BMJ Open

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

Lucie Richard ,[1] Stephen W Hwang,[2] Cheryl Forchuk,[3] Rosane Nisenbaum,[2,4] Kristin Clemens,[5] Kathryn Wiens,[4] Richard Booth,[3] Mahmoud Azimaee,[6] Salimah Z Shariff[1]

To cite: 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**:e030221. doi:10.1136/bmjopen-2019-030221

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136bmjopen-2019-030221).

Received 05 March 2019
Revised 01 August 2019
Accepted 09 September 2019

Check for updates

© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Lucie Richard;
lucie.richard@ices.on.ca

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 values and positive likelihood ratios of 30 case ascertainment algorithms for detecting homelessness using up to eight health service 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.

Strengths and limitations of this study

- ► 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 (eg, First Nations on reserves, individuals newly arrived to Ontario) were 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 healthcare 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 criticised 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 healthcare may offer an alternate avenue to measure and track individuals experiencing homelessness. In particular, several administrative databases such as those for hospital services are standardised nationwide, allowing for population-level tracking of health and healthcare delivery of Canadians.[16] Health administrative data are already widely used in Canada for population surveillance of health conditions such as diabetes, asthma and ischaemic heart disease,[17–21] permitting counts of the population at any point in time as well as tracking changes in group demographics, health status, healthcare trajectories and gaps in care.[22–24] Currently, the utility of these data in tracking social determinants of health, such as homelessness, is 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 analysed 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 authorises 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.

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 data set publicly available, access to the data may be granted to those who meet prespecified criteria for confidential access, available at www.ices.on.ca/DAS. The full data set creation plan and underlying analytic code detailing all analysis procedures are available from the authors on request, understanding that computer programs rely on 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 organised into consecutive self-reported housing episodes, ranging from an earliest date of 31 January 2007 to a latest date of 14 March 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 categorised 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 hospitalisation 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 categorised 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') (online supplementary table 1). We assessed housing status indicators present in: the Discharge Abstract Database, the National Ambulatory Care Reporting System, the Ontario Mental Health Reporting System, the Home Care Database, the Resident Assessment Instrument Contact Assessment Database, the National Rehabilitation Reporting System and the Canadian Organ Replacement Registry. 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 online supplementary 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 from 0 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 healthcare 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 online supplementary table 3) included the ICES RPDB, Ontario Health Insurance Physicians claims database, the Immigration, Refugee and Citizenship Canada Permanent Resident Database and several ICES-derived population surveillance data sets, including: the chronic obstructive pulmonary disease,[39] Ontario Diabetes Dataset,[40] congestive heart failure[41] and Ontario HIV[42] derived cohorts.

Statistical analysis

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

We then applied the optimal annual housing experience algorithm to identify Ontarians experiencing homelessness in each of the 2007–2016 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 V.9.4.[44]

RESULTS

Cohort

We identified 686 eligible HHiT participants (6948 housing episodes, 3443 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 (IQR: 29–181 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 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 service databases increased sensitivity, while marginally decreasing specificity. The use of all databases, as opposed to only CIHI databases, resulted in negligible gains in sensitivity. LR+ 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 maximised sensitivity, specificity and PPV. 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)

