assess the impact of housing transitions on health. Participants were randomly selected at shelters, meal programmes, community health centres, drop-in centres, rooming houses, and single-room occupancy hotels and interviewed once per year until the end of the study or until the individual withdrew.. Collected participant data from the two Ontario cities (Toronto and Ottawa) were organized into consecutive self-reported housing episodes, ranging from an earliest date of January 31, 2007 to a latest date of March 14, 2014.

Due to the low prevalence (<5%) of exclusively housed individuals in this cohort, an additional group of adults presumed housed (the "general population sample") was randomly selected from the ICES Registered Persons Database (RPDB), which includes all individuals eligible for Ontario health coverage. A similar approach was used in previous validation studies (31, 32). To ensure our general population sample had a high likelihood of being housed, we deemed individuals eligible if they were not part of the HHiT study, resided in Toronto or Ottawa throughout the study period and did not reside in a postal code associated with shelter services. We randomly selected 200 individuals for each HHiT participant to approximate the nearest available Canadian homelessness prevalence estimate (33).

Reference standard

The period over which housing status is assessed substantially impacts any analysis of agreement between the reference standard and case ascertainment algorithms. Thus, we *a priori* selected two reference standard definitions (units of analysis) based on their expected utility: a) the housing episode and b) the annual housing experience. Within the HHiT cohort, housing episodes were categorized as *housed* or *homeless* based on pre-established criteria. (34) Responses about housing status were classified into one of 25 categories, and then resolved into housed, institution and homeless categories. "Institution" episodes (which include situations like hospitalization or prison) were then resolved into either housed or homeless categories based on the preceding and subsequent housing episodes: episodes flanked by any homelessness were generally also classified as homeless, as the individual was not stably housed either at the time of entry or exit (or both) from the institution. The general population sample was assumed housed for the entirety of their observation period. For the annual housing experience definition, individuals were categorized as homeless if a homeless episode occurred during the calendar year.

Case Ascertainment Algorithms and Data Sources

Homeless indicators were identified by searching the ICES data dictionary (35) for data elements indicative of housing status (search terms included: 'homeless', 'shelter', 'housing', 'residence', 'transient')(Supplement Table 1). We assessed housing status indicators present in: the Discharge Abstract Database (DAD), the National Ambulatory Care Reporting System emergency (NACRS), the Ontario Mental Health Reporting System (OMHRS), the Home Care Database (HCD), the Resident Assessment Instrument Contact Assessment Database (RAICA), the National Rehabilitation Reporting System (NRS) and the Canadian Organ Replacement Registry (CORR). The first three sources report hospital encounters and are tracked by the Canadian Institute for Health Information (CIHI)(13); for brevity these are hereafter referred to as "CIHI databases".

Postal codes are also often recorded in the above records; therefore, we additionally assessed postal codes where present and in the ICES PSTLYEAR database (which provides a yearly postal code for individuals with Ontario health coverage) against Toronto and Ottawa-based postal codes identifying shelter services or hospitals (which are sometimes erroneously coded instead of shelters)(36). Postal codes which included residential addresses, as determined through a Geographic Information System, were not used to avoid misclassifying housed individuals as homeless.

We tested 30 case ascertainment algorithms (described in Supplement Table 2) which varied by: 1) databases included (all vs. CIHI only); 2) inclusion or exclusion of postal code indicators (none, in health service databases or in PSTLYEAR) and 3) extension of time intervals (ranging 0 days to ±180 days) before and after the reference period. The practice of extending time intervals is known to enhance the sensitivity of case ascertainment algorithms (37, 38). Reference housing episodes or calendar years without overlapping health care encounters were coded as test negative ("housed") by default, to reflect the administrative data's inability to identify homelessness for such reference periods.

Other data sources used to describe the cohort (all data sources are further described in Supplement Table 3) included the ICES RPDB, Ontario Health Insurance Physicians (OHIP) claims database, the Immigration, Refugee and Citizenship Canada (IRCC) Permanent Residents database, and several ICES-derived population-surveillance datasets including: the Chronic Obstructive Pulmonary Disease (COPD)(39), Ontario Diabetes Dataset (ODD)(40), Congestive Heart Failure (CHF)(41) and Ontario HIV (42) derived cohorts.

Statistical analysis

We provided cohort demographics, comorbidities and recent health services usage (variables defined in Supplement Table 4). Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and positive likelihood ratio (LR+) were calculated for all algorithms (formulae in Supplement Table 5). Confidence intervals (95% CIs) were calculated using the Wilson score method (43). For each reference standard, we deemed the algorithm with maximized sensitivity, specificity and PPV to be optimal, while also considering its scalability (i.e. applicability of the algorithm outside Ontario).

We then applied the optimal annual housing experience algorithm to identify Ontarians experiencing homelessness in each of the 2007 to 2016 calendar years, further describing those identified during 2016. Finally, we estimated population-prevalence of homelessness between 2007 and 2016, correcting for sensitivity by dividing the number of identified homeless by the algorithm's sensitivity. Prevalence rates were calculated by dividing estimated population prevalence by the total adult Ontario population for each year. A Poisson regression model was used to estimate the annual change in prevalence over time.

All analyses were conducted using SAS, version 9.4 (44).

## Results

Cohort

We identified 686 eligible HHiT participants (6,948 housing episodes, 3,443 of which were homeless) and randomly selected a further 137,200 individuals from the RPDB (137,200 housing episodes) to generate a total cohort of 137,886 individuals contributing 144,148 housing episodes (Figure 1). HHiT participants were followed for, on average, 64 months, and experienced homelessness for, on average, 40.4% of their overall participation period, with a median homeless episode of 75 days (Interquartile range [IQR]: 29 to 181 days)(Table 1). Individuals in the general population sample were followed for an average of 52 months. We found substantial differences between the HHiT and general population samples, with HHiT participants being younger, more likely male, less likely to have recently immigrated and having more chronic health conditions and recent health care use.

Validation Results

Algorithm sensitivities when identifying a homeless housing episode (among 144,148 total episodes) ranged from 10.8% to 28.9%, with specificities exceeding 99% (Table 2). Extending time intervals or including postal code indicators in health services databases increased sensitivity, while marginally decreasing specificity. The use of all databases, as opposed to only CIHI databases, resulted in negligible gains in sensitivity. Positive likelihood ratios were all in excess of 10, indicating a substantial increase in the likelihood of homelessness following a positive test (45). Based on these findings, we chose *any CIHI database indicator +/- 45 days* as the optimal algorithm based on its scalability and maximized sensitivity, specificity and positive predictive values. More false-positives (n=595) using this algorithm came from the HHiT sample (n=397, or 66.7% of false positives) than the general population sample (n=238) (Supplement Table 6A). Absence of a health care encounter during the reference period accounted for 64.5% (n= 1,825) of false negatives.

Algorithm sensitivities when identifying homeless annual housing experiences (n=491,213 total calendar years) ranged from 18.5% to 35.6%, with specificities at 99.9% (Table 2). Positive likelihood ratios were all in excess of 200, indicating a very substantial increase in the probability of homelessness following a positive test (45). Sensitivity increased without impacting specificity when time windows were extended or when postal code indicators during health care encounters or in PSTLYEAR were included. The use of all databases, as opposed to solely CIHI databases, resulted in negligible gains in sensitivity.

The algorithm that maximized validation statistics was *any CIHI database indicator +/- 15 days or a PSTLYEAR postal code*. Most false-positives (n=365) using this algorithm were sourced from the general population sample (n=250; 68.5% of false positives overall)(Supplement Table 6B). Absence of a health encounter within the reference period accounted for 62.7% (or 997) of false negatives. However, because this algorithm requires a comprehensive database of postal codes uniquely identifying shelters or hospitals to be scaled, we deemed this algorithm suboptimal and therefore opted to use *any CIHI database indicator +/- 15 days* for generating provincial estimates.

Estimates of homelessness

Applying the optimal annual housing experience algorithm, we identified 11,731 Ontarians experiencing homelessness during 2016 (Table 3). Flagged individuals were predominantly male (70%) and between the ages of 25 to 65. One in ten were recent immigrants, about one third resided in Metropolitan Toronto, and a large proportion recently received mental or substance use-related health care (25.7% for psychotic disorders; 54.8% for non-psychotic disorders and 41.9% for substance use disorders). Over 10 years, we identified a total of 54,873 adults who experienced homelessness, of which 18,217 (33.2%) were detected in more than one year (Supplement Table 6C).

As specificity for our chosen algorithm is near 100%, we corrected for sensitivity by dividing our identified cohort count by sensitivity to estimate a total 2016 homeless population of 59,974 (95% CI: 55,231 to 65,208) Ontarians, or 0.53% of the adult Ontario population (Figure 2). Between 2007 and 2016, the number and rate of individuals experiencing homelessness increased by 67.3% and 48.1%, respectively, with an annual percentage increase of 4.4% in the estimated rate of homelessness (95% CI: 4.2% - 4.7%).

## Discussion

We validated health administrative database algorithms for homelessness against the known housing status of individuals in a longitudinally collected, representative sample at risk for homelessness and a random sample of housed individuals in Ontario, Canada. We tested our algorithms' ability to identify individuals during an experience of homelessness and during a year in which homelessness occurred, as either definition could be used for different purposes (research and surveillance, respectively). In both cases, algorithms exhibited low sensitivity but excellent specificities and positive likelihood ratios.

The low sensitivity of the algorithms can be partially explained by the large proportion of reference periods without a health care encounter, which increased false-negatives by default. This reaffirms the consensus that homelessness is ephemeral for many individuals, making it difficult to capture in health administrative data (1, 3, 5). Although homeless individuals are known to access acute care services at a much higher rate than the general population (1, 2), a substantial subgroup in our homeless cohort did not access hospital-based health care services during specific housing periods, and therefore could not be identified as so using the algorithms. We observed that homeless individuals more frequently accessed care through outpatient physician clinics, which are captured through fee-for-service billings. This data holding (the Ontario Health Insurance Plan), currently lacks housing status information and therefore could not be included in our validation.

Our population prevalence estimates suggest substantial increases in homelessness between 2007 and 2016, both in absolute and relative terms. Case sensitivity did not noticeably change over time in our validation cohort (less than a 4% variation throughout, with no trend), but we cannot know for certain whether case sensitivity increased across Ontario during this period, partially or fully accounting for the observed increase. However, a recent presentation by Employment and Social Development Canada indicates that, among Canadian communities who conducted point in time counts in 2016 and 2018, homelessness increased 14% (46); the estimates generated by the 2013 and 2016 *State of Homelessness in Canada* reports indicate similar increases (10, 33). These results suggest that our observed increase may reflect a true increase in the prevalence of homelessness in Ontario.

No Ontario-specific statistics exist against which to directly compare our most recent population prevalence estimate (47); however, if we assume Canadian homelessness as recently reported (10) is proportionally distributed among the 13 Canadian provinces and territories population (Ontario accounted for 38.3% of Canada's population in 2016)(48), approximately 90,000 homeless individuals would be attributable to Ontario in 2016. This prevalence estimate is greater than the 2016 estimate concluded in this study (of approximately 60,000), but individuals identified as homeless in our algorithm share similar demographics with individuals in that report: approximately 25% in both sources are ages 50 and older; 16-19% are youth; and roughly 30% are women (10). Furthermore, one in three individuals were identified in multiple years, similar to the proportion of individuals using shelters in multiple years reported recently (49). Therefore, the gap between methodologies does not appear to reflect a bias in the types of individuals identified in these two sources.

