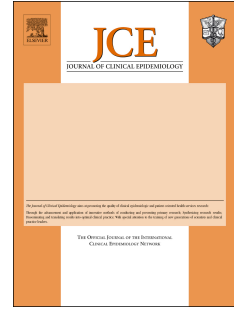


# Journal Pre-proof

Identification of homelessness using health administrative data in Ontario, Canada following a national coding mandate: a validation study

Lucie Richard, MA, Brooke Carter, MSc, Rosane Nisenbaum, PhD, Michael Liu, MPhil, Stephen W. Hwang, MD MPH



PII: S0895-4356(24)00185-9

DOI: <https://doi.org/10.1016/j.jclinepi.2024.111430>

Reference: JCE 111430

To appear in: *Journal of Clinical Epidemiology*

Received Date: 18 March 2024

Revised Date: 4 June 2024

Accepted Date: 10 June 2024

Please cite this article as: Richard L, Carter B, Nisenbaum R, Liu M, Hwang SW, Identification of homelessness using health administrative data in Ontario, Canada following a national coding mandate: a validation study, *Journal of Clinical Epidemiology* (2024), doi: <https://doi.org/10.1016/j.jclinepi.2024.111430>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier Inc.

## Identification of homelessness using health administrative data in Ontario, Canada following a national coding mandate: a validation study

\*Lucie Richard<sup>1</sup>, MA; Brooke Carter<sup>2</sup>, MSc; Rosane Nisenbaum, PhD<sup>1</sup>; Michael Liu, MPhil<sup>3</sup>; Stephen W. Hwang<sup>1,4,5,6</sup>, MD MPH

1. MAP Centre for Urban Health Solutions, Unity Health Toronto, Toronto, ON, Canada
2. ICES Western, London Health Sciences Research Institute, London, ON, Canada
3. Harvard Medical School, Boston, MA, USA
4. Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
5. Division of General Internal Medicine, Department of Medicine, University of Toronto, Toronto, ON, Canada
6. ICES, Toronto, Ontario, Canada

**Corresponding author:**

Lucie Richard  
MAP Centre for Urban Health Solutions, Unity Health Toronto  
30 Bond St.,  
Toronto ON, M5B 1X1  
Telephone: 416-864-5991  
Email: [lucie.richard@unityhealth.to](mailto:lucie.richard@unityhealth.to)  
ORCID: [0000-0001-6577-5067](https://orcid.org/0000-0001-6577-5067)

**ABSTRACT**

**Objective:** Conducting longitudinal health research about people experiencing homelessness poses unique challenges. Identification through administrative data permits large, cost-effective studies; however, case validity in Ontario is unknown after a 2018 Canada-wide policy change mandating homelessness coding in hospital databases. We validated case definitions for identifying homelessness using Ontario health administrative databases after introduction of this coding mandate.

**Study Design and Setting:** We assessed 42 case definitions in a representative sample of people experiencing homelessness in Toronto (n=640) from whom longitudinal housing history (ranging from 2018 to 2022) was obtained; and a randomly selected sample of presumably housed people (n=128,000) in Toronto. We evaluated sensitivity, specificity, positive and negative predictive values, and positive likelihood ratios to select an optimal definition, and compared the resulting true positives against false positives and false negatives to identify potential causes of misclassification.

**Results:** The optimal case definition included any homelessness indicator during a hospital-based encounter within 180 days of a period of homelessness (sensitivity=52.9%; specificity=99.5%). For periods of homelessness with  $\geq 1$  hospital-based healthcare encounter, the optimal case definition had greatly improved sensitivity (75.1%) while retaining excellent specificity (98.5%). Review of false positives suggested that homeless status is sometimes erroneously carried forward in healthcare databases after an individual transitioned out of homelessness.

**Conclusions:** Case definitions to identify homelessness using Ontario health administrative data exhibit moderate to good sensitivity and excellent specificity. Sensitivity has more than doubled since the implementation of a national coding mandate. Mandatory collection and reporting of homelessness information within administrative data present invaluable opportunities for advancing research on the health and healthcare needs of people experiencing homelessness.

Keywords: Homeless persons; Coding mandate; Health services; Validation studies; Canada;

Word count: 3301

## INTRODUCTION

Homelessness is a significant and growing public health issue in Canada. People experiencing homelessness face substantial stigma and structural inequities that result in adverse health and quality of life outcomes [1]. Yet, effective response to this complex issue is hindered by challenges in acquiring reliable data related to this population. Longitudinal studies [2-4], while generally rigorous, are highly resource-intensive and face unique challenges in recruitment and retention, thus typically yielding small sample sizes [5, 6]. Cross-sectional studies and point-in-time collections, by contrast, are less resource-intensive, but may capture non-representative samples, such as when conducting convenience or other non-random sampling, working at specific shelters or collecting data at specific times of year [7-10]. Cross-sectional studies also, by definition, cannot follow individuals over time to assess changes in health status or healthcare utilization.

A growing number of studies have leveraged health administrative data to conduct research within this population [for example, 11-16]. In areas where databases are standardized across large populations or service areas (as in Canada), health administrative data have many benefits, including low cost and broad coverage. For this reason, they are already widely used for population surveillance of health conditions [17, 18]. However, administrative data are not designed for research and can therefore be prone to misclassification errors [19]. Therefore, validation is needed to ensure adequate performance. This is particularly true for social determinants of health, such as homelessness, as this information may be reported infrequently or inconsistently [20-22]. In Ontario, we previously validated indicators of homelessness in Canadian Institute for Health Information (CIHI)

administrative databases, finding that they had low sensitivity [23]. However, beginning in 2018, CIHI required that all Canadian hospitals record homelessness in data submissions using International Classification of Disease-10<sup>th</sup> Revision (ICD-10) code 'Z59.0' ('Homelessness')[24]. This resulted in substantial increases in the number of patients coded as experiencing homelessness across Canada [25]. However, whether this policy change affected case validity for identifying homelessness is currently unknown.

This study used data from a recent prospective cohort study to validate case definitions to identify homelessness in Ontario, Canada, after CIHI's mandate seeking to improve documentation of homelessness in health administrative data.

## **METHODS**

### **Study Design, Setting and Data Sources**

We conducted this validation study in Toronto, Canada, a city on treaty 13 territory in the province of Ontario. In Ontario, healthcare is administered through a single-payer model, with universal coverage of medical services provided through the Ontario Health Insurance Plan (OHIP)[26] and other federal plans for individuals ineligible for OHIP [27]. Thus, administrative data for health services provided in this region covers the vast majority of the population.

This study used data from the *Ku-gaa-gii pimitizi-win* study [4], a representative cohort of people experiencing homelessness in Toronto; and from ICES, an independent, non-profit research institute whose legal status under Ontario's health information privacy law allows it to collect and analyze health care and demographic data for health system evaluation and improvement. The following ICES databases, described in Supplement A, were used: the ICES

Registered Persons Database (RPDB); the Discharge Abstract Database (DAD)[28]; the National Ambulatory Care Reporting System (NACRS) database [29]; the Ontario Mental Health Reporting System (OMHRS) database [30]; the OHIP claims database [31]; the Community Health Centre (CHC) database [32]; the Chronic Obstructive Pulmonary Disease Database [33]; the Ontario Asthma Database [34]; the Ontario Diabetes Database [35]; the Congestive Heart Failure Database [36]; and the Ontario Hypertension Database [37].

All databases were linked using unique encoded identifiers and analyzed at ICES [38]. This study is reported following the STARD guidelines for reporting diagnostic accuracy studies (Supplement B).

### **Ethical Review and Data Availability**

This study received ethics approval from the Research Ethics Board at Unity Health Toronto (REB# 20-272). Due to legal data sharing agreements between ICES and its data providers, we cannot make the dataset underlying analysis publicly available; however, access for confidential access may be granted to parties meeting pre-specified criteria through the ICES DAS program. The protocol and code underlying the analysis is available upon request to the Corresponding Author.