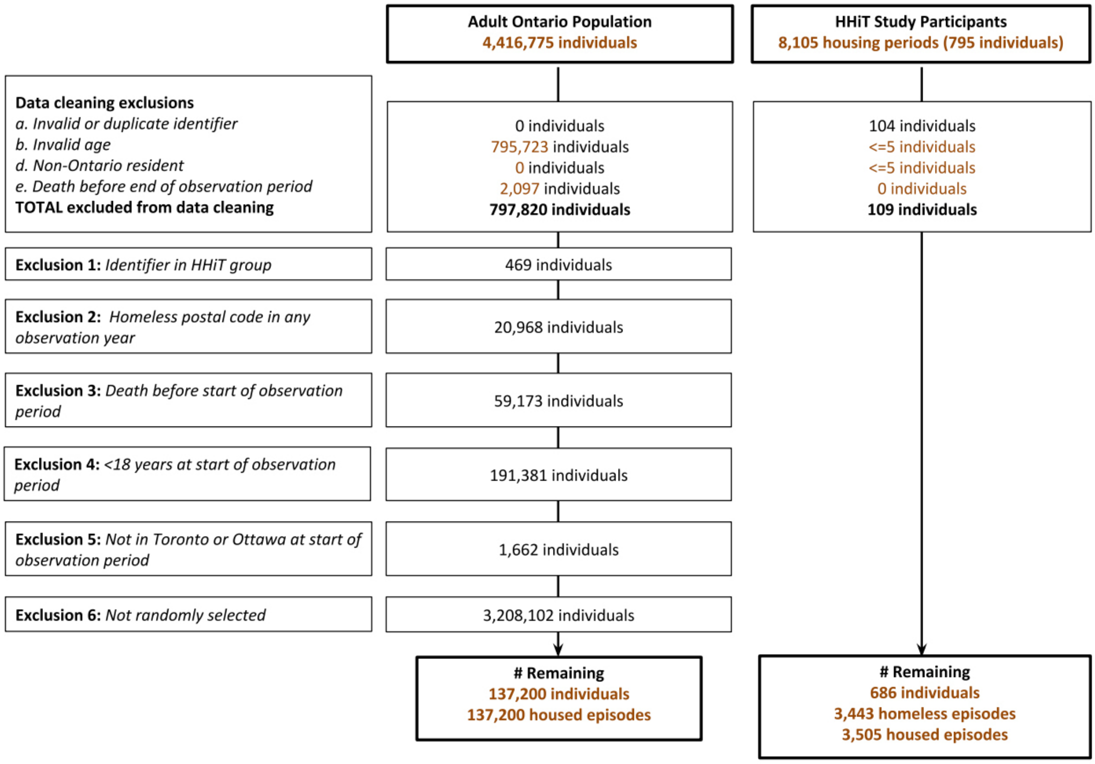

Figure 1 Cohort build. HHiT, Health and Housing in Transition study.

(online supplementary table 6A). Absence of a healthcare encounter during the reference period accounted for 64.5% (n=1825) 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). LR+ 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 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 maximised 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) (online supplementary 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 and 65. One in 10 were recent immigrants, about one-third resided in Metropolitan Toronto, and a large proportion recently received mental or substance use-related healthcare (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 1 year (online supplementary 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% to 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

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	1152 (0.8)	≤5	NR	<0.001
1–3 years	2381 (1.7)	≤5	NR	
4–10 years	9606 (7.0)	9 (1.3)	9597 (7.0)	
Over 10 years	19 518 (14.2)	33 (4.8)	19 485 (14.2)	
Refugee status, n (%)	5907 (4.3)	18 (2.6)	5889 (4.3)	<0.001
Congestive heart failure, n (%)	2186 (1.6)	6 (0.9)	2180 (1.6)	0.14
Chronic obstructive pulmonary disease, n (%)	6627 (4.8)	91 (13.3)	6536 (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*, n (%)	2431 (1.8)	20 (2.9)	2411 (1.8)	0.02
Chronic liver disease*, n (%)	2939 (2.1)	87 (12.7)	2852 (2.1)	<0.001
Mental health-related care†, 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	1640 (1.2)	204 (29.7)	1436 (1.0)	<0.001
Charlson Comorbidity Index, n (%)				
0	7866 (5.7)	86 (12.5)	7780 (5.7)	<0.001
1	1589 (1.2)	25 (3.6)	1564 (1.1)	
2+	2476 (1.8)	25 (3.6)	2451 (1.8)	
No hospitalisation	125 955 (91.3)	550 (80.2)	125 405 (91.4)	
Primary care visits†, mean (SD)	13.0 (17.5)	21.1 (31.7)	12.9 (17.4)	<0.001
Emergency department visits†, mean (SD)	1.6 (1.7)	3.9 (5.1)	1.6 (1.5)	<0.001
Hospitalisations†, mean (SD)	1.3 (0.9)	1.7 (1.4)	1.3 (0.9)	<0.001

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.
*Within the past 3 years.
†Occurring in the past year.
HHiT, Health and Housing in Transition study; n/a, not applicable; NR, not reportable, due to associated small cell suppression; NS, not significant.

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 LR+.

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, 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 healthcare services