This is the first study to validate health administrative data algorithms against a reference standard with the intended purpose of population-surveillance. Most prior work (50-57) identified homelessness using homeless indicators or shelter addresses given during health care encounters, assuming these data represented true housing status. Recently, Vickery et al. validated addresses indicative of homelessness during health care encounters against self-reported housing status in a sample of Medicaid recipients, finding sensitivities between 30% and 76% and specificities between 79% and 97% (58). However, this study required the use of location- and time- specific shelter address registries, making the methodology challenging to scale or generalize. Moreover, this study's results refer to the population using health care (rather than the population overall) and assumed self-reported housing status did not vary over the nearly four year study period. Our study recognized changes in housing status and deliberately included individuals who may not have used health care, in order to estimate the algorithm's ability to count the complete homeless population.

We readily acknowledge some limitations to this validation. First, because it was conducted in a universal, single payer health care system, this validation's applicability is limited to jurisdictions with similar settings who collect similar types of standardized information. Even so, before implementation policy makers should undertake a validation similar to that described here to determine how data sources available to them perform. However, among such jurisdictions this methodology can permit inexpensive, population-level research and surveillance.

Second, as this validation relied on health administrative data with housing indicators, algorithm sensitivity was significantly reduced due to the number of individuals who did not access hospital-based health care services during their housing period and were thus automatically considered housed. Other jurisdictions having access to housing status variables in standardized health services data and the ability to link non-health administrative data containing housing variables (such as in social services, law enforcement, or shelter service data) may realize improved algorithm performance through increased opportunities for encounters during a homeless episode.

Third, we could only validate homelessness among adults eligible for Ontario health care coverage, which although near-complete (>99%) does not include recent arrivals to Ontario, First Nations on reserves, Inuit, certain refugee claimant groups, inmates in federal penitentiaries, eligible veterans and serving members of the Canadian Forces. Since veterans and First Nations, Metis and Inuit individuals are believed to be over-represented among homeless people (10, 33, 49, 59), our algorithms almost certainly underestimate homelessness in these populations, which (in conjunction with the lack of youth in the count) may account for much of the gap between our population estimate and the estimate loosely calculated from the *State of Homelessness in Canada 2016* (10). However, this gap is the result of linkage through Ontario-specific identifiers rather than an inherent limitation of the indicators: future pan-Canadian homelessness surveillance and research can include these populations by accessing these indicators through CIHI.

Fourth, we were forced to assume our general population sample was housed during the entirety of their assigned housing period. It is possible, despite our screening efforts, that some individuals experienced homelessness during their participation in this study. Upon review of the false positives, we identified 238 individuals from the general population sample (0.17% of that sample) who might have thus been misclassified as housed when they were, in fact, homeless. We deemed misclassifying up to a few hundred individuals from a pool of over 140,000 to be preferable to excluding or re-coding such individuals on the basis of the same administrative data we are attempting the validate. Moreover, given the low prevalence of homelessness in the population, the impact of such individuals should be negligible to our overall findings.

Despite the recent Canadian federal government commitment of $2.2 billion over 10 years to tackle homelessness (60), current costs associated with enumeration (11-12) and program evaluation are high, necessarily reducing funding for program implementation. Overall, our algorithms present, despite their low sensitivity, important potential cost-savings opportunities as a homelessness enumeration and surveillance tool. Moreover, these algorithms can track individuals over time and be used to evaluate efforts to improve housing and health status, similar to applications from other previous validation work for population surveillance (20-25). Introduction of mandatory reporting of homelessness among hospital and non-hospital based health care encounters may result in increased identification of homelessness in Ontario.

## Acknowledgements

This study was conducted at the ICES Western Site, which receives core funding from the Academic Medical Organization of Southwestern Ontario (AMOSO), the Schulich School of Medicine and Dentistry, Western University, and the Lawson Health Research Institute. ICES is funded by an annual grant from the Ontario Ministry of Health and Long-Term Care. This project was funded by the Academic Health Science Centre Alternate Funding Innovation Fund, the Ontario Trillium Foundation – Local Poverty Reduction Fund, and two Canadian Institutes of Health Research grants. The opinions, results, and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by ICES or the funding agencies is intended or should be inferred. Parts of this material are based on data and/or information compiled and provided by CIHI. However, the analyses, conclusions, opinions and statements expressed in the material are those of the author(s), and not necessarily those of CIHI. Information in the Immigration, Refugees and Citizenship Canada Permanent Resident Dataset were provided by Immigration, Refugees and Citizenship Canada. We gratefully acknowledge the HHiT team for the use of their data, and Emily Kovacs for compiling postal codes identifying shelter services in Toronto and Ottawa.

# References


1. Fazel S, Geddes JR, Kushel M. The health of homeless people in high-income countries: descriptive epidemiology, health consequences, and clinical and policy recommendations. The Lancet. 2014 Oct 25;384(9953):1529-40.

2. Hwang SW, Chambers C, Chiu S, Katic M, Kiss A, Redelmeier DA, Levinson W. A comprehensive assessment of health care utilization among homeless adults under a system of universal health insurance. American Journal of Public Health. 2013 Dec;103(S2):S294-301.

3. Fitzgerald ST, Shelley MC, Dail PW. Research on homelessness: Sources and implications of uncertainty. American Behavioral Scientist. 2001 Sep;45(1):121-48.

4. Stefancic A, Schaefer-McDaniel NJ, Davis AC, Tsemberis S. Maximizing follow-up of adults with histories of homelessness and psychiatric disabilities. Evaluation and Program Planning. 2004 Nov 1;27(4):433-42.

5. Booth S. Researching health and homelessness: methodological challenges for researchers working with a vulnerable, hard to reach, transient population. Australian Journal of Primary Health. 1999;5(3):76-81.

6. Laflamme F. Survey of homeless in Canada: Street component feasibility study. Ottawa, ON: Statistics Canada Statistical Consultation Group; 2001.

7. Canadian Definition of Homelessness. Canadian Observatory on Homelessness. www.homelesshub.ca/CHRNhomelessdefinition. Published 2012. Accessed January 2, 2019.

8. Williams, M. Can we measure homelessness?A critical evaluation of 'Capture–Recapture'. Methodological Innovations Online. 2010. Aug 5(2): 49-59

9. Employment and Social Development Canada. Everyone counts: A Guide to Point-in-Time Counts in Canada – 2nd Edition. http://www12.esdc.gc.ca/sgpe-pmps/servlet/sgpp-pmpspub?lang=eng&curjsp=p.5bd.2t.1.3ls@eng.jsp&curactn=dwnld&pid=55638&did=4956. Published 2017. Accessed January 3, 2019.

10. Gaetz S, Dej E, Richter T, Redman M. The State of Homelessness in Canada 2016. Toronto: Canadian Observatory on Homelessness Press.

11. City of Toronto. Backgrounder: Results of the 2018 Street Needs Assessment. https://www.toronto.ca/home/media-room/backgrounders-other-resources/backgrounder-results-of-the-2018-street-needs-assessment/ Published 2018. Accessed December 3, 2018.

12. Rabinovitch, H. Snapshot of an Object in Motion: Quantifying Homelessness. Simon Fraser University. 2005. http://summit.sfu.ca/item/15297. Accessed July 10 2019.
13. Burnes D, DiLeo D, eds. Ending homelessness: why we haven't, how we can. Boulder, Colorado: Lynne Rienner; 2016.

14. Gabbard WJ, Snyder CS, Lin MB, Chadha JH, May JD, Jaggers J. Methodological issues in enumerating homeless individuals. Journal of Social Distress and the Homeless. 2007 May 1;16(2):90-103.

15. Canada's National Housing Strategy. https://www.placetocallhome.ca/pdfs/Canada-National-Housing-Strategy.pdf. Updated January 30 2018. Accessed July 2 2019.

16. CIHI. Canadian Institute for Health Information. https://www.cihi.ca/en. Updated 2018. Accessed January 3, 2019.

17. Public Health Agency of Canada. *The Canadian Chronic Disease Surveillance System – An Overview*. https://www.canada.ca/en/public-health/services/publications/canadian-chronic-disease-surveillance-system-factsheet.html. Published 2018. Accessed January 3, 2019.

18. Government of Canada. *Publications – Diseases and Conditions.* https://www.canada.ca/en/services/health/publications/diseases-conditions.html. Published 2018. Accessed January 3, 2019.

19. Lix L, Ayles J, Bartholomew S, et al. The Canadian chronic disease surveillance system: A model for collaborative surveillance. International Journal of Population Data Science. 2018;*3*(3). doi: https://doi.org/10.23889/ijpds.v3i3.433

20. El Emam K, Samet S, Hu J, et al. A protocol for the secure linking of registries for HPV surveillance. PLoS One. 2012;7(7):e39915.

21. Wilson K, Hawken S, Kwong JC, Deeks SL, Crowcroft NS, Manuel D. Vaccine and immunization surveillance in Ontario (VISION): Using linked health administrative databases to monitor vaccine safety. Vaccine. 2018;30(43):6115-6120.

22. Greiver M, Williamson T, Barber D, et al. Prevalence and epidemiology of diabetes in Canadian primary care practices: a report from the Canadian Primary Care Sentinel Surveillance Network. Canadian journal of diabetes. 2014 Jun 1;38(3):179-85.

23. Allison F, Lix Lisa M, Kim R. Estimating multimorbidity prevalence with the Canadian chronic disease surveillance system. Health promotion and chronic disease prevention in Canada: research, policy and practice. 2017 Jul;37(7):215.

24. Ospina MB, Voaklander D, Senthilselvan A et al. Incidence and prevalence of chronic obstructive pulmonary disease among aboriginal peoples in Alberta, Canada. PLoS One. 2015 Apr 13;10(4):e0123204.

25. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. Journal of clinical epidemiology. 2011 Aug 1;64(8):821-9.

26. Carrara G, Scirè CA, Zambon A, Cimmino MA, Cerra C, Caprioli M, et al. A validation study of a new classification algorithm to identify rheumatoid arthritis using administrative health databases: Case–

control and cohort diagnostic accuracy studies. Results from the RECord linkage On Rheumatic Diseases study of the Italian Society for Rheumatology. BMJ open. 2015 Jan 1;5(1):e006029.

27. Butt DA, Tu K, Young J, Green D, Wang M, Ivers N, et al. A validation study of administrative data algorithms to identify patients with Parkinsonism with prevalence and incidence trends. Neuroepidemiology. 2014;43(1):28-37.

28. Widdifield J, Bernatsky S, Paterson JM, Tu K, Ng R, Thorne JC, et al. Accuracy of Canadian health administrative databases in identifying patients with rheumatoid arthritis: a validation study using the medical records of rheumatologists. Arthritis care & research. 2013 Oct 1;65(10):1582-91.