### **Participants and Reference Standard**

The reference standard was housing history obtained longitudinally from participants of the *Ku-gaa-gii pimitizi-win* study [4]. The design and recruitment strategy are available in the protocol [4]. Briefly, between June and September 2021, individuals 16 years or older

experiencing homelessness were randomly selected from 62 sites in Toronto; individuals were included if they provided informed consent and agreed to follow-up interviews. Participants were interviewed at baseline and every three months over approximately one year. Self-reported housing history was organized into consecutive housing episodes [39], which together constituted the participant's observation period. Housing information prior to baseline was truncated to begin no earlier than April 1, 2018.

Housing status was determined by classifying type of housing episodes into one of three categories: "housed", "institution" or "homeless" (Supplement C provides a classification scheme). In line with the most widely accepted definition of homelessness in Canada [40], sheltered, unsheltered and hidden homeless housing episode types were included within the homeless category. 'Institution' episodes (e.g. hospitals, correctional facilities) were resolved into housed or homeless categories based on the preceding and subsequent episodes, with institutional episodes flanked on either side by a homelessness episode also classified as homeless (since the individual was unhoused at entry and/or exit from the institution).

Because of the high prevalence of homelessness within this cohort over the observation period, we further included 200 presumably housed individuals (each having a single 'housed' episode) for each *Ku-gaa-gii pimitizi-win* participant. This approach yielded a study population with a prevalence of homelessness approximately similar to that of the Canadian population over the course of one year (~0.5%)[41], an approach used in previous validation studies [23, 42, 43]. Presumably housed individuals were randomly selected from the RPDB, which lists all persons in Ontario eligible for OHIP. Non-residents of Toronto or those who were homeless, as indicated by participation in the *Ku-gaa-gii pimitizi-win* study or by having a postal code



associated with a shelter service [44], were excluded. Each presumably housed individual was assigned a random start and end date following distributions found among *Ku-gaa-gii pimitizwin* participants.

### **Case definitions**

Case definitions included homelessness indicators present in DAD, NACRS, OMHRS and CHC databases. The first three are maintained by CIHI and report standardized information during hospital-based encounters (acute care hospitalizations, emergency department visits and psychiatric hospitalizations, respectively). The fourth reports encounters at Community Health Centres, a network of non-profit organizations that provide outpatient primary health care, with a particular mission to care for populations facing barriers to healthcare access [32]. Indicators included ICD-10-CA codes Z59.0 ('Homelessness') and Z59.1 ('Inadequate housing') as well as other database-specific indicators, such as residential type in NACRS or living arrangement at discharge in OMHRS. Postal codes are also recorded in all four databases; therefore, we additionally assessed postal codes against a list of known shelter organizations. Shelter postal codes also including residential addresses, as determined using a geographic information system through a process described elsewhere [44], were not used to avoid misclassifying housed individuals as homeless. A list of the homelessness indicators assessed is available in Supplement D.

A full list of case definitions is presented in Supplement E. Case definitions varied by period of time in which eligible healthcare encounters occurred during or after episode end (during or within 90 days after the episode; during or within 180 days after the episode; within

the same calendar year); by type of database included (any; hospital-based only; inpatient only; emergency department [ED] only; CHC only); and by indicators included (all, including postal codes; all, excluding postal codes; code Z59.0 only).

### **Covariates**

We obtained participant characteristics at the start of follow-up, including age, sex assigned at birth, Charlson comorbidity index category, and past diagnosis of hypertension [37], congestive heart failure [36], chronic obstructive pulmonary disease [33] or asthma [34], diabetes [35], chronic kidney disease, or liver disease. We additionally obtained recent healthcare utilization overall (hospital admissions, emergency department visits or outpatient visits) and for mental health or substance use disorders (overall and by subtype). Finally, for *Kugaa-gii pimitizi-win* participants the survey provided additional sociodemographic and behavioural information. Each covariate is fully defined in Supplement F.

### **Statistical analysis**

We present cohort demographics, comorbidities and recent healthcare utilization at the start of observation, using standardized differences to assess group differences [45] (where  $\geq 10\%$  was considered substantial). Validation statistics including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and positive likelihood ratio (LR+) (formulae listed in Supplement G) are then provided for a total of 42 case definition permutations. These case definitions are tested within the cohort overall, as well as among specific subgroups (male assigned at birth; female assigned at birth; youth and adults under 30

years of age; adults over 60 years of age; and individuals with recent substance use-related healthcare).

The primary analysis considers all housing episodes, validating the ability of health administrative data to identify experiences of homelessness overall (here, housing episodes without eligible healthcare encounters were coded as case negative, by default). However, as health or health services research often concerns itself only with the subset of people experiencing homelessness who interact with the healthcare system, a secondary analysis considering only housing episodes with at least one healthcare encounter was conducted to validate performance among patients experiencing homelessness who interact with the healthcare system.

Finally, using the case definition deemed optimal on the basis of its maximized validation characteristics and scalability across Canada, housing episodes identified as true positives (TP) are compared to false positives (FP) and false negatives (FN) to identify characteristics associated with misclassification, with  $\chi^2$ , one-way ANOVA, or Kruskal-Wallis tests used for statistical comparisons.

All analyses were conducted using SAS Enterprise Guide v8.3. Throughout, p-values of  $\leq 0.05$  were deemed significant, and cell counts  $\leq 5$  were suppressed to protect participant and patient privacy.

## RESULTS

A total of 640 *Ku-gaa-gii pimitizi-win* participants contributing 1,163 housing episodes and 128,000 presumably housed individuals were included (Figure 1). Supplement H, table 1

shows that *Ku-gaa-gii pimitizi-win* participants successfully linked to ICES were representative of *Ku-gaa-gii pimitizi-win* participants overall (n=736), except for a lower linkage rate among refugees and people with temporary or other citizenship status. *Ku-gaa-gii pimitizi-win* participants experienced homelessness for an average of 36.1% of their observation period, with a median homeless episode of 52 days (IQR: 0 - 293 days) (Table 1). We found substantial differences between *Ku-gaa-gii pimitizi-win* and housed samples, with *Ku-gaa-gii pimitizi-win* participants being more likely to be male, have higher Charlson comorbidity scores, use more healthcare for all service types (inpatient, ED and non-emergency outpatient), and more likely to have recent healthcare related to mental health or substance use.

[FIGURE 1]

[TABLE 1]

In the primary analysis, case definitions overall had sensitivities ranging between 1.5% and 59.2% and specificities all exceeding 99.5% (Table 2). The most sensitive case definitions included ED visits: CHC-only definitions ranged from 1.5 to 5.6% and inpatient-only definitions ranged from 8.5 to 18.8%. LR+ were all in excess of 70, indicating very substantial increases in the likelihood of being homeless when identified by the case definition [46]. Case definitions were slightly more sensitive among males assigned at birth (Supplement H, table 2a) and adults over 60 years old (Supplement H, table 2c) than among females assigned at birth (Supplement H, table 2b) and youth/adults under 30 years old (Supplement H, table 2d), but otherwise validity was similar among these subgroups. However, case definitions for individuals with recent substance use-related healthcare (Supplement H, table 2e) demonstrated higher sensitivity (up to 79.2%) but lower specificity (as low as 76.5%) and LR+ values well below 5. The

case definition including all indicators (excluding postal codes) from any hospital database within 180 days (sensitivity=52.9%; specificity=99.7%) was deemed optimal, based on its validation characteristics and scalability to other parts of Canada, and was therefore used for further evaluation.

[TABLE 2]

In the secondary analysis (restricting to housing episodes with  $\geq 1$  eligible healthcare encounter), case definitions overall (Table 3) and among subgroups (Supplement H Table 3a-3e) were much more sensitive, but otherwise had similar validity. The optimal case definition had a sensitivity of 75.1%, specificity of 98.5% and PPV of 35.9%, an absolute increase in sensitivity of more than 20% compared to the primary analysis, with only negligible decreases in specificity and PPV.