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

Reference standard definition: housing episode (n=144 148 overall, with 3443 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	3071	140 177	10.8 (9.8 to 11.9)	99.6 (99.6 to 99.7)	41.3 (38.2 to 44.6)	97.9 (97.8 to 97.9)	28.8
1 indicator+/−15 days	482	591	2961	140 114	14.0 (12.9 to 15.2)	99.6 (99.5 to 99.6)	44.9 (42.0 to 47.9)	97.9 (97.9 to 98.0)	33.3
1 indicator+/−45 days	619	665	2824	140 040	18.0 (16.7 to 19.3)	99.5 (99.5 to 99.6)	48.2 (45.5 to 50.9)	98.0 (98.0 to 98.1)	38.0
1 indicator+/−90 days	718	765	2725	139 940	20.9 (19.5 to 22.2)	99.5 (99.4 to 99.5)	48.4 (45.9 to 51.0)	98.1 (98.0 to 98.2)	38.4
1 indicator+/−180 days	861	897	2582	139 808	25.0 (23.6 to 26.5)	99.4 (99.3 to 99.4)	49.0 (46.6 to 51.3)	98.2 (98.1 to 98.3)	39.2
1 indicator OR postal code+/−0 days	450	679	2993	140 026	13.1 (12.0 to 14.2)	99.5 (99.5 to 99.6)	39.9 (37.0 to 42.7)	97.9 (97.8 to 98.0)	27.1
1 indicator OR postal code+/−15 days	572	758	2871	139 947	16.6 (15.4 to 17.9)	99.5 (99.4 to 99.5)	43.0 (40.4 to 45.7)	98.0 (97.9 to 98.1)	30.8
1 indicator OR postal code+/−45 days	714	845	2729	139 860	20.7 (19.4 to 22.1)	99.4 (99.4 to 99.4)	45.8 (43.3 to 48.3)	98.1 (98.0 to 98.2)	34.5
1 indicator OR postal code+/−90 days	824	967	2619	139 738	23.9 (22.5 to 25.4)	99.3 (99.3 to 99.4)	46.0 (43.7 to 48.3)	98.2 (98.1 to 98.2)	34.8
1 indicator OR postal code+/−180 days	994	1135	2449	139 570	28.9 (27.4 to 30.4)	99.2 (99.1 to 99.2)	46.7 (44.6 to 48.8)	98.3 (98.2 to 98.3)	35.8
1 CIHI indicator+/−0 days	368	466	3075	140 239	10.7 (9.7 to 11.8)	99.7 (99.6 to 99.7)	44.1 (40.8 to 47.5)	97.9 (97.8 to 97.9)	36.9
1 CIHI indicator+/−15 days	477	528	2966	140 177	13.9 (12.7 to 15.0)	99.6 (99.6 to 99.7)	47.5 (44.4 to 50.6)	97.9 (97.9 to 98.0)	39.6
1 CIHI indicator+/−45 days	**613**	**595**	**2830**	**140 110**	**17.8 (16.6 to 19.1)**	**99.6 (99.5 to 99.6)**	**50.7 (47.9 to 53.6)**	**98.0 (97.9 to 98.1)**	**42.0**
1 CIHI indicator+/−90 days	710	693	2733	140 012	20.6 (19.3 to 22.0)	99.5 (99.5 to 99.5)	50.6 (48.0 to 53.2)	98.1 (98.0 to 98.2)	41.7
1 CIHI indicator+/−180 days	852	822	2591	139 883	24.8 (23.3 to 26.2)	99.4 (99.4 to 99.5)	50.9 (48.5 to 53.3)	98.2 (98.1 to 98.3)	41.8
1 CIHI indicator OR postal code+/−0 days	444	575	2999	140 130	12.9 (11.8 to 14.1)	99.6 (99.6 to 99.6)	43.6 (40.6 to 46.6)	97.9 (97.8 to 98.0)	32.3
1 CIHI indicator OR postal code+/−15 days	566	652	2877	140 053	16.4 (15.2 to 17.7)	99.5 (99.5 to 99.6)	46.5 (43.7 to 49.3)	98.0 (97.9 to 98.1)	36.9
1 CIHI indicator OR postal code+/−45 days	707	734	2736	139 971	20.5 (19.2 to 21.9)	99.5 (99.4 to 99.5)	49.1 (46.5 to 51.6)	98.1 (98.0 to 98.2)	42.1
1 CIHI indicator OR postal code+/−90 days	817	852	2626	139 853	23.7 (22.3 to 25.2)	99.4 (99.4 to 99.4)	49.0 (46.6 to 51.3)	98.2 (98.1 to 98.2)	41.9
1 CIHI indicator OR postal code+/−180 days	985	1017	2458	139 688	28.6 (27.1 to 30.1)	99.3 (99.2 to 99.3)	49.2 (47.0 to 51.4)	98.3 (98.2 to 98.3)	42.4