29. ICES. www.ices.on.ca. Updated 2019. Accessed January 2, 2019.

30. Hwang SW, Aubry T, Palepu A, Farrell S, Nisenbaum R, Hubley AM, et al. The health and housing in transition study: a longitudinal study of the health of homeless and vulnerably housed adults in three Canadian cities. International journal of public health. 2011 Dec 1;56(6):609-23.

31. Chan J, Mack DR, Manuel DG, Mojaverian N, de Nanassy J, Benchimol EI. Validation of an algorithm to identify children with biopsy-proven celiac disease from within health administrative data: An assessment of health services utilization patterns in Ontario, Canada. PloS one. 2017 Jun 29;12(6):e0180338.

32. Ducharme R, Benchimol EI, Deeks SL, Hawken S, Fergusson DA, Wilson K. Validation of diagnostic codes for intussusception and quantification of childhood intussusception incidence in Ontario, Canada: a population-based study. The Journal of pediatrics. 2013 Oct 1;163(4):1073-9.

33. Gaetz S, Donaldson J, Richter T, Gulliver T. The State of Homelessness in Canada 2013. Toronto: Canadian Homelessness Research Network Press.

34. Harris M, Gadermann A, Norena M, et al. Residential moves and its association with substance use, healthcare needs, and acute care use among homeless and vulnerably housed persons in Canada. International Journal of Public Health. 2018;1-11. doi: 10.1007/s00038-018-1167-6.

35. ICES. Data Dictionary. https://datadictionary.ices.on.ca/Applications/DataDictionary/Default.aspx. Updated 2019. Accessed January 3, 2019.

36. Canadian Institute for Health Information. Recording Postal Codes in the DAD and NACRS. https://www.cihi.ca/en/recording-postal-codes-in-dad-and-nacrs_en.pdf. Published N.D. Accessed December 27, 2018.

37. Robinson JR, Young TK, Roos LL, Gelskey DE. Estimating the burden of disease: comparing administrative data and self-reports. Medical care. 1997 Sep 1:932-47.

38. Hoogenveen R, Westert G, Dijkgraaf M, Schellevis F, de Bakker D. Disease prevalence estimations based on contact registrations in general practice. Statistics in medicine. 2002 Aug 15;21(15):2271-85.

39. Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying individuals with physician diagnosed COPD in health administrative databases. COPD. 2009;6(5):388-394.

40. Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: Determination of prevalence and incidence of diabetes using a validated administrative data algorithm. Diabetes Care. 2002;25(3):512-516.

41. Schultz SE, Rothwell DM, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: A validation study using primary care patient records. Chronic Dis Inj Can. 2013;33(3):160-166.

42. Antoniou T, Zagorski B, Loutfy MR, Strike C, Glazier RH. Validation of case-finding algorithms derived from administrative data for identifying adults living with human immunodeficiency virus infection. Plos One. 2011;6(6):e21748.

43. Wilson EB. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association. 1927;22(158):209-212.

44. SAS Enterprise Guide. V9.4. Cary, NC: SAS Institute Inc.

45. McGee, S. Simplifying likelihood ratios. Journal of general internal medicine. 2002; 17(8): 647-650.

46. Duchesne, A. & Cooper, I. National Homelessness Data National Shelter Data 2005-2016 Point-in-Time Count 2018. http://datathatmakesadifference.com/wp-content/uploads/2019/06/Annie-Duchesne-Ian-Cooper.pdf Published May 28 2019. Accessed July 10 2019.

47. Ministry of Municipal Affairs and Housing. A place to call home: report of the expert advisory panel on homelessness. 2015. http://www.mah.gov.on.ca/AssetFactory.aspx?did=11038 Accessed July 15 2019.

48. Statistics Canada. Ontario [Province] and Canada [Country] (table) Census Profile. 2016 Census. Statistics Canada Catalogue no. 98-316-X2016001. https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E. Released November 29, 2017.

49. Gaetz S, Gulliver T, Richter T. The state of homelessness in Canada 2014. Toronto: Canadian Observatory on Homelessness Press.

50. Peterson R, Gundlapalli AV, Metraux S, et al. Identifying homelessness among veterans using VA administrative data: Opportunities to expand detection criteria. PloS one. 2015;10(7):e0132664.

51. Tsai J, Doran KM, Rosenheck RA. When health insurance is not a factor: National comparison of homeless and nonhomeless US veterans who use Veterans Affairs Emergency Departments. Am J Public Health. 2013;103(S2):S225-S231.

52. Bharel M, Lin WC, Zhang J, O'Connell E, Taube R, Clark RE. Health care utilization patterns of homeless individuals in Boston: Preparing for Medicaid expansion under the Affordable Care Act. Am J Public Health. 2013;103(S2):S311-S317.

53. Zech J, Husk G, Moore T, Kuperman GJ, Shapiro JS. Identifying homelessness using health information exchange data. Journal of the American Medical Informatics Association. 2015;22(3): 682-687.

54. Romero-Ortuño R, O'Riordan D, Silke B. Profiling the medical admissions of the homeless. Acute Med. 2012;11:197-204.

55. Gundlapalli AV, Carter ME, Palmer M et al. Using Natural Language Processing on the Free Text of Clinical Documents to Screen for Evidence of Homelessness Among US Veterans. AMIA Annual Symposium Proceedings. 2013: 537–546.

56. Erickson J, Abbott K, Susienka L. Automatic address validation and health record review to identify homeless Social Security disability applicants. Journal of biomedical informatics. 2018;2: 41-46.

57. Government of Alberta. Total Inpatient Costs by Homeless Status. Health Trends Alberta. 2018.

58. Vickery KD, Shippee ND, Bodurtha P, et al. Identifying homeless Medicaid enrollees using enrollment addresses. Health services research. 2018;53(3):1992-2004.

59. Patrick C. Aboriginal Homelessness in Canada: A Literature Review. http://homelesshub.ca/sites/default/files/AboriginalLiteratureReview.pdf. Published 2014. Accessed January 3, 2019.

60. Government of Canada. Reaching Home: Canada’s Homelessness Strategy. https://www.canada.ca/en/employment-social-development/news/2018/11/reaching-home-canadas-homelessness-strategy.html. Published 2018. Accessed January 3, 2019.

Table 1 – Cohort characteristics at the start of a randomly selected housing episode, by source

| Characteristic | Validation Participants (N=137,886) | HHiT sample Study (N=686) | General population sample (N=137,200) | P-value |
|---|---|---|---|---|
| Mean % (SD) of period spent homeless | n/a | 40.4% (29.4%) | n/a | n/a |
| Median days (IQR) of homelessness episode | n/a | 75 (29 – 181) | n/a | n/a |
| Age, mean (SD) | 46.1 (18.0) | 43.5 (10.6) | 46.1 (18.0) | < 0.001 |
| Female, n (%) | 70,535 (51.2%) | 208 (30.3%) | 70,327 (51.3%) | < 0.001 |
| Located in Ottawa, n (%) | 104,059 (75.5%) | 357 (52%) | 103,702 (75.6%) | < 0.001 |
| Located in Toronto, n (%) | 33,827 (24.5%) | 329 (48%) | 33,498 (24.4%) | < 0.001 |
| Recent immigrant, n (%) | 32,657 (23.7%) | 45 (6.6%) | 32,612 (23.8%) | < 0.001 |
| Date of immigration, n (%) | | | | |
| < 1 year | 1,152 (0.8%) | <=5 | NR | < 0.001 |
| 1 to 3 years | 2,381 (1.7%) | <=5 | NR | |
| 4-10 years | 9,606 (7.0%) | 9 (1.3%) | 9,597 (7.0%) | |
| Over 10 years | 19,518 (14.2%) | 33 (4.8%) | 19,485 (14.2%) | |
| Refugee status, n (%) | 5,907 (4.3%) | 18 (2.6%) | 5,889 (4.3%) | < 0.001 |
| Congestive heart failure, n (%) | 2,186 (1.6%) | 6 (0.9%) | 2,180 (1.6%) | 0.14 |
| Chronic obstructive pulmonary disease, n (%) | 6,627 (4.8%) | 91 (13.3%) | 6,536 (4.8%) | < 0.001 |
| Diabetes, n (%) | 11,332 (8.2%) | 67 (9.8%) | 11,265 (8.2%) | 0.14 |
| HIV, n (%) | 402 (0.3%) | 30 (4.4%) | 372 (0.3%) | < 0.001 |
| Chronic kidney disease[1], n (%) | 2,431 (1.8%) | 20 (2.9%) | 2,411 (1.8%) | 0.02 |
| Chronic liver disease[1], n (%) | 2,939 (2.1%) | 87 (12.7%) | 2,852 (2.1%) | < 0.001 |
| Mental health related care[2], n (%) | | | | |
| Psychotic disorders | 928 (0.7%) | 76 (11.1%) | 852 (0.6%) | < 0.001 |
| Non-psychotic disorders | 15,128 (11.0%) | 248 (36.2%) | 14,880 (10.8%) | < 0.001 |
| Substance use disorders | 1,640 (1.2%) | 204 (29.7%) | 1,436 (1.0%) | < 0.001 |
| Charlson comorbidity index, n (%) | | | | |
| 0 | 7,866 (5.7%) | 86 (12.5%) | 7,780 (5.7%) | < 0.001 |
| 1 | 1,589 (1.2%) | 25 (3.6%) | 1,564 (1.1%) | |
| 2+ | 2,476 (1.8%) | 25 (3.6%) | 2,451 (1.8%) | |
| No Hospitalizations | 125,955 (91.3%) | 550 (80.2%) | 125,405 (91.4%) | |
| Primary care visits[2], mean (SD) | 13.0 (17.5) | 21.1 (31.7) | 12.9 (17.4) | < 0.001 |
| Emergency department visits[2], mean (SD) | 1.6 (1.7) | 3.9 (5.1) | 1.6 (1.5) | < 0.001 |
| Hospitalizations[2], mean (SD) | 1.3 (0.9) | 1.7 (1.4) | 1.3 (0.9) | < 0.001 |

1. Within past 3 years; 2. Occurring in the past year. Cells representing <=5 individuals are suppressed to protect participant privacy. Individual immigration status defined based on presence of a landing date in the Immigration, Refugees and Citizenship Canada Permanent Resident Database from 1985 to 2018. NR = Not reportable, due to associated small cell suppression; NS=Not significant; HIV=Human immunodeficiency virus