[TABLE 3]

Table 4 presents characteristics of individuals during housing episodes (with  $\geq 1$  eligible healthcare encounter) characterized as TP, FP and FN using the optimal case definition. Individuals more likely to be incorrectly identified as homeless (FPs) were youths or adults under 30 years old (24.4% [FPs] vs 10.9% [TPs]). Among *Ku-gaa-gii pimitizi-win* participants (Supplement H, table 4), FPs also had much shorter homeless episodes (median 27 days [FPs] vs 201 days [TPs]). Overall, 57.5% (257/447) of FPs were *Ku-gaa-gii pimitizi-win* participants who had become housed, and 35% (67/190) of the remaining FPs (from the housed group) had at least one healthcare encounter with a homelessness indicator within the past 5 years.

[TABLE 4]

Conversely, individuals were more likely to be incorrectly identified as housed (FNs) if they were assigned female at birth (34.3% [FNs] vs 22.3% [TPs]), or did not have recent mental health or substance use-related healthcare (Table 4). FNs also had far fewer healthcare encounters overall. Among *Ku-gaa-gii pimitizi-win* participants (Supplement H, table 4), FNs were more likely to self-identify as Black (22.8% [FNs] vs 9.5% [TPs]), and less likely to smoke tobacco (71.5% [FNs] vs 83.4% [TPs]), or use illegal drugs or prescriptions for non-medical reasons (26.6% [FNs] vs 46.4% [TPs]).

## DISCUSSION

We validated case definitions to identify homelessness in a longitudinally collected, representative sample of people experiencing homelessness and a random sample of housed individuals in Toronto, Canada, following the introduction of mandatory coding of homelessness in Canadian hospitals in 2018. Case definitions overall exhibited moderate sensitivities (optimal definition: 52.9%) and excellent specificities (>99%) and LR+ (over 70), but poor to moderate PPV (optimal definition: 36.2%). Sensitivity increased substantially compared to our prior validation [23], conducted before introduction of the national coding mandate [24]. The nearest equivalent case definition (1 CIHI indicator +/- 180 days) from this prior analysis demonstrated a sensitivity of only 24.8% [23].

Historically, ICD-10 “Z-codes” for social determinants of health have only been infrequently used [20-22]. Because of this, such data previously would result in limited and potentially biased cohorts, and thus be of limited value for research on the health and healthcare of people experiencing homelessness. However, ICD-10 codes have the major

benefit of being standardized and utilized by health systems in over 100 countries [47]. The CIHI policy built upon these benefits by introducing mandatory coding of Z59.0 based on documentation in patients' clinical records. This relatively simple initiative [48] clearly improved identification and reporting of homelessness in Canada, and may serve as a model for other jurisdictions and organizations working to improve collection of homelessness and other social determinants of health data, such as the US Centers for Medicare and Medicaid Services [49] or the UK National Health Service [50, 51].

Nevertheless, case definition sensitivities remain moderate. This is mostly explained by homeless housing periods occurring without eligible healthcare encounters (which were coded case negative, by default). We confirm in our secondary analysis that sensitivity improves significantly (between 20-25% higher) with negligible changes to specificity when we only consider housing episodes with at least one healthcare encounter. In this subgroup, optimal case validity (75.1% sensitivity; 98.5% specificity) approaches or is on par with well-accepted chronic disease case definitions like dementia, which also require patients interact with the healthcare system [52].

The low positive predictive value, reflecting a relatively large number of false positives, is less easily explained. A majority (~72%) of these were *Ku-gaa-gii pimitizi-win* participants who became housed or were presumed-housed individuals having a homelessness indicator prior to the observation period. This, combined with the significantly lower duration of homelessness episodes among FPs, suggests the low PPV are, at least in part, a result of previously documented homelessness being incorrectly carried forward during healthcare encounters after an individual transitions out of homelessness. Considering the large windows of time

during which we considered identification after an episode ended, this carry-forward problem is likely substantial, although it presents an obvious area for improvement of homelessness identification during the registration and admission process in Canadian hospitals.

Aside from this probable carryforward issue, some groups were modestly associated with being misclassified as housed, such as people identifying as Black, people without recent mental health or substance use-related healthcare, and people who do not smoke tobacco or do not use non-prescription or illicit drugs. Some characteristics (namely, self-identification as Black and lack of history of mental health or substance use-related healthcare) are also significantly associated with fewer healthcare encounters, and therefore these characteristics might be confounded by fewer opportunities for identification. However, it is also possible that fear of stigma or other barriers to disclosure may be contributing factors [53].

Our validation benefits from a reliable reference standard, derived from one of the largest representative cohorts of people experiencing homelessness with longitudinally-collected housing history since the At Home/Chez Soi study [3, 4]. However, we acknowledge that our data are not openly available nor was our study prospectively registered. Further, our results' generalizability is limited to Canada, where universal health insurance and standardized collection of administrative information create a unique data environment. As our reference standard was sampled from a highly urbanized region (Toronto), case validity might differ in small cities or rural areas.

Additionally, our methodology is reliant on interaction with the healthcare system. Where health research targets current clients of the healthcare system, this limitation is minor, as evidenced by the good sensitivity of case definitions for patients experiencing homelessness.



However, this methodology is not suitable for assessing the health needs and experiences of individuals who do not receive care at traditional healthcare settings. As people experiencing homelessness are more likely to avoid traditional healthcare settings due to past experiences of exclusion or stigma [53, 54], it is crucial that researchers explicitly recognize the potential impact of this factor on the generalizability of their results. Additionally, where studies compare health or healthcare between cohorts of homeless and housed individuals, special care should be taken to ensure cohorts are sourced in a similar manner (for example, by requiring all participants have at least one eligible healthcare encounter) to avoid spurious associations [55].

Finally, researchers using administrative data must also consider the impact of linkage methodology (both between databases and within databases for longitudinal work) on the representativeness of their cohorts. In our study, only individuals eligible for OHIP were linkable between and within ICES databases, due to the method used to encrypt identifiers [38]. While most (>99%) people in Ontario are eligible for OHIP, certain groups such as refugee claimants (asylum seekers whose claims are not yet adjudicated), recent interprovincial migrants to Ontario, temporary workers with short-term work permits, and individuals without legal status are not covered [26], and therefore are not linkable at ICES. These groups are all believed to be over-represented among people experiencing homelessness, especially refugee claimants who make up a significant proportion of shelter clients across Canada [56] and up to 50% of shelter clients in Toronto [57]. Explicit consideration of these groups and the resulting impact on findings are vital for appropriate interpretation of results.

In conclusion, health administrative data present an invaluable opportunity to conduct health research about people experiencing homelessness in Canada, particularly about

individuals who interact with the healthcare system. The notable increase in sensitivity since 2018 serves as strong evidence that coding mandates can improve capture of housing and other social determinants of health data. Because of this policy change, it is now possible to create reliable cohorts of people experiencing homelessness in Ontario to assess healthcare utilization and outcomes, which in turn facilitates evaluation of policies and programs aimed at addressing persistent health inequities among this population.

### **Acknowledgements**

This research was funded by the Canadian Institutes of Health Research (CIHR) (FDN VR5-173211) in partnership with the Public Health Agency of Canada as part of the COVID-19 Immunity Task Force (CITF). This study was supported by ICES, which is funded by an annual grant from the Ontario Ministry of Health and Ministry of Long-Term Care. The study was completed at the ICES Western site, whose core funding is provided by the Academic Medical Organization of Southwestern Ontario, the Schulich School of Medicine and Dentistry, Western University, and the Lawson Health Research Institute. Parts of this material are based on data and information compiled and provided by the Ontario Ministry of Health, Ministry of Long-term Care and Canadian Institute for Health Information. The analyses, conclusions, opinions, and statements expressed herein are those of the authors and do not reflect those of the funding sources or data providers; no endorsement is intended or should be inferred.

The authors thank the *Ku-gaa-gii pimitizi-win* study's research participants, without whom this study would not have been possible and who participated even while navigating numerous other priorities. We also wish to thank Survey Research Unit staff, including Olivia

Spandier, Alexandra Carasco, Annika Khan, and the entire SRU interview team for undertaking recruitment and data collection; shelter and COVID-19 physical distancing shelter hotel staff for facilitating visits; Dr. Suzanne Stewart from the Waakebiness-Bryce Institute for Indigenous Health, Joe Hester from Anishnawbe Health Toronto and members of MAP's Community Expert Group for participating in the community review of this project; and finally, staff at the MAP Centre for Urban Health Solutions who contributed in numerous ways to the study.