Reference standard definition: annual housing experience (n=491 213 calendar years overall, with 2290 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	1861	488 589	18.7 (17.2 to 20.4)	99.9 (99.9 to 99.9)	56.2 (52.7 to 59.7)	99.6 (99.6 to 99.6)	274.2
1 indicator+/−15 days	454	352	1836	488 571	19.8 (18.2 to 21.5)	99.9 (99.9 to 99.9)	56.3 (52.9 to 59.7)	99.6 (99.6 to 99.6)	275.4
1 indicator+/−45 days	487	406	1803	488 517	21.3 (19.6 to 23.0)	99.9 (99.9 to 99.9)	54.5 (51.3 to 57.8)	99.6 (99.6 to 99.6)	256.1
1 indicator+/−90 days	529	472	1761	488 451	23.1 (21.4 to 24.9)	99.9 (99.9 to 99.9)	52.8 (49.8 to 55.9)	99.6 (99.6 to 99.7)	239.3
1 indicator+/−180 days	590	588	1700	488 335	25.8 (24.0 to 27.6)	99.9 (99.9 to 99.9)	50.1 (47.2 to 52.9)	99.7 (99.6 to 99.7)	214.2
1 indicator OR postal code+/−0 days	512	433	1778	488 490	22.4 (20.7 to 24.1)	99.9 (99.9 to 99.9)	54.2 (51.0 to 57.3)	99.6 (99.6 to 99.7)	252.5
1 indicator OR postal code+/−15 days	543	458	1747	488 465	23.7 (22.0 to 25.5)	99.9 (99.9 to 99.9)	54.2 (51.1 to 57.3)	99.6 (99.6 to 99.7)	253.1
1 indicator OR postal code+/−45 days	581	525	1709	488 398	25.4 (23.6 to 27.2)	99.9 (99.9 to 99.9)	52.5 (49.6 to 55.5)	99.7 (99.6 to 99.7)	236.3
1 indicator OR postal code+/−90 days	629	610	1661	488 313	27.5 (25.7 to 29.3)	99.9 (99.9 to 99.9)	50.8 (48.0 to 53.5)	99.7 (99.6 to 99.7)	220.2
1 indicator OR postal code+/−180 days	707	754	1583	488 169	30.9 (29.0 to 32.8)	99.9 (99.8 to 99.9)	48.4 (45.8 to 51.0)	99.7 (99.7 to 99.7)	200.2
1 indicator+/−0 day OR PSTLYEAR postal code	588	356	1702	488 567	25.7 (23.9 to 27.5)	99.9 (99.9 to 99.9)	62.3 (59.2 to 65.3)	99.7 (99.6 to 99.7)	352.6

Continued

Table 2 Continued

Algorithm definition	TP	FP	FN	TN	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	PPV (%)(95% CI)	NPV (%)(95% CI)	LR+
Reference standard definition: annual housing experience (n=491 213 calendar years overall, with 2290 homeless years)									
1 indicator+/−15 days OR PSTLYEAR postal code	706	402	1584	488 521	30.8 (29.0 to 32.8)	99.9 (99.9 to 99.9)	63.7 (60.8 to 66.5)	99.7 (99.7 to 99.7)	375.0
1 indicator+/−45 days OR PSTLYEAR postal code	734	452	1556	488 471	32.1 (30.2 to 34.0)	99.9 (99.9 to 99.9)	61.9 (59.1 to 64.6)	99.7 (99.7 to 99.7)	346.7
1 indicator+/−90 days OR PSTLYEAR postal code	766	518	1524	488 405	33.4 (31.5 to 35.4)	99.9 (99.9 to 99.9)	59.7 (56.9 to 62.3)	99.7 (99.7 to 99.7)	315.7
1 indicator+/−180 days OR PSTLYEAR postal code	816	633	1474	488 290	35.6 (33.7 to 37.6)	99.9 (99.9 to 99.9)	56.3 (53.7 to 58.8)	99.7 (99.7 to 99.7)	275.2
1 CIHI indicator+/−0 days	423	300	1867	488 623	18.5 (16.9 to 20.1)	99.9 (99.9 to 99.9)	58.5 (54.9 to 62.0)	99.6 (99.6 to 99.6)	301.0
1 CIHI indicator+/−15 days	**448**	**315**	**1842**	**488 608**	**19.6 (18.0 to 21.2)**	**99.9 (99.9 to 99.9)**	**58.7 (55.2 to 62.2)**	**99.6 (99.6 to 99.6)**	**303.6**
1 CIHI indicator+/−45 days	480	358	1810	488 565	21.0 (19.3 to 22.7)	99.9 (99.9 to 99.9)	57.3 (53.9 to 60.6)	99.6 (99.699.6)	286.3
1 CIHI indicator+/−90 days	521	405	1769	488 518	22.8 (21.1 to 24.5)	99.9 (99.9 to 99.9)	56.3 (53.0 to 59.4)	99.6 (99.6 to 99.7)	274.7
1 CIHI indicator+/−180 days	581	519	1709	488 404	25.4 (23.6 to 27.2)	99.9 (99.9 to 99.9)	52.8 (49.9 to 55.8)	99.7 (99.6 to 99.7)	239.0
1 CIHI indicator OR postal code+/−0 days	508	370	1782	488 553	22.2 (20.5 to 23.9)	99.9 (99.9 to 99.9)	57.9 (54.6 to 61.1)	99.6 (99.6 to 99.7)	293.1
1 CIHI indicator OR postal code+/−15 days	539	390	1751	488 533	23.5 (21.8 to 25.3)	99.9 (99.9 to 99.9)	58.0 (54.8 to 61.2)	99.6 (99.6 to 99.7)	295.1
1 CIHI indicator OR postal code+/−45 days	576	442	1714	488 481	25.2 (23.4 to 27.0)	99.9 (99.9 to 99.9)	56.6 (53.5 to 59.6)	99.7 (99.6 to 99.7)	278.2
1 CIHI indicator OR postal code+/−90 days	622	502	1668	488 421	27.2 (25.4 to 29.0)	99.9 (99.9 to 99.9)	55.3 (52.4 to 58.2)	99.7 (99.6 to 99.7)	264.5
1 CIHI indicator OR postal code+/−180 days	699	634	1591	488 289	30.5 (28.7 to 32.4)	99.9 (99.9 to 99.9)	52.4 (49.8 to 55.1)	99.7 (99.7 to 99.7)	235.4
1 CIHI indicator+/-0 day OR PSTLYEAR postal code	583	322	1707	488 601	25.5 (23.7 to 27.3)	99.9 (99.9 to 99.9)	64.4 (61.2 to 67.5)	99.7 (99.6 to 99.7)	386.6
1 CIHI indicator+/−15 days OR PSTLYEAR postal code	701	365	1589	488 558	30.6 (28.8 to 32.5)	99.9 (99.9 to 99.9)	65.8 (62.9 to 68.5)	99.7 (99.7 to 99.7)	410.0
1 CIHI indicator+/−45 days OR PSTLYEAR postal code	728	404	1562	488 519	31.8 (29.9 to 33.7)	99.9 (99.9 to 99.9)	64.3 (61.5 to 67.0)	99.7 (99.7 to 99.7)	384.7
1 CIHI indicator+/−90 days OR PSTLYEAR postal code	760	451	1530	488 472	33.2 (31.3 to 35.1)	99.9 (99.9 to 99.9)	62.8 (60.0 to 65.4)	99.7 (99.7 to 99.7)	359.8
1 CIHI indicator+/−180 days OR PSTLYEAR postal code	809	564	1481	488 359	35.3 (33.4 to 37.3)	99.9 (99.9 to 99.9)	58.9 (56.3 to 61.5)	99.7 (99.7 to 99.7)	306.2