**Table 2 – Accuracy of case ascertainment algorithms in identifying individuals experiencing homelessness**

| Algorithm Definition | TP | FP | FN | TN | Sensitivity (%) (95% CI) | Specificity (%) (95% CI) | PPV (%) (95% CI) | NPV (%) (95% CI) | LR+ |
|---|---|---|---|---|---|---|---|---|---|
| *Reference Standard Definition: Housing Episode (n = 144,148 overall, with 3,443 homeless episodes)* | | | | | | | | | |
| 1 indicator +/- 0 days | 372 | 528 | 3,071 | 140,177 | 10.8 (9.8 - 11.9) | 99.6 (99.6 - 99.7) | 41.3 (38.2 - 44.6) | 97.9 (97.8 - 97.9) | 28.8 |
| 1 indicator +/- 15 days | 482 | 591 | 2,961 | 140,114 | 14.0 (12.9 - 15.2) | 99.6 (99.5 - 99.6) | 44.9 (42.0 - 47.9) | 97.9 (97.9 – 98.0) | 33.3 |
| 1 indicator +/- 45 days | 619 | 665 | 2,824 | 140,040 | 18.0 (16.7 - 19.3) | 99.5 (99.5 - 99.6) | 48.2 (45.5 - 50.9) | 98.0 (98.0 - 98.1) | 38.0 |
| 1 indicator +/- 90 days | 718 | 765 | 2,725 | 139,940 | 20.9 (19.5 - 22.2) | 99.5 (99.4 - 99.5) | 48.4 (45.9 – 51.0) | 98.1 (98.0 - 98.2) | 38.4 |
| 1 indicator +/- 180 days | 861 | 897 | 2,582 | 139,808 | 25.0 (23.6 - 26.5) | 99.4 (99.3 - 99.4) | 49.0 (46.6 - 51.3) | 98.2 (98.1 - 98.3) | 39.2 |
| 1 indicator OR postal code +/- 0 days | 450 | 679 | 2,993 | 140,026 | 13.1 (12.0 - 14.2) | 99.5 (99.5 - 99.6) | 39.9 (37.0 - 42.7) | 97.9 (97.8 – 98.0) | 27.1 |
| 1 indicator OR postal code +/- 15 days | 572 | 758 | 2,871 | 139,947 | 16.6 (15.4 - 17.9) | 99.5 (99.4 - 99.5) | 43.0 (40.4 - 45.7) | 98.0 (97.9 - 98.1) | 30.8 |
| 1 indicator OR postal code +/- 45 days | 714 | 845 | 2,729 | 139,860 | 20.7 (19.4 - 22.1) | 99.4 (99.4 - 99.4) | 45.8 (43.3 - 48.3) | 98.1 (98.0 - 98.2) | 34.5 |
| 1 indicator OR postal code +/- 90 days | 824 | 967 | 2,619 | 139,738 | 23.9 (22.5 - 25.4) | 99.3 (99.3 - 99.4) | 46.0 (43.7 - 48.3) | 98.2 (98.1 - 98.2) | 34.8 |
| 1 indicator OR postal code +/- 180 days | 994 | 1,135 | 2,449 | 139,570 | 28.9 (27.4 - 30.4) | 99.2 (99.1 - 99.2) | 46.7 (44.6 - 48.8) | 98.3 (98.2 - 98.3) | 35.8 |
| 1 CIHI indicator +/- 0 days | 368 | 466 | 3,075 | 140,239 | 10.7 (9.7 - 11.8) | 99.7 (99.6 - 99.7) | 44.1 (40.8 - 47.5) | 97.9 (97.8 - 97.9) | 36.9 |
| 1 CIHI indicator +/- 15 days | 477 | 528 | 2,966 | 140,177 | 13.9 (12.7 – 15.0) | 99.6 (99.6 - 99.7) | 47.5 (44.4 - 50.6) | 97.9 (97.9 – 98.0) | 39.6 |
| **1 CIHI indicator +/- 45 days** | **613** | **595** | **2,830** | **140,110** | **17.8 (16.6 - 19.1)** | **99.6 (99.5 - 99.6)** | **50.7 (47.9 - 53.6)** | **98.0 (97.9 - 98.1)** | **42.0** |
| 1 CIHI indicator +/- 90 days | 710 | 693 | 2,733 | 140,012 | 20.6 (19.3 – 22.0) | 99.5 (99.5 - 99.5) | 50.6 (48.0 - 53.2) | 98.1 (98.0 - 98.2) | 41.7 |
| 1 CIHI indicator +/- 180 days | 852 | 822 | 2,591 | 139,883 | 24.8 (23.3 - 26.2) | 99.4 (99.4 - 99.5) | 50.9 (48.5 - 53.3) | 98.2 (98.1 - 98.3) | 41.8 |
| 1 CIHI indicator OR postal code +/- 0 days | 444 | 575 | 2999 | 140130 | 12.9 (11.8 - 14.1) | 99.6 (99.6 - 99.6) | 43.6 (40.6 - 46.6) | 97.9 (97.8 – 98.0) | 32.3 |
| 1 CIHI indicator OR postal code +/- 15 days | 566 | 652 | 2877 | 140,053 | 16.4 (15.2 - 17.7) | 99.5 (99.5 - 99.6) | 46.5 (43.7 - 49.3) | 98.0 (97.9 - 98.1) | 36.9 |
| 1 CIHI indicator OR postal code +/- 45 days | 707 | 734 | 2736 | 139,971 | 20.5 (19.2 - 21.9) | 99.5 (99.4 - 99.5) | 49.1 (46.5 - 51.6) | 98.1 (98.0 - 98.2) | 42.1 |
| 1 CIHI indicator OR postal code +/- 90 days | 817 | 852 | 2626 | 139,853 | 23.7 (22.3 - 25.2) | 99.4 (99.4 - 99.4) | 49.0 (46.6 - 51.3) | 98.2 (98.1 - 98.2) | 41.9 |
| 1 CIHI indicator OR postal code +/- 180 days | 985 | 1017 | 2458 | 139,688 | 28.6 (27.1 - 30.1) | 99.3 (99.2 - 99.3) | 49.2 (47.0 - 51.4) | 98.3 (98.2 - 98.3) | 42.4 |

***Reference Standard Definition: Annual Housing Experience (n = 491,213 calendar years overall, with 2,290 homeless years)***

| Algorithm Definition | TP | FP | FN | TN | Sensitivity (%) (95% CI) | Specificity (%) (95% CI) | PPV (%) (95% CI) | NPV (%) (95% CI) | LR+ |
|---|---|---|---|---|---|---|---|---|---|
| 1 indicator +/- 0 days | 429 | 334 | 1,861 | 488,589 | 18.7 (17.2 - 20.4) | 99.9 (99.9 - 99.9) | 56.2 (52.7 - 59.7) | 99.6 (99.6 - 99.6) | 274.2 |
| 1 indicator +/- 15 days | 454 | 352 | 1,836 | 488,571 | 19.8 (18.2 - 21.5) | 99.9 (99.9 - 99.9) | 56.3 (52.9 - 59.7) | 99.6 (99.6 - 99.6) | 275.4 |
| 1 indicator +/- 45 days | 487 | 406 | 1,803 | 488,517 | 21.3 (19.6 – 23.0) | 99.9 (99.9 - 99.9) | 54.5 (51.3 - 57.8) | 99.6 (99.6 - 99.6) | 256.1 |
| 1 indicator +/- 90 days | 529 | 472 | 1,761 | 488,451 | 23.1 (21.4 - 24.9) | 99.9 (99.9 - 99.9) | 52.8 (49.8 - 55.9) | 99.6 (99.6 - 99.7) | 239.3 |
| 1 indicator +/- 180 days | 590 | 588 | 1,700 | 488,335 | 25.8 (24.0 - 27.6) | 99.9 (99.9 - 99.9) | 50.1 (47.2 - 52.9) | 99.7 (99.6 - 99.7) | 214.2 |
| 1 indicator OR postal code +/- 0 days | 512 | 433 | 1,778 | 488,490 | 22.4 (20.7 - 24.1) | 99.9 (99.9 - 99.9) | 54.2 (51.0 - 57.3) | 99.6 (99.6 - 99.7) | 252.5 |
| 1 indicator OR postal code +/- 15 days | 543 | 458 | 1,747 | 488,465 | 23.7 (22.0 - 25.5) | 99.9 (99.9 - 99.9) | 54.2 (51.1 - 57.3) | 99.6 (99.6 - 99.7) | 253.1 |
| 1 indicator OR postal code +/- 45 days | 581 | 525 | 1,709 | 488,398 | 25.4 (23.6 - 27.2) | 99.9 (99.9 - 99.9) | 52.5 (49.6 - 55.5) | 99.7 (99.6 - 99.7) | 236.3 |
| 1 indicator OR postal code +/- 90 days | 629 | 610 | 1,661 | 488,313 | 27.5 (25.7 - 29.3) | 99.9 (99.9 - 99.9) | 50.8 (48.0 - 53.5) | 99.7 (99.6 - 99.7) | 220.2 |
| 1 indicator OR postal code +/- 180 days | 707 | 754 | 1,583 | 488,169 | 30.9 (29.0 - 32.8) | 99.9 (99.8 - 99.9) | 48.4 (45.8 – 51.0) | 99.7 (99.7 - 99.7) | 200.2 |
| 1 indicator +/- 0 days OR PSTLYEAR postal code | 588 | 356 | 1,702 | 488,567 | 25.7 (23.9 - 27.5) | 99.9 (99.9 - 99.9) | 62.3 (59.2 - 65.3) | 99.7 (99.6 - 99.7) | 352.6 |
| 1 indicator +/- 15 days OR PSTLYEAR postal code | 706 | 402 | 1,584 | 488,521 | 30.8 (29.0 - 32.8) | 99.9 (99.9 - 99.9) | 63.7 (60.8 - 66.5) | 99.7 (99.7 - 99.7) | 375.0 |
| 1 indicator +/- 45 days OR PSTLYEAR postal code | 734 | 452 | 1,556 | 488,471 | 32.1 (30.2 – 34.0) | 99.9 (99.9 - 99.9) | 61.9 (59.1 - 64.6) | 99.7 (99.7 - 99.7) | 346.7 |
| 1 indicator +/- 90 days OR PSTLYEAR postal code | 766 | 518 | 1,524 | 488,405 | 33.4 (31.5 - 35.4) | 99.9 (99.9 - 99.9) | 59.7 (56.9 - 62.3) | 99.7 (99.7 - 99.7) | 315.7 |
| 1 indicator +/- 180 days OR PSTLYEAR postal code | 816 | 633 | 1,474 | 488,290 | 35.6 (33.7 - 37.6) | 99.9 (99.9 - 99.9) | 56.3 (53.7 - 58.8) | 99.7 (99.7 - 99.7) | 275.2 |
| 1 CIHI indicator +/- 0 days | 423 | 300 | 1,867 | 488,623 | 18.5 (16.9 - 20.1) | 99.9 (99.9 - 99.9) | 58.5 (54.9 – 62.0) | 99.6 (99.6 - 99.6) | 301.0 |
| **1 CIHI indicator +/- 15 days** | **448** | **315** | **1,842** | **488,608** | **19.6 (18.0 - 21.2)** | **99.9 (99.9 - 99.9)** | **58.7 (55.2 - 62.2)** | **99.6 (99.6 - 99.6)** | **303.6** |
| 1 CIHI indicator +/- 45 days | 480 | 358 | 1,810 | 488,565 | 21.0 (19.3 - 22.7) | 99.9 (99.9 - 99.9) | 57.3 (53.9 - 60.6) | 99.6 (99.6 - 99.6) | 286.3 |
| 1 CIHI indicator +/- 90 days | 521 | 405 | 1,769 | 488,518 | 22.8 (21.1 - 24.5) | 99.9 (99.9 - 99.9) | 56.3 (53.0 - 59.4) | 99.6 (99.6 - 99.7) | 274.7 |
| 1 CIHI indicator +/- 180 days | 581 | 519 | 1,709 | 488,404 | 25.4 (23.6 - 27.2) | 99.9 (99.9 - 99.9) | 52.8 (49.9 - 55.8) | 99.7 (99.6 - 99.7) | 239.0 |
| 1 CIHI indicator OR postal code +/- 0 days | 508 | 370 | 1,782 | 488,553 | 22.2 (20.5 - 23.9) | 99.9 (99.9 - 99.9) | 57.9 (54.6 - 61.1) | 99.6 (99.6 - 99.7) | 293.1 |