*Ku-gaa-gii pimitizi-win*, which roughly translates in English to "life is always/forever moving", is a spirit name given in ceremony by Elder Dylan Courchene from Anishnawbe Health Toronto. The *Ku-gaa-gii pimitizi-win* study was conducted on Anishnawbe and Haudenosaunee territories. The authors acknowledge and pay respects to the Elders and the communities of past, present and future.

## 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. *Lancet* 2014; 384 (9953): 1529-40. doi: 10.1016/S0140-6736(14)61132-6
2. Hwang SW, Aubry T, Palepu A, 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. *Int J Publ Health* 2011; 56: 609-23. doi: 10.1007/s00038-011-0283-3
3. Goering PN, Streiner DL, Adair C, et al. The At Home/Chez Soi trial protocol: a pragmatic, multi-site, randomised controlled trial of a Housing First intervention for homeless individuals with mental illness in five Canadian cities. *BMJ Open* 2011; 1(2):e000323. doi: 10.1136/bmjopen-2011-000323
4. Richard L, Nisenbaum R, Liu M, et al. Ku-gaa-gii pimitizi-win, the COVID-19 cohort study of people experiencing homelessness in Toronto, Canada: a study protocol. *BMJ Open* 2022; 12(8): e063234. doi: 10.1136/bmjopen-2022-063234corr1
5. Stefancic A, Schaefer-McDaniel NJ, Davis AC, et al. Maximizing follow-up of adults with histories of homelessness and psychiatric disabilities. *Eval Program Plann* 2004; 27:433–42. doi: 10.1016/j.evalprogplan.2004.07.006
6. Booth S. Researching health and homelessness: methodological challenges for researchers working with a vulnerable, hard to reach, transient population. *Aust J Prim Health* 1999; 5:76–81. doi: 10.1071/PY99037
7. Phelan JC, Link BG. Who are "the homeless"? Reconsidering the stability and composition of the homeless population. *AJPH* 1999; 89(9):1334-8. doi: 10.2105/ajph.89.9.1334
8. Schneider M, Brisson D, Burnes D. Do we really know how many are homeless?: An analysis of the point-in-time homelessness count. *Families in Society* 2016; 97(4):321-9. doi: 10.1606/1044-3894.2016.97.39
9. Williams JC. "Stand up and be counted": the politics of a homeless enumeration. *Poverty & Public Policy* 2011;3(3):1-27. doi: 10.2202/1944-2858.1172
10. Infrastructure Canada. Everyone Counts 2020-2022: Preliminary Highlights Report. 2023. Available: <https://www.infrastructure.gc.ca/homelessness-sans-abri/reports-rapports/pit-counts-dp-2020-2022-highlights-eng.html#h2.2>
11. Nielsen SF, Hjorthøj CR, Erlangsen A, et al. Psychiatric disorders and mortality among people in homeless shelters in Denmark: a nationwide register-based cohort study. *Lancet* 2011; 377(9784): 2205-14. doi: 10.1016/S0140-6736(11)60747-2

12. Thomas I, Mackie P. A population level study of SARS-CoV-2 prevalence amongst people experiencing homelessness in Wales, UK. *Int J Popul Data Sci* 2020; 5 (4). doi: 10.23889/ijpds.v5i4.1695
13. Wiens K, Nisenbaum R, Sucha E, et al. Does housing improve health care utilization and costs? a longitudinal analysis of health administrative data linked to a cohort of individuals with a history of homelessness. *Med Care* 2021; 59: S110-6. doi: 10.1097/MLR.0000000000001379
14. Richard L, Golding H, Saskin R, et al. Cold-related injuries among patients experiencing homelessness in Toronto: a descriptive analysis of emergency department visits. *CJEM* 2023; 25(8):695-703. doi: 10.1007/s43678-023-00546-7
15. Liu M, Richard L, Campitelli MA, et al. Hospitalizations during the COVID-19 pandemic among recently homeless individuals: a retrospective population-based matched cohort study. *JGIM* 2022; 37(8):2016-25. doi: 10.1007/s11606-022-07506-4
16. Richard L, Booth R, Rayner J, et al. Testing, infection and complication rates of COVID-19 among people with a recent history of homelessness in Ontario, Canada: a retrospective cohort study. *CMAJ Open* 2021; 9(1):E1-9. doi: 10.9778/cmajo.20200287
17. 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. *Can J Diabetes* 2014; 38(3):179-85. doi: 10.1016/j.jcjd.2014.02.030
18. 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; 10(4): e0123204. doi: 10.1371/journal.pone.0123204
19. Benchimol EI, Manuel DG, To T, et al. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin Epidemiol* 2011; 64(8): 821-9. doi: 10.1016/j.jclinepi.2010.10.006
20. Torres JM, Lawlor J, Colvin JD, et al. ICD social codes: an underutilized resource for tracking social needs. *Medical care* 2017; 55 (9): 810-6. doi: 10.1097/MLR.0000000000000764
21. Guo Y, Chen Z, Xu K, et al. International Classification of Diseases, Tenth Revision, Clinical Modification social determinants of health codes are poorly used in electronic health records. *Medicine* 2020; 99 (52): e23818. doi: 10.1097/MD.00000000000023818
22. Truong HP, Luke AA, Hammond G, Wadhera RK, Reidhead M, Maddox KE. Utilization of social determinants of health ICD-10 Z-codes among hospitalized patients in the United States, 2016–2017. *Medical care*. 2020 Dec 1;58(12):1037-43.
23. Richard L, Hwang SW, Forchuk C, et al. Validation study of health administrative data algorithms to identify individuals experiencing homelessness and estimate population prevalence of homelessness in Ontario, Canada. *BMJ Open* 2019; 9(10): e030221. doi: 10.1136/bmjopen-2019-030221

24. Canadian Institute for Health Information (CIHI). Updated ICD-10-CA Coding Direction: Homelessness, and Falls from an Electric Scooter (Escooter), Mobility Scooter, Segway® or Hoverboard. 2021. Available: <https://www.cihi.ca/en/bulletin/updated-icd-10-ca-coding-direction-homelessness-and-falls-from-an-electric-scooter>
25. De Prophetis E, Beck K, Ridgeway D, et al. Trends in hospital coding for people experiencing homelessness in Canada, 2015-2020: a descriptive study. *CMAJ Open* 2023; 11(6): E1188-96. doi: 10.9778/cmajo.20230044
26. Government of Ontario. *Apply for OHIP and get a health card*. No date. Available: [www.ontario.ca/page/apply-ohip-and-get-health-card](http://www.ontario.ca/page/apply-ohip-and-get-health-card). Accessed 2020 Aug 5
27. Government of Canada. Interim Federal Health Program: Who is eligible. No date. Available: <https://www.canada.ca/en/immigration-refugees-citizenship/services/refugees/help-within-canada/health-care/interim-federal-health-program/eligibility.html>. Accessed 2024 Mar 25.
28. Canadian Institute for Health Information. Discharge Abstract Database (DAD) metadata. No date. Available: <https://www.cihi.ca/en/discharge-abstract-database-dad-metadata>. Accessed 2024 May 24.
29. Canadian Institute for Health Information. National Ambulatory Care Reporting System (NACRS) metadata. No date. Available: <https://www.cihi.ca/en/national-ambulatory-care-reporting-system-nacrs-metadata>. Accessed 2024 May 24
30. Canadian Institute for Health Information. Ontario Mental Health Reporting System Metadata. No date. Available: <https://www.cihi.ca/en/ontario-mental-health-reporting-system-metadata>. Accessed 2024 May 24
31. Government of Ontario. Claims History Database. No date. Available: <https://data.ontario.ca/dataset/claims-history-database>. Accessed 2024 May 24.
32. Ontario Ministry of Health and Long-Term Care. Community Health Centres. No date. Available: <https://www.health.gov.on.ca/en/common/system/services/chc/default.aspx>. Accessed 2024 May 24
33. Gershon AS, Wang C, Guan J, et al. Identifying individuals with physician diagnosed COPD in health administrative databases. *COPD: Journal of Chronic Obstructive Pulmonary Disease*. 2009 Jan 1;6(5):388-94.
34. Gershon AS, Wang C, Guan J, et al. Identifying patients with physician-diagnosed asthma in health administrative databases. *Canadian Respiratory Journal*. 2009 Nov 1;16:183-8.

