Creation of a Case Definition to Identify People Living with HIV using Canadian Primary Care Sentinel Surveillance Network data

Sarah Elizabeth Boyd

Memorial University of Newfoundland Faculty of Medicine <https://orcid.org/0000-0001-6673-4646>

Tao Chen

Canadian Primary Care Sentinel Surveillance Network

Jillian Blackmore

Memorial University of Newfoundland

Marissa Becker

University of Manitoba

Claire Kendall

University of Ottawa

Alexander Singer

University of Manitoba

Laurie Ireland

Winnipeg Regional Health Authority

Shabnam Asghari (✉ shabnam.asghari@med.mun.ca)

<https://orcid.org/0000-0002-4970-0978>

Research article

Keywords: People living with HIV; case definition; Human immunodeficiency Virus; Algorithm Validation; Sensitivity; Specificity, EMR

Posted Date: March 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-18251/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Creation of a Case Definition to Identify People Living with HIV using Canadian Primary Care Sentinel Surveillance Network data

Sarah Boyd, Tao Chen, Jillian Blackmore, Marissa Becker, Claire Kendall, Alexander Singer, Laurie Ireland
Shabnam Asghari (corresponding author)

Abstract

Background: HIV care that has been predominantly specialty based can often be managed in primary care settings. The Canadian Primary Care Sentinel Surveillance Network (CPCSSN) database is a rich source of electronic medical records of primary care practices. A case definition for HIV using CPCSSN allows the examination of epidemiology and healthcare utilization of people living with HIV (PLHIV) whose records are in this database and assists us to better understand the primary care of PLHIV. The objective of this study was to develop a case definition for HIV applicable to the CPCSSN database.

Methods: Electronic medical record data from CPCSSN was used to develop a retrospective cohort between 2009 and 2017. We identified all possible records of PLHIV in the CPCSSN dataset based on the presence of HIV codes, keywords and medications. Multiple combination of codes, keywords and medications were analyzed to see which resulted in the most accurate definition of PLHIV. To assess the validity, we linked the data to external references (a Newfoundland and Labrador (NL) PLHIV cohort and Nine Circles HIV clinic in Manitoba) and an internal reference (a random sample of the CPCSSN database which was reviewed by two experts to confirm HIV status).

Results: We found that the presence of an HIV keyword along with either an ICD code or taking 3 or more HIV medications concurrently was the most accurate algorithm for predicting PLHIV with a sensitivity of 95%, 97.9% and 88.9%, specificity of 63.5%, 100% and 100% for the NL PLHIV cohort, Nine Circles HIV clinic data and internal reference, respectively.

Conclusion: To our knowledge, this is the first study to develop an algorithm for identifying people with HIV applicable to the Canadian Primary Care Sentinel Surveillance Network database. This algorithm-based case definition will support future research investigating the utilization of primary healthcare by PLHIV and facilitate improvements in primary care for PLHIV.

Keywords: People living with HIV; case definition; Human immunodeficiency Virus; Algorithm Validation; Sensitivity; Specificity, EMR

Background

The care of people living with human immunodeficiency virus (PLHIV) has evolved in recent years as HIV has become a chronic condition for those on treatment.¹ With the success of antiretroviral treatment (ART), AIDS-related illnesses are no longer the biggest threat to PLHIV, rather, emerging complications resulting from aging and ART toxicities are of concern.¹ Common comorbidities in the general population such as cardiovascular disease, diabetes, mental health conditions and some cancers are now increasingly associated with chronic HIV infection². In general, the management of multi-morbidity and chronic diseases is the mainstay of primary care providers (PCP) and for this reason, HIV care which has been predominantly speciality based is now transitioning to primary care in many settings.³

The use of disease registries or sentinel surveillance to monitor patient care patterns and outcomes is an effective way to achieve practice improvements using tools such as audit and feedback or patient recalls leading to better quality of care. It has been shown that family physicians comprehensively use electronic medical records and that they are adequate sources for secondary use.⁴ The Canadian Primary Care Sentinel Surveillance Network (CPCCSN) has effectively created and validated algorithms for several chronic diseases. These case definitions are essential to conduct observational studies and pragmatic clinical trials in primary care settings.⁵

There have been previous studies that have created a case definition for HIV. For example, in Massachusetts, Kerr et al. developed an HIV database including individuals with either a positive HIV antibody test, a CD4 count <200 or HIV viral load >5000 copies/mL.⁶ From this cohort, it was found that PCPs were more likely to screen for chronic disease conditions, for example, by doing a complete blood count and measuring cholesterol and glucose levels.⁶ Another database was developed by Paul et al. in