| Algorithm Definition | TP | FP | FN | TN | Sensitivity (%) (95% CI) | Specificity (%) (95% CI) | PPV (%) (95% CI) | NPV (%) (95% CI) | LR+ |
|---|---|---|---|---|---|---|---|---|---|
| 1 CIHI indicator OR postal code +/- 15 days | 539 | 390 | 1,751 | 488,533 | 23.5 (21.8 - 25.3) | 99.9 (99.9 - 99.9) | 58.0 (54.8 - 61.2) | 99.6 (99.6 - 99.7) | 295.1 |
| 1 CIHI indicator OR postal code +/- 45 days | 576 | 442 | 1,714 | 488,481 | 25.2 (23.4 – 27.0) | 99.9 (99.9 - 99.9) | 56.6 (53.5 - 59.6) | 99.7 (99.6 - 99.7) | 278.2 |
| 1 CIHI indicator OR postal code +/- 90 days | 622 | 502 | 1,668 | 488,421 | 27.2 (25.4 – 29.0) | 99.9 (99.9 - 99.9) | 55.3 (52.4 - 58.2) | 99.7 (99.6 - 99.7) | 264.5 |
| 1 CIHI indicator OR postal code +/- 180 days | 699 | 634 | 1,591 | 488,289 | 30.5 (28.7 - 32.4) | 99.9 (99.9 - 99.9) | 52.4 (49.8 - 55.1) | 99.7 (99.7 - 99.7) | 235.4 |
| 1 CIHI indicator +/- 0 days OR PSTLYEAR postal code | 583 | 322 | 1,707 | 488,601 | 25.5 (23.7 - 27.3) | 99.9 (99.9 - 99.9) | 64.4 (61.2 - 67.5) | 99.7 (99.6 - 99.7) | 386.6 |
| 1 CIHI indicator +/- 15 days OR PSTLYEAR postal code | 701 | 365 | 1,589 | 488,558 | 30.6 (28.8 - 32.5) | 99.9 (99.9 - 99.9) | 65.8 (62.9 - 68.5) | 99.7 (99.7 - 99.7) | 410.0 |
| 1 CIHI indicator +/- 45 days OR PSTLYEAR postal code | 728 | 404 | 1,562 | 488,519 | 31.8 (29.9 - 33.7) | 99.9 (99.9 - 99.9) | 64.3 (61.5 – 67.0) | 99.7 (99.7 - 99.7) | 384.7 |
| 1 CIHI indicator +/- 90 days OR PSTLYEAR postal code | 760 | 451 | 1,530 | 488,472 | 33.2 (31.3 - 35.1) | 99.9 (99.9 - 99.9) | 62.8 (60.0 - 65.4) | 99.7 (99.7 - 99.7) | 359.8 |
| 1 CIHI indicator +/- 180 days OR PSTLYEAR postal code | 809 | 564 | 1,481 | 488,359 | 35.3 (33.4 - 37.3) | 99.9 (99.9 - 99.9) | 58.9 (56.3 - 61.5) | 99.7 (99.7 - 99.7) | 306.2 |

Bold lines indicate optimal case algorithm definitions. TP = True Positive (flagged as homeless and truly homeless); FP = False Positive (flagged as homeless but not truly homeless); FN = False Negative (flagged as housed but truly homeless); TN = True Negative (flagged as housed and truly housed); PPV = Positive Predictive Value; NPV = Negative Predictive Value; LR+ = Positive Likelihood Ratio; CIHI=Discharge Abstract Database, National Ambulatory Care Reporting System or Ontario Mental Health Reporting System; PSTLYEAR = ICES PSTLYEAR postal code, indicating the best estimate of an individual's postal code for the year using ICES databases.

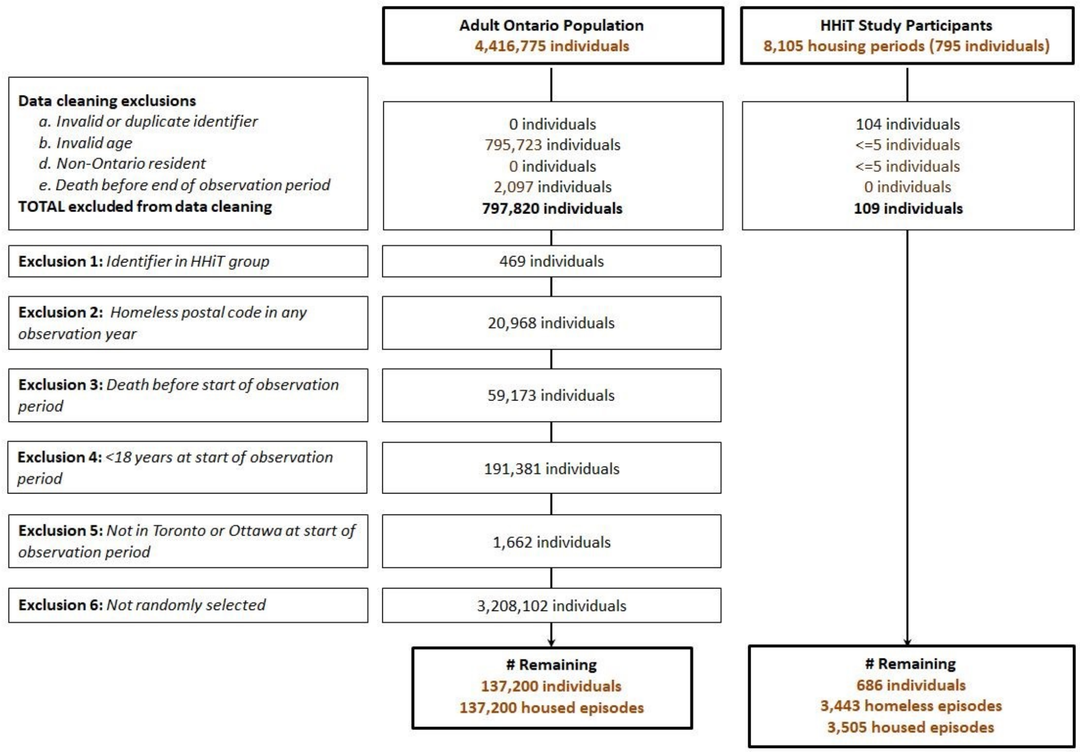
**Table 3 – Characteristics of individuals identified as homeless in 2016 using the optimal annual housing experience algorithm (Any CIHI indicator +/- 15 days)**

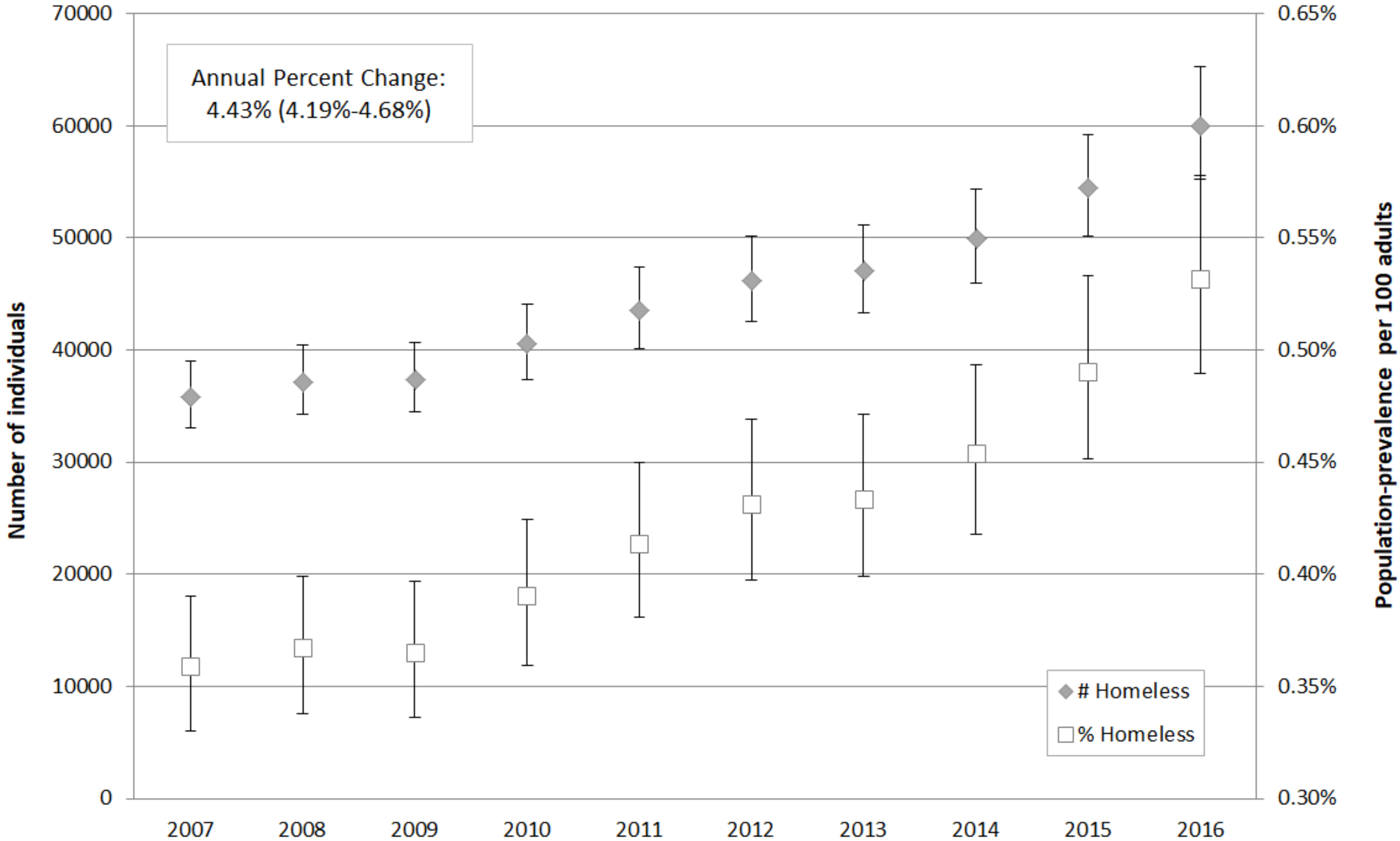
| | Individuals identified as homeless in 2016 (N = 11,731) |
|---|---|
| Age group, in years, N (%) | |
| 18 to 24 | 1,901 (16.2%) |
| 25 to 34 | 3,498 (29.8%) |
| 35 to 50 | 3,246 (27.7%) |
| 51 to 65 | 2,352 (20.1%) |
| Over 65 | 734 (6.3%) |
| Female sex, N (%) | 3,497 (29.8%) |
| City of residence in 2016, N (%) | |
| Toronto | 4,299 (36.7%) |
| Ottawa | 684 (5.8%) |
| In a rural area, N (%) | 667 (5.7%) |
| Recent immigrant, N (%) | 1,172 (10.0%) |
| Immigrated as refugee, N (%) | 366 (3.2%) |
| Charlson comorbidity index, N (%) | |
| 0 | 1,825 (15.6%) |
| 1 | 550 (4.7%) |
| 2+ | 465 (4.0%) |
| No hospitalizations | 8,891 (75.8%) |
| Comorbidities, N (%) | |
| Congestive heart failure | 222 (1.9%) |
| Chronic obstructive pulmonary disease | 1,258 (10.7%) |
| Diabetes | 1,233 (10.5%) |
| Chronic kidney disease[1] | 588 (5.0%) |
| Chronic liver disease[1] | 1,244 (10.6%) |
| HIV positive | 202 (1.7%) |
| Primary care visits[2], mean (SD) | 33.0 (43.6) |
| Emergency department visits[2], mean (SD) | 5.5 (9.2) |
| Admissions to hospital[2], mean (SD) | 1.9 (1.7) |
| Mental health related care[2], N (%) | |
| Psychotic disorders | 3,014 (25.7%) |
| Non-psychotic disorders | 6,433 (54.8%) |
| Substance use disorders | 4,917 (41.9%) |