35. Guttman A, Nakhla M, Henderson M, et al. Validation of a health administrative data algorithm for assessing the epidemiology of diabetes in Canadian children. *Pediatric diabetes*. 2010 Mar;11(2):122-8.
36. Schultz SE, Rothwell DM, Chen Z, et al. Identifying cases of congestive heart failure from administrative data: a validation study using primary care patient records. *Chronic diseases and injuries in Canada*. 2013 Jun 1;33(3).
37. Tu K, Campbell NR, Chen ZL, et al. Accuracy of administrative databases in identifying patients with hypertension. *Open medicine*. 2007;1(1):e18.
38. Schull MJ, Azimae M, Marra M, et al. ICES: data, discovery, better health. *Int J Popul Data Sci* 2019; 4(2). doi: 10.23889/ijpds.v4i2.1135
39. Tsemberis S, McHugo G, Williams V, et al. Measuring homelessness and residential stability: The residential time-line follow-back inventory. *J Community Psychol* 2007; 35(1): 29-42. doi: 10.1002/jcop.20132
40. Canadian Observatory on Homelessness. *Canadian Definition of Homelessness*. 2012. Available: <http://homelesshub.ca/homelessdefinition>. Accessed 2024 May 21.
41. Gaetz S, DeJ E, Richter T, et al. The state of homelessness in Canada. Toronto: Canadian Observatory on Homelessness Press, 2016
42. Chan J, Mack DR, Manuel DG, et al. 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; 12(6):e0180338. doi: 10.1371/journal.pone.0180338
43. Ducharme R, Benchimol EI, Deeks SL, et al. Validation of diagnostic codes for intussusception and quantification of childhood intussusception incidence in Ontario, Canada: a population-based study. *J Pediatr* 2013; 163(4):1073-9. doi: 10.1016/j.jpeds.2013.05.034
44. Richard L, Ouedraogo A, Shariff SZ. Identifying homelessness using health administrative data and postal codes. Not available online (But available upon request to first author). Prepared for Public Health Ontario, 2020.
45. Austin PC. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Commun Stat Simul Comput*. 2009; 38: 1228–34. doi: 10.1080/03610910902859574
46. McGee S. Simplifying likelihood ratios. *J Gen Intern Med*. 2002; 17: 647–50. doi: 10.1046/j.1525-1497.2002.10750.x

47. World Health Organization. International Classification of Diseases (ICD) Information Sheet. Available: <https://web.archive.org/web/20121104064222/http://www.who.int/classifications/icd/factsheet/en/>. Archived from the original on November 4, 2012. Accessed 2024 Mar 25.
48. Moorthie S, Peacey V, Evans S, Phillips V, Roman-Urrestarazu A, Brayne C, Lafortune L. A scoping review of approaches to improving quality of data relating to health inequalities. *International Journal of Environmental Research and Public Health*. 2022 Nov 29;19(23):15874.
49. Sandhu S, Liu M, Wadhera RK. Hospitals and health equity—translating measurement into action. *New England Journal of Medicine*. 2022 Dec 29;387(26):2395-7.
50. National Health Service. 2022/2023 priorities and operational planning guidance. V3 22 Feb 2022. Available: <https://www.england.nhs.uk/wp-content/uploads/2022/02/20211223-B1160-2022-23-priorities-and-operational-planning-guidance-v3.2.pdf>. Accessed 2024 March 28.
51. Buzelli L, Dunn P, Scott S, Gottlieb L, Alderwick H. A framework for NHS action on social determinants of health. *The Health Foundation*. 2022 Oct 4. Accessed 2024 Mar 28.
52. Jaakkimainen L, Duchon R, Lix L, et al. Identification of Early Onset Dementia in Population-Based Health Administrative Data: A Validation Study Using Primary Care Electronic Medical Records. *J of Alzheimer's Disease*. 2022; 89(4):1463-72. doi: 10.3233/JAD-220384
53. Reilly J, Ho I, Williamson A. A systematic review of the effect of stigma on the health of people experiencing homelessness. *Health & Social Care in the Community* 2022; 30(6):2128-41. doi: 10.1111/hsc.13884
54. Wen CK, Hudak PL, Hwang SW. Homeless people's perceptions of welcomeness and unwelcomeness in healthcare encounters. *J Gen Intern Med*. 2007; 22: 1011-7. doi: 10.1007/s11606-007-0183-7
55. Westreich D. Berkson's bias, selection bias, and missing data. *Epidemiology*. 2012 Jan 1;23(1):159-64. doi: 10.1097/EDE.0b013e31823b6296
56. Infrastructure Canada. Homelessness Data Snapshot: Homelessness experienced by Newcomers to Canada. 2022. Available: <https://www.infrastructure.gc.ca/homelessness-sans-abri/reports-rapports/data-newcomers-donnees-nouveaux-arrivants-eng.html>. Accessed 2023 Dec 18
57. City of Toronto. Immediate Federal Government Support Required for Refugee Claimants in Toronto. 2023. Available: <https://www.toronto.ca/legdocs/mmis/2023/cc/bgrd/backgroundfile-239033.pdf> Accessed 2023 Dec 18.



Table 1 – Cohort characteristics at the start of the earliest housing episode, by group

	<b><i>Ku-gaa-gii</i> pimitizi-win participants (N=640)</b>	<b>Presumed housed (N=128,000)</b>	<b>Std Diff</b>
Age at start of observation period, N (%)			
<16 years	<=5 (0.5%)	2,624 (2.1%)	14%
16-29 years	89 (13.9%)	28,372 (22.2%)	22%
30-44 years	208 (32.5%)	33,478 (26.2%)	14%
45-59 years	227 (35.5%)	32,178 (25.1%)	23%
60+ years	NR	31,348 (24.5%)	17%
Sex assigned at birth, N (%)			
Male	447 (69.8%)	63,002 (49.2%)	43%
Female	193 (30.2%)	64,998 (50.8%)	43%
Charlson comorbidity index category, N (%)			
No Hospitalizations	510 (79.7%)	118,914 (92.9%)	39%
0	84 (13.1%)	7,018 (5.5%)	27%
1	24 (3.8%)	905 (0.7%)	21%
2+	22 (3.4%)	1,163 (0.9%)	17%
Presence of specified comorbidities, N (%)			
Hypertension	113 (17.7%)	25,176 (19.7%)	5%
Congestive heart failure	15 (2.3%)	1,882 (1.5%)	6%
COPD or asthma	181 (28.3%)	20,235 (15.8%)	30%
Diabetes	78 (12.2%)	12,773 (10.0%)	7%
Chronic kidney disease	8 (1.3%)	327 (0.3%)	12%
Liver disease	32 (5.0%)	1,815 (1.4%)	20%
Mental health or substance use related concerns, N (%)			
Any mental health or substance use concern	262 (40.9%)	5,944 (4.6%)	96%
Substance use disorder	157 (24.5%)	544 (0.4%)	78%
Psychotic disorders (inc. schizophrenia)	64 (10.0%)	401 (0.3%)	45%
Mood/anxiety disorders	112 (17.5%)	4,360 (3.4%)	47%
OCD or other personality disorders	23 (3.6%)	117 (0.1%)	26%
Intentional self-injury	26 (4.1%)	90 (0.1%)	28%
Admissions in the past year, N (%)			
0	510 (79.7%)	118,914 (92.9%)	39%
1	75 (11.7%)	6,915 (5.4%)	23%
2+	55 (8.6%)	2,171 (1.7%)	32%
ED visits in the past year, N (%)			
0	261 (40.8%)	110,766 (86.5%)	108%
1-3	212 (33.1%)	16,154 (12.6%)	5%
4+	167 (26.1%)	1,080 (0.8%)	8%