Bold lines indicate optimal case algorithm definitions.
CIHI, Discharge Abstract Database, National Ambulatory Care Reporting System or Ontario Mental Health Reporting System; FN, false negative (flagged as housed but truly homeless); FP, false positive (flagged as homeless but not truly homeless); LR+, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; PSTLYEAR, ICES PSTLYEAR postal code, indicating the best estimate of an individual's postal code for the year using ICES databases; TN, true negative (flagged as housed and truly housed); TP, true positive (flagged as homeless and truly homeless).

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 (years), n (%)	
18–24	1901 (16.2)
25–34	3498 (29.8)
35–50	3246 (27.7)
51–65	2352 (20.1)
Over 65	734 (6.3)
Female sex, n (%)	3497 (29.8)
City of residence in 2016, n (%)	
Toronto	4299 (36.7)
Ottawa	684 (5.8)
In a rural area, n (%)	667 (5.7)
Recent immigrant, n (%)	1172 (10.0)
Immigrated as refugee, n (%)	366 (3.2)
Charlson Comorbidity Index, n (%)	
0	1825 (15.6)
1	550 (4.7)
2+	465 (4.0)
No hospitalisation	8891 (75.8)
Comorbidities, n (%)	
Congestive heart failure	222 (1.9)
Chronic obstructive pulmonary disease	1258 (10.7)
Diabetes	1233 (10.5)
Chronic kidney disease*	588 (5.0)
Chronic liver disease*	1244 (10.6)
HIV positive	202 (1.7)
Primary care visits, mean (SD)	33.0 (43.6)
Emergency department visits, mean (SD)	5.5 (9.2)
Admissions to hospital, mean (SD)	1.9 (1.7)
Mental health-related care, n (%)	
Psychotic disorders	3014 (25.7)
Non-psychotic disorders	6433 (54.8)
Substance use disorders	4917 (41.9)

*Within the past 3 years.
†Occurring in the past year.
CIHI, Discharge Abstract Database, National Ambulatory Care Reporting System or Ontario Mental Health Reporting System.

during specific housing periods, and therefore could not be identified as such 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 by 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 aged 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 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%.[58] However, this study required the use of location and time-specific shelter address registries, making the methodology challenging to scale or generalise. 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 4-year study period. Our study recognised changes in housing status and deliberately included individuals who may not have used healthcare, in order to estimate the algorithm's ability to count the complete homeless population.