1. Within the past 3 years; 2. Occurring in the past year; TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative; PPV = Positive Predictive Value; NPV = Negative Predictive Value; LR+ = Positive Likelihood Ratio; CIHI= Discharge Abstract Database, National Ambulatory Care Reporting System or Ontario Mental Health Reporting System

Figure 1. Cohort Build

Figure 2. Estimated number of individuals and population-prevalence (per 100 adults) experiencing homelessness in Ontario from 2007 to 2016 using the optimal annual housing experience case ascertainment algorithm (any CIHI indicator +/-15 days), with 95% confidence intervals, correcting for sensitivity. Annual Percentage Change with confidence interval was calculated using a Poisson regression


Adult Ontario Population
4,416,775 individuals
HHiT Study Participants
8,105 housing periods (795 individuals)
Data cleaning exclusions
a. Invalid or duplicate identifier
b. Invalid age
d. Non-Ontario resident
e. Death before end of observation period
TOTAL excluded from data cleaning
0 individuals
795,723 individuals
0 individuals
2,097 individuals
797,820 individuals
104 individuals
<=5 individuals
<=5 individuals
0 individuals
109 individuals
Exclusion 1: Identifier in HHiT group
469 individuals
Exclusion 2: Homeless postal code in any observation year
20,968 individuals
Exclusion 3: Death before start of observation period
59,173 individuals
Exclusion 4: <18 years at start of observation period
191,381 individuals
Exclusion 5: Not in Toronto or Ottawa at start of observation period
1,662 individuals
Exclusion 6: Not randomly selected
3,208,102 individuals
# Remaining
137,200 individuals
137,200 housed episodes
# Remaining
686 individuals
3,443 homeless episodes
3,505 housed episodes


Annual Percent Change:
4.43% (4.19%-4.68%)
Number of individuals
Population-prevalence per 100 adults
70000
60000
50000
40000
30000
20000
10000
0
0.65%
0.60%
0.55%
0.50%
0.45%
0.40%
0.35%
0.30%
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
# Homeless
% Homeless

## Supplement Table 1 – Data Elements Indicative of Homelessness, Supportive Housing or Shelter Use

| Database | Variable Name | Indicator Value | Description |
|---|---|---|---|
| **DAD** | HOMELESS | “Y” | Homelessness indicator |
| | INSTTYPE | “SH” | Institution Type = Supportive Housing |
| | DX10CODE1 to DX10CODE25 | “Z590” or “Z591” | ICD-10 diagnosis codes for “Homelessness” and “Inadequate housing” |
| | CMGDIAG | “Z590” or “Z591” | ICD-10 diagnosis codes for “Homelessness” and “Inadequate housing” |
| | PSTLCODE | “XX” | Used to indicate transient/homeless patients |
| **NACRS** | DX10CODE1 to DX10CODE10 | “Z590” or “Z591” | ICD-10 diagnosis codes for “Homelessness” and “Inadequate housing” |
| | RESTYPE | “3” or “4” | Residence Type = “Homeless” or “Shelter” |
| | PSTLCODE | “XX” | Used to indicate transient/homeless patients |
| **OMHRS** | PREDX10CODE to PREDX10CODE11 | “Z590” or “Z591” | ICD-10 diagnosis codes for “Homelessness” and “Inadequate housing” |
| | POSTDX10CODE1 to POSTDX10CODE24 | “Z590” or “Z591” | ICD-10 diagnosis codes for “Homelessness” and “Inadequate housing” |
| | PRIOR_RESIDENCE | “6” | Prior residential status = “Homeless (with or without shelter)” |
| | USUAL_RESIDENCE | “8” | Usual residential status = “Homeless (with or without shelter)” |
| | ADMITFROM | “8” | Admitted from = “Homeless (with or without shelter)” |
| | DISCHLIVING | “8” | Living arrangement at discharge = “Homeless (with or without shelter)” |
| | P5_Retired_2009 | “6” | (Variable retired in 2009) Living arrangement = “Homeless (with or without shelter)” |
| | PSTLCODE | “XX” | Used to indicate transient/homeless patients |
| **HCD** | DXCODE | “V600” or “V601” | ICD-9 diagnosis codes for “Lack of housing” or “Inadequate housing” |
| | REQUEST_PROGRAM | “6” | Program Requested = “Supportive Housing” |
| | RESIDENCE_TYPE | “1604”, “2200” or “3400” | Residence Type = “Other Supportive Living Unit”, “Hostel/Shelter” or “No fixed address” |
| **RAICA** | B4 | “8” | Expected residential/living status during service provision = “Homeless (with / without shelter)” |
| **NRS** | ALIVESET | “6” | Admission living setting = “Shelter” |
| | FLIVESET | “6” | Follow-up living setting = “Shelter” |

| Database | Variable Name | Indicator Value | Description |
|---|---|---|---|
| | PRIM_DISCH_WAIT_REASON | "1.1" | Primary Discharge Wait Reason = "Assisted Living/Supportive Housing" |
| | SECND_DISCH_WAIT_REASON | "1.1" | Secondary Discharge Wait Reason = "Assisted Living/Supportive Housing" |
| **CORR** | PROVINCE_CODE | "XX" | "Transient/Homeless" |
| | HEALTH_CARD_PROVINCE_CODE | "XX" | "Transient/Homeless" |

ICD=International Classification of Diseases

## Supplement Table 2: Description of Case Ascertainment Algorithms

| Name | Data Sources included[1] | Time Interval | Case Positive Condition(s) |
|---|---|---|---|
| **1 indicator +/- 0 days** | DAD<br>NACRS<br>OMHRS<br>CORR<br>RAICA<br>HCD<br>NRS | 0 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator[2] in any of the included sources within the specified time frame |
| **1 indicator +/- 15 days** | DAD<br>NACRS<br>OMHRS<br>CORR<br>RAICA<br>HCD<br>NRS | 15 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included sources within the specified time frame |
| **1 indicator +/- 45 days** | DAD<br>NACRS<br>OMHRS<br>CORR<br>RAICA<br>HCD<br>NRS | 45 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included sources within the specified time frame |
| **1 indicator +/- 90 days** | DAD<br>NACRS<br>OMHRS<br>CORR<br>RAICA<br>HCD<br>NRS | 90 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included sources within the specified time frame |
| **1 indicator +/- 180 days** | DAD<br>NACRS<br>OMHRS<br>CORR<br>RAICA<br>HCD<br>NRS | 180 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included sources within the specified time frame |

| Name | Data Sources included[1] | Time Interval | Case Positive Condition(s) |
|---|---|---|---|
| **1 indicator OR postal code +/- 0 days** | DAD<br>NACRS<br>OMHRS<br>CORR<br>RAICA<br>HCD<br>NRS<br>ICES PSTLYEAR | 0 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included health encounter sources or ICES PSTLYEAR-matched facilities providing shelter services. |
| **1 indicator OR postal code +/- 15 days** | DAD<br>NACRS<br>OMHRS<br>CORR<br>RAICA<br>HCD<br>NRS<br>ICES PSTLYEAR | 15 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included health encounter sources or ICES PSTLYEAR-matched facilities providing shelter services. |
| **1 indicator OR postal code +/- 45 days** | DAD<br>NACRS<br>OMHRS<br>CORR<br>RAICA<br>HCD<br>NRS<br>ICES PSTLYEAR | 45 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included health encounter sources or ICES PSTLYEAR-matched facilities providing shelter services. |
| **1 indicator OR postal code +/- 90 days** | DAD<br>NACRS<br>OMHRS<br>CORR<br>RAICA<br>HCD<br>NRS<br>ICES PSTLYEAR | 90 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included health encounter sources or ICES PSTLYEAR-matched facilities providing shelter services. |

| Name | Data Sources included[1] | Time Interval | Case Positive Condition(s) |
|---|---|---|---|
| **1 indicator OR postal code +/- 180 days** | DAD<br>NACRS<br>OMHRS<br>CORR<br>RAICA<br>HCD<br>NRS<br>ICES PSTLYEAR | 180 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included health encounter sources or ICES PSTLYEAR-matched facilities providing shelter services. |
| **1 CIHI indicator +/- 0 days** | DAD<br>NACRS<br>OMHRS | 0 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included sources within the specified time frame |
| **1 CIHI indicator +/- 15 days** | DAD<br>NACRS<br>OMHRS | 15 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included sources within the specified time frame |
| **1 CIHI indicator +/- 45 days** | DAD<br>NACRS<br>OMHRS | 45 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included sources within the specified time frame |
| **1 CIHI indicator +/- 90 days** | DAD<br>NACRS<br>OMHRS | 90 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included sources within the specified time frame |
| **1 CIHI indicator +/- 180 days** | DAD<br>NACRS<br>OMHRS<br>ICES PSTLYEAR | 180 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included sources within the specified time frame |
| **1 CIHI indicator OR postal code +/- 0 days** | DAD<br>NACRS<br>OMHRS<br>ICES PSTLYEAR | 0 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included health encounter sources or ICES PSTLYEAR-matched facilities providing shelter services. |
| **1 CIHI indicator OR postal code +/- 15 days** | DAD<br>NACRS<br>OMHRS<br>ICES PSTLYEAR | 15 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included health encounter sources or ICES PSTLYEAR-matched facilities providing shelter services. |

| Name | Data Sources included[1] | Time Interval | Case Positive Condition(s) |
|---|---|---|---|
| **1 CIHI indicator OR postal code +/- 45 days** | DAD<br>NACRS<br>OMHRS<br>ICES PSTLYEAR | 45 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included health encounter sources or ICES PSTLYEAR-matched facilities providing shelter services. |
| **1 CIHI indicator OR postal code +/- 90 days** | DAD<br>NACRS<br>OMHRS<br>ICES PSTLYEAR | 90 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included health encounter sources or ICES PSTLYEAR-matched facilities providing shelter services. |
| **1 CIHI indicator OR postal code +/- 180 days** | DAD<br>NACRS<br>OMHRS<br>ICES PSTLYEAR | 180 days before the encounter start or after the encounter end | 1 positive ("homeless") indicator in any of the included health encounter sources or ICES PSTLYEAR-matched facilities providing shelter services. |