Outpatient visits <sup>1</sup> in the past year, N (%)				
	0	126 (19.7%)	46,654 (36.4%)	38%
	1-5	109 (17.0%)	33,197 (25.9%)	22%
	6-10	84 (13.1%)	19,389 (15.1%)	6%
	11+	321 (50.2%)	28,760 (22.5%)	60%
% of observation period spent homeless				
	Mean $\pm$ SD	36.1 $\pm$ 40.0	N/A	N/A
	Median (IQR)	17 (0-75)	N/A	N/A
Duration of homelessness episode				
	Mean $\pm$ SD	207.6 $\pm$ 445.5	N/A	N/A
	Median (IQR)	52 (0-293)	N/A	N/A

Std Diff=Standardized difference; SD=Standard deviation; IQR=Interquartile range; COPD=Chronic obstructive pulmonary disease; OCD=Obsessive compulsive disorder; ED=Emergency department; <=5 and NR are unreportable to protect participant and patient privacy

<sup>1</sup> Visits occurring in an outpatient setting as recorded in OHIP or CHC databases

Table 2 – Validity of case definitions identifying homelessness in health administrative data

Case definition	TP	FP	FN	TN	Sens %	Spec %	PPV %	NPV %	LR+
<b><i>Within the same calendar year</i></b>									
Any indicator (inc postal codes) during any healthcare encounter	264	426	216	128,257	55.0	99.7	38.3	99.8	166.1
Any indicator (exc postal codes) during any healthcare encounter	246	408	234	128,275	51.3	99.7	37.6	99.8	161.6
ICD-10-CA code 'Z59.0' during any healthcare encounter	242	358	238	128,325	50.4	99.7	40.3	99.8	181.2
Any indicator (inc postal codes) during any hospital-based healthcare encounter	253	365	227	128,318	52.7	99.7	40.9	99.8	185.8
Any indicator (exc postal codes) during any hospital-based healthcare encounter	235	347	245	128,336	49.0	99.7	40.4	99.8	181.6
ICD-10-CA code 'Z59.0' during any hospital-based healthcare encounter	231	324	249	128,359	48.1	99.8	41.6	99.8	191.1
Any indicator (inc postal codes) during any inpatient healthcare encounter	73	138	407	128,545	15.2	99.9	34.6	99.7	141.8
Any indicator (exc postal codes) during any inpatient healthcare encounter	72	137	408	128,546	15.0	99.9	34.5	99.7	140.9
ICD-10-CA code 'Z59.0' during any inpatient healthcare encounter	41	70	439	128,613	8.5	99.9	36.9	99.7	157.0
Any indicator (inc postal codes) during any ED visit	244	334	236	128,349	50.8	99.7	42.2	99.8	195.9
Any indicator (exc postal codes) during any ED visit	225	314	255	128,369	46.9	99.8	41.7	99.8	192.1
ICD-10-CA code 'Z59.0' during any ED visit	225	314	255	128,369	46.9	99.8	41.7	99.8	192.1
ICD-10-CA code 'Z59.0' or 'Z59.1' during any CHC visit	27	81	453	128,602	5.6	99.9	25.0	99.7	89.4
ICD-10-CA code 'Z59.0' during any CHC visit	26	53	454	128,630	5.4	100.0	32.9	99.7	131.5
<b><i>Between episode start and within 90 days of episode end</i></b>									
Any indicator (inc postal codes) during any healthcare encounter	261	503	219	128,180	54.4	99.6	34.2	99.8	139.1
Any indicator (exc postal codes) during any healthcare encounter	243	484	237	128,199	50.6	99.6	33.4	99.8	134.6
ICD-10-CA code 'Z59.0' during any healthcare encounter	237	413	243	128,270	49.4	99.7	36.5	99.8	153.8
Any indicator (inc postal codes) during any hospital-based healthcare encounter	250	416	230	128,267	52.1	99.7	37.5	99.8	161.1

Any indicator (exc postal codes) during any hospital-based healthcare encounter	232	397	248	128,286	48.3	99.7	36.9	99.8	156.7
ICD-10-CA code 'Z59.0' during any hospital-based healthcare encounter	227	370	253	128,313	47.3	99.7	38.0	99.8	164.5
Any indicator (inc postal codes) during any inpatient healthcare encounter	79	159	401	128,524	16.5	99.9	33.2	99.7	133.2
Any indicator (exc postal codes) during any inpatient healthcare encounter	78	158	402	128,525	16.3	99.9	33.1	99.7	132.3
ICD-10-CA code 'Z59.0' during any inpatient healthcare encounter	51	76	429	128,607	10.6	99.9	40.2	99.7	179.9
Any indicator (inc postal codes) during any ED visit	240	377	240	128,306	50.0	99.7	38.9	99.8	170.7
Any indicator (exc postal codes) during any ED visit	220	356	260	128,327	45.8	99.7	38.2	99.8	165.7
ICD-10-CA code 'Z59.0' during any ED visit	220	356	260	128,327	45.8	99.7	38.2	99.8	165.7
ICD-10-CA code 'Z59.0' or 'Z59.1' during any CHC visit	10	27	470	128,656	2.1	100	27.0	99.6	99.3
ICD-10-CA code 'Z59.0' during any CHC visit	7	16	473	128,667	1.5	100	30.4	99.6	117.3
<b><i>Between episode start and within 180 days of episode end</i></b>									
Any indicator (inc postal codes) during any healthcare encounter	284	562	196	128,121	59.2	99.6	33.6	99.8	135.5
Any indicator (exc postal codes) during any healthcare encounter	265	539	215	128,144	55.2	99.6	33.0	99.8	131.8
ICD-10-CA code 'Z59.0' during any healthcare encounter	258	460	222	128,223	53.8	99.6	35.9	99.8	150.4
Any indicator (inc postal codes) during any hospital-based healthcare encounter	273	470	207	128,213	56.9	99.6	36.7	99.8	155.7
<b>***Any indicator (exc postal codes) during any hospital-based healthcare encounter</b>	<b>254</b>	<b>447</b>	<b>226</b>	<b>128,236</b>	<b>52.9</b>	<b>99.7</b>	<b>36.2</b>	<b>99.8</b>	<b>152.3</b>
ICD-10-CA code 'Z59.0' during any hospital-based healthcare encounter	248	417	232	128,266	51.7	99.7	37.3	99.8	159.4
Any indicator (inc postal codes) during any inpatient healthcare encounter	90	181	390	128,502	18.8	99.9	33.2	99.7	133.3
Any indicator (exc postal codes) during any inpatient healthcare encounter	89	181	391	128,502	18.5	99.9	33.0	99.7	131.8
ICD-10-CA code 'Z59.0' during any inpatient healthcare encounter	58	87	422	128,596	12.1	99.9	40.0	99.7	178.7
Any indicator (inc postal codes) during any ED visit	262	424	218	128,259	54.6	99.7	38.2	99.8	165.7
Any indicator (exc postal codes) during any ED visit	241	398	239	128,285	50.2	99.7	37.7	99.8	162.3
ICD-10-CA code 'Z59.0' during any ED visit	241	398	239	128,285	50.2	99.7	37.7	99.8	162.3

ICD-10-CA code 'Z59.0' or 'Z59.1' during any CHC visit	14	51	466	128,632	2.9	100	21.5	99.6	73.6
ICD-10-CA code 'Z59.0' during any CHC visit	13	29	467	128,654	2.7	100	31.0	99.6	120.2

TP=True positive; TN=True negative; FP=False positive; FN=False negative; Sens=Sensitivity; Spec=Specificity; LR+=Positive likelihood ratio; \*\*\*represents the optimal case definition

Journal Pre-proof

Table 3: Validity of case definitions identifying homelessness in health administrative data, restricted to housing episodes having at least one eligible healthcare encounter ('people experiencing homelessness who use healthcare services')