North Carolina in which a computed phenotype for HIV infection was developed.⁷ In this study, a panel of physicians created two algorithms for detecting PLHIV using electronic medical records (EMR) consisting of different laboratory tests, HIV-specific medications and International Classification of Diseases (ICD) billing codes. It was found that first having an ICD-9 code followed by a positive laboratory test or taking a combination of HIV-specific medications resulted in a higher sensitivity and specificity for determining PLHIV.⁷ These algorithms included individuals who were likely PLHIV, but they may not be generalizable to the Canadian population. A third example, Antoniou et al., created a case finding algorithm for HIV infection using administrative health databases in Ontario.⁸ They applied 48 case-finding algorithms based on combinations of physician billing claims, hospital and emergency room separations and prescription drug claims and found that with the exception of algorithms using only a single physician claim, the specificity of all algorithms exceeded 99%. They recommend that an algorithm consisting of three physician claims over a three year period would be optimal for identification of PLHIV.⁸

The objective of this study was to develop an algorithm for identifying PLHIV within CPCSSN. To our knowledge, this is the first study that validates the case definition of PLHIV in primary care using CPCSSN data.

Methods

CPCSSN Database

We obtained a dataset from CPCSSN - Canada's largest multi-system primary care database with EMR data from over 1.7 million patients nationwide in 8 provinces/territories.⁹ Participating family physicians across Canada voluntarily share EMR data from their practices which supports chronic disease surveillance, research and quality improvement efforts.⁹

Possible HIV cases in CPCSSN

We included EMR data for all CPCSSN patients seen between January 1, 2009 and December 31, 2017 and identified those with any of the following: All individuals whose CPCSSN EMR contained either (1) An ICD-9 or ICD-10 code related to HIV from billing or encounter diagnosis and health condition (See Table 1 of the Appendix), (2) An HIV-related keyword in provider entered text in billing, encounter diagnosis or health condition tables (See Table 2 of the Appendix) or (3) Presence of any HIV drug listed under prescribed medication in the CPCSSN EMR (See Table 3 of the Appendix). Our keywords were HIV-specific so we did not include "viral load" or "VL" as it would result in unrelated conditions being included. From this list, the keywords were scanned for any unrelated keywords (e.g. hearing aids) and negative keywords (e.g. HIV negative). These records were subsequently removed to make the final study population. A list of unrelated and negative HIV keywords can be found in Table 4 of the appendix.

External Reference Databases

We used two external *references: one from NL and one from Manitoba*. As an external reference for the NL population, we used data from a prior study of PLHIV in NL.¹⁰ This cohort used the Public Health Laboratory data, Health Administrative data and the HIV Clinic data. This data linkage was conducted by the NL Centre for Health Information.¹⁰

The second external reference was the data from Nine Circles Community Health Centre, the primary care site of the Manitoba HIV Program. The Manitoba HIV Program has a centralized referral process in which providers are encouraged to refer to the program for HIV care provided at 3 sites: Nine Circles (where PCPs provide care in Winnipeg), 7th Street Access Centre (where PCPs provide care in Brandon), or Health Sciences Centre (where Infectious Disease specialists provide care in Winnipeg). Nine Circles has the largest population of PLHIV receiving care in a primary care setting in Manitoba. Nine Circles PCPs all participate in CPCSSN, therefore their clinic's data is held in the central repository. Related to a prior quality improvement effort, Nine Circles supported PCPs with time to review all patients living with HIV in their case load, update their medical history information entered into the EMR, and utilize a single recommended ICD-9 code of 042 for all patients with HIV. For this reason, we used presence of 042 in the CPCSSN derived Health Condition record as an external reference. Approximately 50% of patients receiving care at Nine Circles are PLHIV whereas the rest receive their general care at the community health centre.

132 *Internal Reference Database*

133 As we only had data from two provinces as external references, we wanted to assess whether our
134 algorithm could apply to all of the CPCSSN Canada database. To develop an internal reference, a random
135 sample of 100 individuals with at least one HIV keyword, code or medication in CPCSSN was selected.
136 The sample size of 100 cases was estimated based on prevalence of HIV in Canada at the time of this
137 study, confidence level of 95% and power of 80%. Two independent researchers with expertise in HIV
138 reviewed codes, keywords and medications, and assigned the cases as either PLHIV or not.
139 Disagreement between the reviewers was discussed in a meeting and resolved by a third reviewer.

140

141 *Algorithm Development*

142 We generated a list of possible algorithms based on any combination of HIV-related ICD codes,
143 keywords or 3 or more HIV medications. Furthermore, we considered the presence of multiple HIV
144 codes within two years and again within three years. For the algorithm including medications, we
145 considered the presence of three or more medications taken concurrently instead of just one or two
146 medications as some medications used to treat HIV are also used to treat hepatitis, whereas the
147 combination of three or more antiretroviral therapies are more specific to HIV.¹¹ When counting the
148 number of HIV medications, we included the number of different medications and not the number of
149 pills taken. For example, if someone took one pill containing 3 medications they would be counted as
150 taking 3 HIV medications.