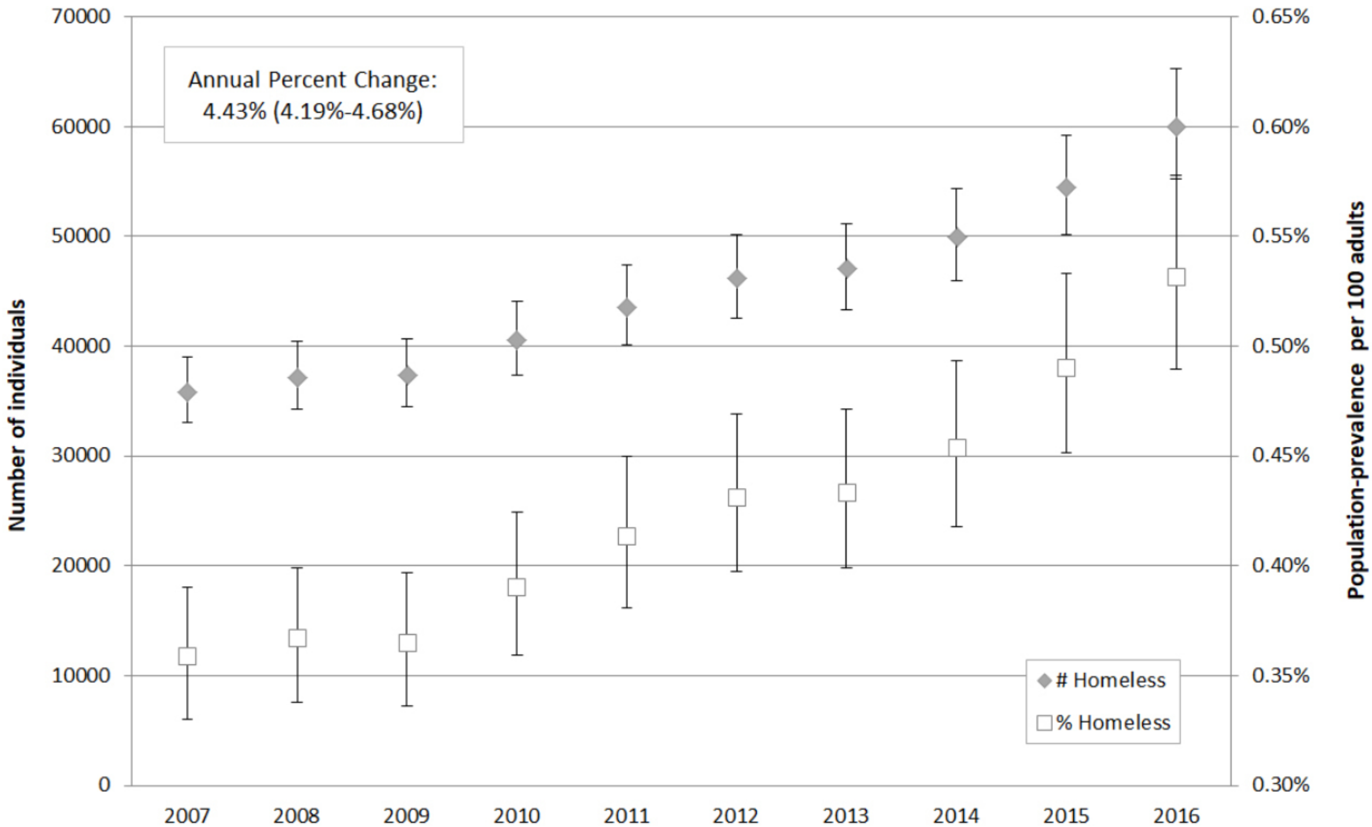

Figure 2 Estimated number of individuals and population prevalence (per 100 adults) experiencing homelessness in Ontario from 2007 to 2016 using the optimal annual housing experience case ascertainment algorithm (any CIHI indicator +/−15 days), with 95% CIs, correcting for sensitivity. Annual percentage change with CI was calculated using a Poisson regression. CIHI, Canadian Institute for Health Information.

We readily acknowledge some limitations to this validation. First, because it was conducted in a universal, single-payer healthcare system, this validation's applicability is limited to jurisdictions with similar settings who collect similar types of standardised information. Even so, before implementation policymakers 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 healthcare services during their housing period and were thus automatically considered housed. Other jurisdictions having access to housing status variables in standardised 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 realise improved algorithm performance through increased opportunities for encounters during a homeless episode.

Third, we could only validate homelessness among adults 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 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. On 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 recoding such individuals on the basis of the same administrative data we are attempting to 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 programme evaluation are high, necessarily reducing funding for programme implementation. Overall, our algorithms present, despite their low sensitivity, important potential cost-saving 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 healthcare encounters may result in increased identification of homelessness in Ontario.