Case definition	TP	FP	FN	TN	Sens %	Spec %	PPV %	NPV %	LR+
<b><i>Within the same calendar year</i></b>									
Any indicator (inc postal codes) during any healthcare encounter	217	346	64	24,375	77.2	98.6	38.5	99.7	55.2
Any indicator (exc postal codes) during any healthcare encounter	202	333	79	24,388	71.9	98.7	37.8	99.7	53.4
ICD-10-CA code 'Z59.0' during any healthcare encounter	199	289	82	24,432	70.8	98.8	40.8	99.7	60.6
Any indicator (inc postal codes) during any hospital-based healthcare encounter	207	289	74	24,432	73.7	98.8	41.7	99.7	63.0
Any indicator (exc postal codes) during any hospital-based healthcare encounter	192	276	89	24,445	68.3	98.9	41.0	99.6	61.2
ICD-10-CA code 'Z59.0' during any hospital-based healthcare encounter	189	259	92	24,462	67.3	99.0	42.2	99.6	64.2
Any indicator (inc postal codes) during any inpatient healthcare encounter	59	114	222	24,607	21.0	99.5	34.1	99.1	45.5
Any indicator (exc postal codes) during any inpatient healthcare encounter	58	113	223	24,608	20.6	99.5	33.9	99.1	45.2
ICD-10-CA code 'Z59.0' during any inpatient healthcare encounter	34	59	247	24,662	12.1	99.8	36.6	99.0	50.7
Any indicator (inc postal codes) during any ED visit	199	266	82	24,455	70.8	98.9	42.8	99.7	65.8
Any indicator (exc postal codes) during any ED visit	184	252	97	24,469	65.5	99.0	42.2	99.6	64.2
ICD-10-CA code 'Z59.0' during any ED visit	184	252	97	24,469	65.5	99.0	42.2	99.6	64.2
ICD-10-CA code 'Z59.0' or 'Z59.1' during any CHC visit	24	74	257	24,647	8.5	99.7	24.5	99.0	28.5
ICD-10-CA code 'Z59.0' during any CHC visit	23	46	258	24,675	8.2	99.8	33.3	99.0	44.0
<b><i>Between episode start and within 90 days of episode end</i></b>									
Any indicator (inc postal codes) during any healthcare encounter	231	454	50	24,267	82.2	98.2	33.7	99.8	44.8
Any indicator (exc postal codes) during any healthcare encounter	215	437	66	24,284	76.5	98.2	33.0	99.7	43.3
ICD-10-CA code 'Z59.0' during any healthcare encounter	210	372	71	24,349	74.7	98.5	36.1	99.7	49.7

Any indicator (inc postal codes) during any hospital-based healthcare encounter	220	369	61	24,352	78.3	98.5	37.4	99.8	52.5
Any indicator (exc postal codes) during any hospital-based healthcare encounter	204	352	77	24,369	72.6	98.6	36.7	99.7	51.0
ICD-10-CA code 'Z59.0' during any hospital-based healthcare encounter	200	331	81	24,390	71.2	98.7	37.7	99.7	53.2
Any indicator (inc postal codes) during any inpatient healthcare encounter	69	144	212	24,577	24.6	99.4	32.4	99.1	42.2
Any indicator (exc postal codes) during any inpatient healthcare encounter	68	143	213	24,578	24.2	99.4	32.2	99.1	41.8
ICD-10-CA code 'Z59.0' during any inpatient healthcare encounter	48	72	233	24,649	17.1	99.7	40.0	99.1	58.7
Any indicator (inc postal codes) during any ED visit	211	337	70	24,384	75.1	98.6	38.5	99.7	55.1
Any indicator (exc postal codes) during any ED visit	194	319	87	24,402	69.0	98.7	37.8	99.6	53.5
ICD-10-CA code 'Z59.0' during any ED visit	194	319	87	24,402	69.0	98.7	37.8	99.6	53.5
ICD-10-CA code 'Z59.0' or 'Z59.1' during any CHC visit	10	27	271	24,694	3.6	99.9	27.0	98.9	32.6
ICD-10-CA code 'Z59.0' during any CHC visit	7	16	274	24,705	2.5	99.9	30.4	98.9	38.5
<b><i>Between episode start and within 180 days of episode end</i></b>									
Any indicator (inc postal codes) during any healthcare encounter	237	482	44	24,239	84.3	98.1	33.0	99.8	43.3
Any indicator (exc postal codes) during any healthcare encounter	221	464	60	24,257	78.6	98.1	32.3	99.8	41.9
ICD-10-CA code 'Z59.0' during any healthcare encounter	215	395	66	24,326	76.5	98.4	35.2	99.7	47.9
Any indicator (inc postal codes) during any hospital-based healthcare encounter	227	395	54	24,326	80.8	98.4	36.5	99.8	50.6
Any indicator (exc postal codes) during any hospital-based healthcare encounter	211	377	70	24,344	75.1	98.5	35.9	99.7	49.2
ICD-10-CA code 'Z59.0' during any hospital-based healthcare encounter	206	356	75	24,365	73.3	98.6	36.7	99.7	50.9
Any indicator (inc postal codes) during any inpatient healthcare encounter	76	156	205	24,565	27.0	99.4	32.8	99.2	42.9
Any indicator (exc postal codes) during any inpatient healthcare encounter	75	156	206	24,565	26.7	99.4	32.5	99.2	42.3
ICD-10-CA code 'Z59.0' during any inpatient healthcare encounter	51	79	230	24,642	18.1	99.7	39.2	99.1	56.8
Any indicator (inc postal codes) during any ED visit	217	362	64	24,359	77.2	98.5	37.5	99.7	52.7

Any indicator (exc postal codes) during any ED visit	200	342	81	24,379	71.2	98.6	36.9	99.7	51.4
ICD-10-CA code 'Z59.0' during any ED visit	200	342	81	24,379	71.2	98.6	36.9	99.7	51.4
ICD-10-CA code 'Z59.0' or 'Z59.1' during any CHC visit	14	50	267	24,671	5.0	99.8	21.9	98.9	24.6
ICD-10-CA code 'Z59.0' during any CHC visit	237	482	44	24,239	84.3	98.1	33.0	99.8	43.3

TP=True positive; TN=True negative; FP=False positive; FN=False negative; Sens=Sensitivity; Spec=Specificity; LR+=Positive likelihood ratio; <=5 and NR are unreportable to protect participant and patient privacy



Table 4 – Comparison of characteristics at start of housing episode (those with at least one eligible healthcare encounter), by identification using the ‘optimal’ health administrative data case definition

Characteristics	True Positive (TP) n=211	False Positive (FP) n=377	True Negative (TN) n=24,344	False Negative (FN) n=70	P value TP vs FP	P value TP vs FN
Age, N (%)						
< 29 yrs old	23 (10.9%)	92 (24.4%)	5,593 (23.0%)	14 (20.0%)	0.001	0.18
30-44 yrs	64 (30.3%)	106 (28.1%)	6,241 (25.6%)	21 (30.0%)		
45-59 yrs	85 (40.3%)	123 (32.6%)	5,574 (22.9%)	21 (30.0%)		
60+ yrs	39 (18.5%)	56 (14.9%)	6,936 (28.5%)	14 (20.0%)		
Sex assigned at birth, N (%)						
Female	47 (22.3%)	103 (27.3%)	13,478 (55.4%)	24 (34.3%)	0.18	0.045
Male	164 (77.7%)	274 (72.7%)	10,866 (44.6%)	46 (65.7%)		
Charlson comorbidity index category, N (%)					0.73	0.101
No Hospitalizations	146 (69.2%)	273 (72.4%)	20,858 (85.7%)	59 (84.3%)		
0	43 (20.4%)	66 (17.5%)	2,498 (10.3%)	8 (11.4%)		
1	NR	19 (5.0%)	396 (1.6%)	<=5		
2+	<=5	19 (5.0%)	592 (2.4%)	<=5		
Presence of specified comorbidities, N (%)						
Hypertension	47 (22.3%)	88 (23.3%)	7,020 (28.8%)	13 (18.6%)	0.77	0.51
Congestive heart failure	6 (2.8%)	12 (3.2%)	778 (3.2%)	<=5	0.82	0.99
COPD or asthma	77 (36.5%)	117 (31.0%)	5,680 (23.3%)	21 (30.0%)	0.18	0.32
Diabetes	31 (14.7%)	52 (13.8%)	3,743 (15.4%)	9 (12.9%)	0.76	0.70
Chronic kidney disease	<=5	8 (2.1%)	198 (0.8%)	<=5	0.55	0.99
Liver disease	NR	33 (8.8%)	543 (2.2%)	<=5	0.18	0.65
Mental health or substance use related concerns						
Any mental health/substance use concern	129 (61.1%)	210 (55.7%)	2,242 (9.2%)	25 (35.7%)	0.20	<.001
Substance use disorder	93 (44.1%)	141 (37.4%)	257 (1.1%)	13 (18.6%)	0.11	<.001