1. Data sources are named and described in Supplement Table 3
2. indicators in each data sources are presented in Supplement Table 1

## Supplement Table 3: Databases Used

| Name | Data Source | Description |
|---|---|---|
| **Health and Housing in Transition Study** | Primary data collection | A longitudinal study conducted from 2009-2014 in three Canadian cities (Toronto, Ontario; Ottawa, Ontario; and Vancouver, British Columbia) aiming to assess the impact of housing transitions on health. Participants were randomly selected at shelters, meal programmes, community health centres, drop-in centres, rooming houses, and single-room occupancy hotels from January to December 2009 and were interviewed every 12 months.<br><br>Data on housing status were initially classified into one of 25 types of residence, which were then further classified into one of three mutually exclusive residence categories: housed, institution or homeless. To determine if periods of time spent in institutions (e.g. hospitals, prison, etc.) should be considered periods of homelessness or housing, housing status prior and subsequent to the period of institutionalization were reviewed, and institution housing episodes flanked by any period of homelessness was also considered homelessness. |
| **Canadian Institute for Health Information Discharge Abstract Database (DAD)** | Canadian Institute for Health Information (CIHI) | The DAD contains administrative, clinical (diagnoses and procedures/interventions), demographic, and administrative information for all admissions to acute care hospitals in Ontario. At ICES, consecutive DAD records are linked together to form 'episodes of care' among the hospitals to which patients have been transferred after their initial admission |
| **Ontario Mental Health Reporting System (OMHRS)** | Canadian Institute for Health Information (CIHI) | The OMHRS contains administrative, clinical (diagnoses and procedures), demographic, and administrative information for all admissions to adult designated inpatient mental health beds. This includes beds in general hospitals, provincial psychiatric facilities, and specialty psychiatric facilities. Clinical assessment data is ascertained using the Resident Assessment Instrument for Mental Health (RAI-MH), but different amounts of information are collected using this instrument depending on the length of stay in the mental health bed. Multiple assessments may occur during the length of a mental health admission. |
| **National Ambulatory Care Reporting System (NACRS)** | Canadian Institute for Health Information (CIHI) | The NACRS contains administrative, clinical (diagnoses and procedures), demographic, and administrative information for all patient visits made to hospital- and community-based ambulatory care centres (emergency departments, day surgery units, hemodialysis units, and cancer care clinics) in Ontario. At ICES, NACRS records are linked with other data sources (DAD, Ontario Mental Health Reporting System [OMHRS]) to identify transitions to other care settings, such as inpatient acute care or psychiatric care. |

| Name | Data Source | Description |
|---|---|---|
| **Home Care Database (HCD)** | Ontario Association of Community Care Access Centres | The Home Care Database contains administrative data about the patients, episodes, and services who receive home care through CCACs. The data included here is extracted from the CCAC administrative data system (CHRIS). |
| **Resident Assessment Instrument Contact Assessment Database (RAICA)** | Ontario Association of Community Care Access Centres | The interRAIContact Assessment (interRAICA) is a short screening assessment completed for adults at the time of intake to CCAC service (i.e. home care and / or palliative care) from community or hospital (including ED). It was designed to support decision-making about the urgency for immediate service provision, record essential clinical information on persons who would not be receiving comprehensive assessment at a later stage, and provide the minimum clinical information to enable short-term services to be put in place before completion of a full RAI assessment (ie. RAI-HC) |
| **National Rehabilitation Reporting System (NRS)** | Ministry of Health and Long-Term Care | The National Rehabilitation Reporting System (NRS) contains client data collected from participating adult inpatient rehabilitation facilities and programs across Canada. Data elements include socio-demographic information, administrative data, patient health characteristics, activities and participation and interventions. |
| **Canadian Organ Replacement Registry (CORR)** | Canadian Institute for Health Information (CIHI) | The Ontario portion of the Canadian Organ Replacement Register (CORR) records activity and outcomes of vital organ transplantation and renal dialysis activities. |
| **ICES-derived PSTLYEAR database** | ICES; Ministry of Health and Long-Term Care | The ICES-derived PSTLYEAR database contains the best known postal code for persons in the OHIP Registered Persons Database on July 1$^{st}$ of each year starting from year 1991. Postal codes supplied by the Ministry of Health and Long-Term Care are enriched with information in CIHI and other ICES-housed datasets to take advantage of the postal code information recorded each time an individual accesses certain healthcare services. |
| **OHIP Registered Persons Database** | Ministry of Health and Long-Term Care | The OHIP RPDB provides basic demographic information (age, sex, location of residence, date of birth, and date of death for deceased individuals) for those issued an Ontario health insurance number. The RPDB also indicates the time periods for which an individual was eligible to receive publicly funded health insurance benefits and provides the best known postal code for each registrant on July 1st of each year. |

| Name | Data Source | Description |
| --- | --- | --- |
| **Ontario Health Insurance Plan (OHIP)** | Ministry of Health and Long-Term Care | The OHIP claims database contains information on inpatient and outpatient services provided to Ontario residents eligible for the province's publicly funded health insurance system by fee-for-service health care practitioners (primarily physicians) and "shadow billings" for those paid through non-fee-for-service payment plans. Billing codes on the claims (OHIP fee codes) identify the care provider, their area of specialization and the type and location of service. OHIP billing claims also contain a 3-digit diagnosis code - the main reason for the service - captured using a modified version of the ICD, 8th revision coding system. |
| **Immigration, Refugees, and Citizenship Canada's Permanent Resident database (IRCC)** | Immigration, Refugees and Citizenship Canada | The Ontario portion of the IRCC Permanent Resident Database includes immigration application records for people who initially applied to land in Ontario since 1985. The dataset contains permanent residents' demographic information such as country of citizenship, level of education, mother tongue, and landing date. New immigrants who are currently residing in Ontario but originally landed in another province are not captured in this dataset. |

| Name | Data Source | Description |
|---|---|---|
| **Ontario COPD Database (COPD)** | Canadian Institute for Health Information (CIHI) | The Ontario COPD Database is created using two separate algorithms applied to inpatient hospitalization (DAD), same day surgery (SDS) records, and physician billing claims (OHIP) data to determine the diagnosis date for incident cases of chronic obstructive pulmonary disease in Ontario.<br><br>In an algorithm which maximizes sensitivity, the definition for COPD is any physician billing claim with a diagnosis for COPD (OHIP diagnosis codes: 491, 492, 496) or any inpatient hospitalization or same day surgery record with a diagnosis for COPD (ICD-9 diagnosis codes: 491, 492, 496; ICD-10 diagnosis codes: J41- J44; in any diagnostic code space). When using expert panel review of primary care charts as the reference standard, this definition has been shown to have the following performance characteristics: Sensitivity (85.0%), Specificity (78.4%), Positive Predictive Value (57.5%), and Negative Predictive Value (93.8%).(7)<br><br>In an algorithm which maximizes specificity, the definition for COPD is ≥3 physician billing claims with a diagnosis for COPD (OHIP diagnosis codes: 491, 492, 496) or ≥1 inpatient hospitalization or same day surgery record with a diagnosis for COPD (ICD-9 diagnosis codes: 491, 492, 496; ICD-10 diagnosis codes: J41, J42, J43, J44; in any diagnostic code space) in a two-year period. When using expert panel review of primary care charts as the reference standard, this definition has been shown to have the following performance characteristics: Sensitivity (57.5%), Specificity (95.4%), Positive Predictive Value (81.3%), and Negative Predictive Value (86.7%).(1) |
| **Ontario Diabetes Database (ODD)** | Canadian Institute for Health Information (CIHI) | The ODD is created using algorithms applied to inpatient hospitalization (DAD) records, same day surgery (SDS) records, and physician billing claims (OHIP) data to determine the diagnosis date for incident cases of diabetes in Ontario. For adults aged 19 years and greater, the definition for diabetes is 2 physician billing claims with a diagnosis for diabetes (OHIP diagnosis code: 250) or 1 inpatient hospitalization or same day surgery record with a diagnosis for diabetes (ICD-9 diagnosis code: 250; ICD-10 diagnosis codes: E10, E11, E13, E14; in any diagnostic code space) within a 2 year period. Physician claims and hospitalizations with a diagnosis of diabetes occurring within 120 prior to and 180 days after a gestational hospitalization record were excluded. When using primary care chart abstraction as the reference standard, this definition has been shown to have the following performance characteristics: Sensitivity (86.1%), Specificity (97.1%), Positive Predictive Value (79.8%), and Negative Predictive Value (98.1%).(2) |

| Name | Data Source | Description |
|---|---|---|
| **Ontario CHF Database (CHF)** | Canadian Institute for Health Information (CIHI) | The Ontario CHF Database is created using a definition of ≥2 physician billing claims with a diagnosis of congestive heart failure (OHIP diagnosis code: 428) and/or ≥1 inpatient hospitalization or same day surgery record with a diagnosis of congestive heart failure (ICD-9 diagnosis code: 428; ICD-10 diagnosis code: I50; in the primary diagnostic code space) in a two-year period applied to hospitalization (DAD), same day surgery (SDS), and physician billing claims (OHIP) data to determine the diagnosis date for incident cases of CHF in Ontario.<br><br>When using electronic medical record data abstraction as the reference standard, the above definition has been demonstrated to have the following performance characteristics: Sensitivity (84.8%), Specificity (97.0%), and Positive Predictive Value (55.3%).(3) |
| **Ontario HIV Database (HIV)** | Canadian Institute for Health Information (CIHI) | The Ontario HIV Database contains all Ontario HIV positive patients identified since 1992. HIV positive patients are defined as persons having received at least 3 physician claims with OHIP diagnosis code 042, 043, or 044 within 3 years. The diagnosis date is the first of these claims, unless a prior OHIP record with the above diagnosis codes or a hospitalization having an ICD-10 diagnosis code of B20, B21, B22, B23, or B24 occurs earlier.<br><br>This definition has been shown to have high sensitivity (96.2%) and specificity (99.6%)(4) |