Author affiliations
[1]ICES Western, London, Ontario, Canada
[2]St Michael's Hospital, Toronto, Ontario, Canada
[3]Arthur Labatt Family School of Nursing, Western University, London, Ontario, Canada
[4]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada
[5]Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada
[6]ICES, Toronto, Ontario, Canada

Acknowledgements 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.

Contributors LR conceived 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 the study, participated in the study design, interpretation of study results and provided feedback on the manuscript. SWH, 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.

Funding This study was supported by an Ontario Trillium Foundation–Local Poverty Reduction Fund grant (held by CF, RB and SZS), an Academic Medical Organization of Southwestern Ontario Academic Health Science Centre Alternate Funding Innovation Fund grant (held by KC and SZS) and two Canadian Institutes of Health Research grants, CIHR MOP-86765 and HOA-80066 (held by SWH and RN). 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. Information in the Immigration, Refugees and Citizenship Canada Permanent Resident Dataset was provided by Immigration, Refugees and Citizenship Canada.

Disclaimer The opinions, results and conclusions reported in this paper are those of the authors and are independent of 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.

Competing interests KC reports grants from AstraZeneca, and attendance to a Merck conference, occurring outside the submitted work. All other authors have no competing interest to declare.

Patient consent for publication Not required.

Ethics approval This study was approved by the St Michael's Hospital Research Ethics Board, and follows the STARD guidelines for reporting diagnostic accuracy studies.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD
Lucie Richard http://orcid.org/0000-0001-6577-5067

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;384:1529–40.
2. Hwang SW, Chambers C, Chiu S, *et al*. A comprehensive assessment of health care utilization among homeless adults under a system of universal health insurance. *Am J Public Health* 2013;103:S294–S301.
3. Fitzgerald ST, Shelley MC, Dail PW. Research on homelessness: sources and implications of uncertainty. *Am Behav Sci* 2001;45:121–48.
4. 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.
5. 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.
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, 2012 [Accessed 2 January 2019].
8. Williams M. Can We Measure Homelessness? A Critical Evaluation of 'Capture-Recapture'. *Methodol Innov Online* 2010;5:49.1–59.
9. Employment and Social Development Canada. Everyone counts: a guide to Point-in-Time counts in Canada – 2nd edition, 2017. Available: 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 [Accessed 3 January 2019].
10. Gaetz S, Dej E, Richter T, *et al*. *The state of homelessness in Canada*. Toronto: Canadian Observatory on Homelessness Press, 2016.
11. City of Toronto. Backgrounder: results of the 2018 street needs assessment, 2018. Available: https://www.toronto.ca/home/media-room/backgrounders-other-resources/backgrounder-results-of-the-2018-street-needs-assessment/ [Accessed 3 December 2018].
12. Rabinovitch H. Snapshot of an object in motion: quantifying homelessness. Simon Fraser university, 2005. Available: http://summit.sfu.ca/item/15297 [Accessed 10 July 2019].
13. Burnes D, DiLeo D. *Ending homelessness: why we haven't, how we can*. Boulder, Colorado: Lynne Rienner, 2016.
14. Gabbard WJ, Snyder CS, Lin MB, *et al*. Methodological issues in enumerating homeless individuals. *J Soc Distress Homeless* 2007;16:90–103.
15. Canada's National Housing Strategy, 2018. Available: https://www.placetocallhome.ca/pdfs/Canada-National-Housing-Strategy.pdf [Accessed 2 July 2019].
16. CIHI. Canadian Institute for health information, 2018. Available: https://www.cihi.ca/en [Accessed 3 January 2019].
17. Public Health Agency of Canada. The Canadian chronic disease surveillance system – an overview, 2018. Available: https://www.canada.ca/en/public-health/services/publications/canadian-chronic-disease-surveillance-system-factsheet.html [Accessed 3 January 2019].
18. Government of Canada. Publications – diseases and conditions, 2018. Available: https://www.canada.ca/en/services/health/publications/diseases-conditions.html [Accessed 3 January 2019].
19. Lix L, Ayles J, Bartholomew S, *et al*. The Canadian chronic disease surveillance system: a model for collaborative surveillance. *Int J Popul Data Sci* 2018;3.