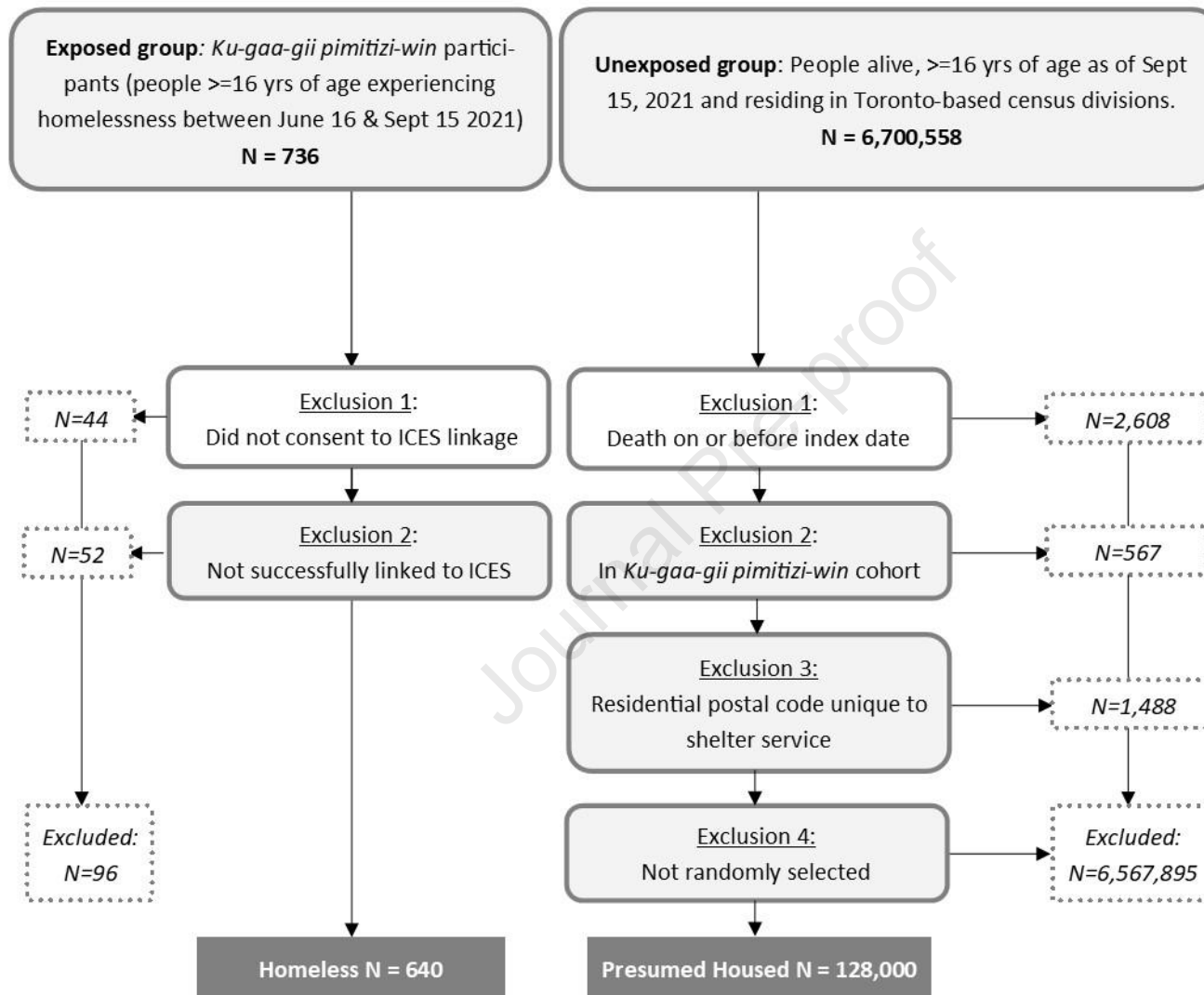
Psychotic disorders (inc. schizophrenia)	34 (16.1%)	51 (13.5%)	168 (0.7%)	6 (8.6%)	0.39	0.12
Mood/anxiety disorders	55 (26.1%)	97 (25.7%)	1,610 (6.6%)	11 (15.7%)	0.93	0.08
OCD or other personality disorders	NR	26 (6.9%)	56 (0.2%)	<=5	0.47	0.04
Intentional self-injury	NR	26 (6.9%)	49 (0.2%)	<=5	0.47	0.04
Admissions in past year, N (%)						
0	146 (69.2%)	273 (72.4%)	20,858 (85.7%)	59 (84.3%)	0.66	0.021
1	37 (17.5%)	56 (14.9%)	2,476 (10.2%)	9 (12.9%)		
2 or more	NR	48 (12.7%)	1,010 (4.1%)	<=5		
ED visits in past year, N (%)						
0	53 (25.1%)	101 (26.8%)	17,069 (70.1%)	33 (47.1%)	0.44	<.001
1-3	60 (28.4%)	121 (32.1%)	6,536 (26.8%)	22 (31.4%)		
4 or more	98 (46.4%)	155 (41.1%)	739 (3.0%)	15 (21.4%)		
Outpatient visits <sup>1</sup> in past year, N (%)						
0	19 (9.0%)	34 (9.0%)	2,606 (10.7%)	11 (15.7%)	0.96	0.13
1-5	29 (13.7%)	53 (14.1%)	6,223 (25.6%)	10 (14.3%)		
6-10	25 (11.8%)	50 (13.3%)	4,828 (19.8%)	13 (18.6%)		
11 or more	138 (65.4%)	240 (63.7%)	10,687 (43.9%)	36 (51.4%)		
% of observation period spent homeless, <i>Ku-gaa-gii pimitizi-win</i> participants only						
Mean ± SD	66.4 ± 31.0	24.9 ± 29.7	16.21 ± 25.0	65.8 ± 30.9	<.001	0.88
Median (IQR)	69 (41-100)	6 (0-47)	0 (0-26)	70 (40-100)	<.001	0.73
Duration of homelessness episode, <i>Ku-gaa-gii pimitizi-win</i> participants only						
Mean ± SD	391.4 ± 671.8	89.6 ± 167.1	130.6 ± 442.7	354.9 ± 271.8	<.001	0.66
Median (IQR)	201 (113-461)	27 (0-128)	0 (0-138)	330 (166-478)	<.001	0.05

TP=True positive; TN=True negative; FP=False positive; FN=False negative; COPD=Chronic obstructive pulmonary disease; OCD=Obsessive compulsive disorder; ED=Emergency department;

<sup>1</sup> Visits occurring in an outpatient setting as recorded in OHIP or CHC databases

<=5 and NR are unreportable to protect participant and patient privacy

Figure 1 – Cohort Build



**Research Support**

This research received no external financial or non-financial support.

**Relationships**

There are no additional relationships to disclose.

**Patents and Intellectual Property**

There are no patents to disclose.

**Other Activities**

There are no additional activities to disclose.

Journal Pre-proof

## Highlights

- Homelessness status in Canadian healthcare data historically had low sensitivity
- Beginning April 2018, Canadian hospitals must code homelessness where on the chart
- Case definitions now have moderate to good sensitivity, doubling from before 2018
- Future potential improvements include updating status after homelessness ends
- Canadian healthcare data is an important resource for studies about this population

Journal Pre-proof