# Supplement Table 4: Variable Definitions

| Variable | Data Source | Definition Description |
|---|---|---|
| **Age** | RPDB | Age of the individual at the index date |
| **Sex** | RPDB | Sex of the individual |
| **Rural status** | RPDB | Resides in a rural area as defined as a settlement of <10 000 individuals |
| **Location (city)** | RPDB | City in which the individual is believed to reside as of July 1st of the index year, based on their census division information |
| **Recent immigrant** | IRCC | Presence of a landing date in the Immigration, Refugees and Citizenship Canada Permanent Database indicates immigration to Ontario between 1985 to 2018 |
| **Date of immigration** | IRCC | Time, in years, since immigration to Ontario from outside Canada occurred |
| **Refugee status** | IRCC | Class of immigration status = Refugee |
| **Congestive heart failure** | CHF | Presence in the database indicates the individual has a history of congestive heart failure[1] |
| **Chronic obstructive pulmonary disease** | COPD | Presence in the database indicates the individual has a history of COPD[2] |
| **Diabetes** | ODD | Presence in the database indicates the individual has a history of diabetes[3] |
| **HIV status** | HIV | Presence in the database indicates the individual is HIV positive[4]. |
| **Chronic kidney disease** | DAD, NACRS, OHIP | 1 hospitalization or 3 ED visit or physician claims in 1 year within 3 years of the index date with any of the following eligible codes:<br>ICD-10: E102, E112, E132, E142, I12, I13, N00, N01, N02, N03, N04, N05, N06, N07, N08, N1, N20, N21, N22, N23<br>OHIP dx: 403, 585 |
| **Chronic liver disease** | DAD, NACRS, OHIP | 1 hospitalization, ED visit or physician claim within 3 years of the index date with any of the following eligible codes:<br>ICD-10: B16, B17, B18, B19, B942, E830, E831, I85, K70, K713, K714, K715, K717, K721, K729, K73, K74, K753, K754, K758, K759, K76, K77, R160, R162, R17, R18, Z225<br>OHIP dx: 070, 571, 573<br>OHIP fee: Z551, Z554 |
| **Psychosis related mental health care** | DAD, NACRS, OMHRS, OHIP | 1 hospitalization, ED visit or physician claim within 1 year of the index date with any of the following eligible codes:<br>ICD-10: F20, F22, F23, F24, F25, F28, F29<br>DSM-IV: 295, 297, 298<br>OHIP dx: 295, 297, 298 |

| Variable | Data Source | Definition Description |
|---|---|---|
| **Non-psychotic disorders related mental health care** | DAD, NACRS, OMHRS, OHIP | 1 hospitalization, ED visit or physician claim within 1 year of the index date with any of the following eligible codes:<br>ICD-10: F30, F31, F32, F33, F34, F38, F39, F40, F41, F42, F43, F48, F60, F93<br>DSM-IV: 296, 300, 301<br>OHIP dx: 296, 300, 301, 309, 311 |
| **Substance use related mental health care** | DAD, NACRS, OMHRS, OHIP | 1 hospitalization, ED visit or physician claim within 1 year of the index date with any of the following eligible codes:<br>ICD-10: F10, F11, F12, F13, F14, F15, F16, F17, F18, F19, F55<br>DSM-IV: 291, 292, 303, 304, 305<br>OHIP dx: 291, 292, 303, 304, 305 |
| **Outpatient visits** | OHIP | Number of physician visits within 1 year prior to the index date, defined as one visit per day per physician |
| **Emergency department visits** | NACRS | Number of ED visits within 1 year prior to the index date |
| **Hospitalizations** | DAD | Number of admissions to acute care hospitals within 1 year prior to the index date. |

(1) Gershon AS, Wang C, Guan J, Vasilevska-Ristovska J, Cicutto L, To T. Identifying individuals with physcian diagnosed COPD in health administrative databases. COPD 2009; 6(5):388-394.

(2) Hux JE, Ivis F, Flintoft V, Bica A. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. Diabetes Care 2002; 25(3):512-516.

(3) Schultz SE, Rothwell DM, Chen Z, Tu K. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. Chronic Dis Inj Can 2013; 33(3):160-166.

(4) Tony Antoniou, Brandon Zagorski, Mona R. Loutfy, Carol Strike, Richard H. Glazier. Validation of Case-Finding Algorithms Derived from Administrative Data for Identifying Adults Living with Human Immunodeficiency Virus Infection. Plos One. 2011;6(6):e21748. Epub 2011 Jun 30.

# Supplement Table 5 – Validation Statistic Formulae

The following diagnostic tests were used to assess the validity of each case ascertainment algorithm.

| Validation Statistic | Formula |
|---|---|
| **Sensitivity** | TP / (TP + FN) |
| **Specificity** | TN / (FP + TN) |
| **Positive Predictive Value** | TP / (TP + FP) |
| **Negative Predictive Value** | TN / (FN + TN) |
| **Positive Likelihood Ratio** | Sensitivity / (1 - Specificity) |

TP=True positive (truly experiencing homelessness and flagged as homeless by the case ascertainment algorithm)
FP=False positive (truly housed but flagged as homeless by the case ascertainment algorithm)
FN=False negative (truly experiencing homelessness but not flagged as homeless by the case ascertainment algorithm)
TN=True negative (truly housed and flagged as housed by the case ascertainment algorithm)

# Supplement Table 6 – Additional Tables

### Table 6A – Characteristics of true positives, false positives and false negatives using the optimal housing episode algorithm

| | True Positives (N=613) | False Positives (N=595) | False Negatives (N=2,830) |
|---|---|---|---|
| **Episodes without encounters, n (% of group)** | 0 (0%) | 0 (0%) | 1,825 (64.5%) |
| **Cohort source = HHiT study, n (% of group)** | 613 (100%) | 397 (66.7%) | 2,830 (100%) |

Optimal housing episode algorithm = 1 CIHI indicator +/-45 days of the housing episode start and end dates

### Table 6B – Characteristics of true positives, false positives and false negatives using the (non-scalable) optimal annual housing experience algorithm

| | True Positives (N=701) | False Positives (N=365) | False Negatives (N=1,589) |
|---|---|---|---|
| **Episodes without encounters, n (% of group)** | 0 (0%) | 0 (0%) | 997 (62.7%) |
| **Cohort source = HHiT study, n (% of group)** | 701 (100%) | 115 (31.5%) | 2,830 (100%) |

Optimal annual housing experience algorithm = 1 CIHI indicator +/-15 days of the calendar year start and end dates or one postal code from PSTLYEAR

**Table 6C – Number of adult Ontarians identified as experiencing homelessness by the optimal annual housing experience algorithm between 2007 and 2016**

| Year | # identified (95% CI) | Adult ON Population | Unadjusted Rate (95% CI) |
|---|---|---|---|
| **2007** | 7,012 (6,850-7,178) | 9,995,143 | 0.07% (0.069% - 0.072%) |
| **2008** | 7,271 (7,106-7,440) | 10,125,078 | 0.072% (0.07% - 0.073%) |
| **2009** | 7,318 (7,152-7,488) | 10,250,718 | 0.071% (0.07% - 0.073%) |
| **2010** | 7,934 (7,761-8,110) | 10,393,961 | 0.076% (0.075% - 0.078%) |
| **2011** | 8,521 (8,342-8,704) | 10,529,817 | 0.081% (0.079% - 0.083%) |
| **2012** | 9,028 (8,844-9,216) | 10,699,090 | 0.084% (0.083% - 0.086%) |
| **2013** | 9,202 (9,016-9,392) | 10,859,071 | 0.085% (0.083% - 0.086%) |
| **2014** | 9,769 (9,577-9,965) | 11,001,544 | 0.089% (0.087% - 0.091%) |
| **2015** | 10,658 (10,458-10,862) | 11,117,135 | 0.096% (0.094% - 0.098%) |
| **2016** | 11,731 (11,521-11,945) | 11,287,810 | 0.104% (0.102% - 0.106%) |
| **Total individuals identified over 10 years** | | | **54,873** |
| **Individuals present in > 1 year estimate** | | | **18,217 (33.2% of total)** |

Adult ON Population derived from Ontario inter-censal population estimates.

Optimal annual housing experience algorithm = 1 CIHI indicator +/-15 days of the calendar year start and end dates.

Confidence intervals calculated using the Wilson score method.

# Reporting checklist for diagnostic test accuracy study.

Based on the STARD guidelines.

## Instructions to authors

Complete this checklist by entering the page numbers from your manuscript where readers will find each of the items listed below.

Your article may not currently address all the items on the checklist. Please modify your text to include the missing information. If you are certain that an item does not apply, please write "n/a" and provide a short explanation.

Upload your completed checklist as an extra file when you submit to a journal.

In your methods section, say that you used the STARD reporting guidelines, and cite them as:

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, LijmerJG Moher D, Rennie D, de Vet HCW, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF, For the STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies.

| | | Reporting Item | Page Number |
|---|---|---|---|
| | #1 | Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC) | 1 |
| | #2 | Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts) | 2 |
| | #3 | Scientific and clinical background, including the intended use and clinical role of the index test | 3-5 |
| | #4 | Study objectives and hypotheses | 3 |
| Study design | #5 | Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study) | 5 |
| Participants | #6 | Eligibility criteria | 6 |

| | | | |
|---|---|---|---|
| | #7 | On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry) | 6 |
| | #8 | Where and when potentially eligible participants were identified (setting, location and dates) | 6 |
| | #9 | Whether participants formed a consecutive, random or convenience series | 6 |
| Test methods | #10a | Index test, in sufficient detail to allow replication | 7-8 |
| | #10b | Reference standard, in sufficient detail to allow replication | 7 |
| | #11 | Rationale for choosing the reference standard (if alternatives exist) | 7 |
| | #12a | Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory | See note 1 |
| | #12b | Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory | 7-8 |
| | #13a | Whether clinical information and reference standard results were available to the performers / readers of the index test | See note 2 |
| | #13b | Whether clinical information and index test results were available to the assessors of the reference standard | See note 3 |
| Analysis | #14 | Methods for estimating or comparing measures of diagnostic accuracy | 8 |
| | #15 | How indeterminate index test or reference standard results were handled | See note 4 |
| | #16 | How missing data on the index test and reference standard were handled | 7-8 |
| | #17 | Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory | 7-8 |
| | #18 | Intended sample size and how it was determined | 6 |
| Participants | #19 | Flow of participants, using a diagram | Figure 1 |

| | | | |
|---|---|---|---|
| | #20 | Baseline demographic and clinical characteristics of participants | 9 |
| | #21a | Distribution of severity of disease in those with the target condition | 9 |
| | #21b | Distribution of alternative diagnoses in those without the target condition | See note 5 |
| | #22 | Time interval and any clinical interventions between index test and reference standard | See note 6 |
| Test results | #23 | Cross tabulation of the index test results (or their distribution) by the results of the reference standard | 9-10, Table 2 |
| | #24 | Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals) | 9-10, Table 2 |
| | #25 | Any adverse events from performing the index test or the reference standard | See note 7 |
| | #26 | Study limitations, including sources of potential bias, statistical uncertainty, and generalisability | 11-13 |
| | #27 | Implications for practice, including the intended use and clinical role of the index test | 11-13 |
| | #28 | Registration number and name of registry | See note 8 |
| | #29 | Where the full study protocol can be accessed | See note 9 |
| | #30 | Sources of funding and other support; role of funders | See note 10 |

## Author notes

1. n/a - variables are binary
2. n/a - index test uses administrative data, i.e. there were no index test performers
3. n/a - index test uses administrative data. i.e. by definition the index test was not available to those assessing the reference standard
4. n/a - no indeterminate results were possible
5. n/a - those without target definition were assumed housed by default, as described in the Methods

6. n/a - no clinical interventions are relevant and time intervals were included in case algorithm definitions, as described in the Methods

7. n/a - not a clinical test

8. n/a - not registered

9. n/a - full protocol described in-text

10. 1 (title page), 11 (acknowledgements)