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:e39915.
21. Wilson K, Hawken S, Kwong JC, *et al*. Vaccine and immunization surveillance in Ontario (vision) – using linked health administrative databases to monitor vaccine safety. *Vaccine* 2012;30:6115–20.
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. *Can J Diabetes* 2014;38:179–85.
23. Feely A, Lix LM, Reimer K, Allison F, Lix Lisa M, Kim R. Estimating multimorbidity prevalence with the Canadian chronic disease surveillance system. *Health Promot Chronic Dis Prev Can* 2017;37:215–22.
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;10:e0123204.
25. 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;64:821–9.
26. Carrara G, Scirè CA, Zambon A, *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;5:e006029.
27. Butt DA, Tu K, Young J, *et al*. A validation study of administrative data algorithms to identify patients with parkinsonism with prevalence and incidence trends. *Neuroepidemiology* 2014;43:28–37.
28. Widdifield J, Bernatsky S, Paterson JM, *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 Res* 2013;65:1582–91.
29. ICES, 2019. Available: www.ices.on.ca [Accessed 2 January 2019].
30. 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 Public Health* 2011;56:609–23.
31. 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:e0180338.
32. 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;163:1073–9.
33. Gaetz S, Donaldson J, Richter T, *et al*. *The state of homelessness in Canada*. Toronto: Canadian Homelessness Research Network Press, 2013.
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. *Int J Public Health* 2018:1–11.
35. ICES. Data dictionary, 2019. Available: https://datadictionary.ices.on.ca/Applications/DataDictionary/Default.aspx [Accessed 3 January 2019].
36. Canadian Institute for Health Information. Recording postal codes in the dad and NACRS. Available: https://www.cihi.ca/en/recording-postal-codes-in-dad-and-nacrs_en.pdf [Accessed 27 December 2018].
37. Robinson JR, Young TK, Roos LL, *et al*. Estimating the burden of disease: comparing administrative data and self-reports. *Medical care* 1997;1:932–47.
38. Hoogenveen R, Westert G, Dijkgraaf M, *et al*. Disease prevalence estimations based on contact registrations in general practice. *Stat Med* 2002;21:2271–85.
39. Gershon AS, Wang C, Guan J, *et al*. Identifying individuals with Physcian diagnosed COPD in health administrative databases. *COPD* 2009;6:388–94.
40. Hux JE, Ivis F, Flintoft V, *et al*. Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 2002;25:512–6.
41. 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 Dis Inj Can* 2013;33:160–6.
42. Antoniou T, Zagorski B, Loutfy MR, *et al*. Validation of case-finding algorithms derived from administrative data for identifying adults living with human immunodeficiency virus infection. *PLoS One* 2011;6:e21748.
43. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 1927;22:209–12.
44. SAS Enterprise Guide9.4Cary, NC, 2018SAS Institute Inc. Available: https://www.sas.com/en_ca/home.html
45. McGee S. Simplifying likelihood ratios. *J Gen Intern Med* 2002;17:647–50.
46. Duchesne A, Cooper I. National homelessness data national shelter data 2005-2016 Point-in-Time count 2018, 2019. Available: http://datathatmakesadifference.com/wp-content/uploads/2019/06/Annie-Duchesne-Ian-Cooper.pdf [Accessed 10 July 2019].
47. Ministry of Municipal Affairs and Housing. A place to call home: report of the expert Advisory panel on homelessness, 2015. Available: http://www.mah.gov.on.ca/AssetFactory.aspx?did=11038 [Accessed 15 Jul 2019].
48. Statistics Canada. Ontario [Province] and Canada [Country] (table) Census Profile. 2016 census. statistics Canada catalogue No. 98-316-X2016001. Available: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E [Accessed Released 29 November 2017].
49. Gaetz S, Gulliver T, Richter T. *The state of homelessness in Canada*. Toronto: Canadian Observatory on Homelessness Press, 2014.
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: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:S225–S231.
52. Bharel M, Lin W-C, Zhang J, *et al*. Health care utilization patterns of homeless individuals in Boston: preparing for Medicaid expansion under the Affordable care act. *Am J Public Health* 2013;103:S31 1–S317.
53. Zech J, Husk G, Moore T, *et al*. Identifying homelessness using health information exchange data. *J Am Med Inform Assoc* 2015;26:682–7.
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 Annu Symp Proc* 2013;2013:537–46.
56. Erickson J, Abbott K, Susienka L. Automatic address validation and health record review to identify homeless social security disability applicants. *J Biomed Inform* 2018;82:41–6.
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 Serv Res* 2018;53:1992–2004.
59. Patrick C. Aboriginal homelessness in Canada: a literature review, 2014. Available: http://homelesshub.ca/sites/default/files/AboriginalLiteratureReview.pdf [Accessed 3 January 2019.].
60. Government of Canada. Reaching Home: Canada's Homelessness Strategy, 2018Published. Available: https://www.canada.ca/en/employment-social-development/news/2018/11/reaching-home-canadas-homelessness-strategy.html [Accessed January 3, 2019].