151

152

153

Algorithm Validation

To assess the validity of the algorithm, we compared the cases identified by applying all possible algorithms in the CPCSSN dataset to the cases identified via the three references. First, we compared all potential cases from NL to the reference from the independent study mentioned above.¹² We also compared all potential cases of PLHIV in CPCSSN to patients at Nine Circles in Manitoba with documented HIV. For the internal reference, we compared every possible algorithm in the random sample to the cases identified as PLHIV by the review process.

We calculated the accuracy and performance indices (sensitivity, specificity, negative predictive value, positive predictive value, kappa agreement and area under receiver operating characteristic curve) of all algorithms compared to the internal and external references.

Selection of the Optimal Algorithm

To select the optimal algorithm, we compared the area under receiver operating characteristic curve (AUC) from each dataset (NL, Nine Circles and CPCSSN Canada).¹³⁻¹⁶ First, we considered the algorithms that had AUC >0.79 in all three eligible algorithms. Among them, the algorithms showing Kappa >0.75 in at least two datasets were selected as potentially being the optimal algorithm.^{17,18} We then compared the AUC for the selected algorithms (with AUC >0.79 in three datasets and Kappa >0.75 in at least two datasets) to the other algorithms.¹⁹ To test the statistical significance among the areas under these dependent ROC curves, we used the method suggested by DeLong et al. (1988). A p-value <0.05 was considered significant for all analyses. All analyses were conducted using the STATA software version 13.

Ethics approval

Ethics was obtained from the Health Research Ethics Board of NL. We also received permission from the CPCSSN Standing Committee on Research and Surveillance which provides approval for use of data held in the national repository which includes the Nine Circles data.

Results

There were 6799 individuals with at least one HIV code, keyword or 3 or more concurrent HIV medications in CPCSSN. Figure 1 describes how the final cohort was derived.

*Note: Individuals could be in more than one group

Figure 1: Flow chart showing the individuals in CPCSSN with HIV codes, keywords and medications.

Among the 6799 cases, 83 cases were from NL and linkage to the external reference in NL (n=251) resulted in 20 confirmed cases of PLHIV in NL CPCSSN data. Another 792 cases were from the Nine Circles clinic of which 765 were confirmed positive. Furthermore, 100 cases of people who had at least one HIV keyword, code or medication in CPCSSN were randomly selected of which 63 were determined to be PLHIV by experts.

Algorithms:

Of the individuals who had at least one code, keyword or medication, only 2.9% had more than one HIV code during the study period. Because the numbers were so low, an algorithm comprising at least two HIV codes was not considered as it resulted in a very low sensitivity (2.5-55%). Sensitivity, specificity, positive predictive value, negative predictive value, kappa and area under the curve for other combinations of codes, keywords and medications were calculated, with visual results from the different algorithms being shown in Figures 2, 3 and 4. Numerical results from all algorithms can be found in the appendix. As it was shown in Figures 2-4, the accuracy and performance indices are improved when the number of case finding approaches in an algorithm increases. We found variability in the accuracy based on the region in which the algorithms were tested. For example, a case finding algorithm including only keywords had a sensitivity of 95.0%, 98.3%, and 93.7%, specificity of 19.0%, 100%, and 56.8% and AUC of 0.57, 0.99, 0.81 for NL, Manitoba, and Canada respectively. Combination of the keywords and at least three HIV medications or an ICD code has a sensitivity of 95.0%, 97.9% and 88.9%, specificity of 63.5%, 100% and 100% and AUC of 0.79, 0.99 and 0.94 for NL, Manitoba, and Canada respectively.

Table 1: Accuracy and performance values of cases with an HIV keyword and either an HIV code or 3 or more HIV medications taken concurrently (chosen algorithm) for NL, Nine Circles and Canadian data.

	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Kappa (SE)	ROC (CI)
NL	95.0 (75.1-99.9)	63.5 (50.4-75.3)	45.2 (29.8-61.3)	97.6 (87.1-99.9)	0.43 (0.09)	0.79 (0.72-0.87)
Nine Circles Clinic	97.9 (96.6-98.8)	100 (87.2-100)	100 (99.5-100)	62.8 (46.7-77.0)	0.76 (0.03)	0.99 (0.98-0.99)
Canadian Random Sample	88.9 (78.4-95.4)	100 (90.5-100)	100 (93.6-100)	84.1 (69.9-93.4)	0.86 (0.09)	0.94 (0.90-0.98)

Figure 2: Accuracy and performance values of all HIV algorithms compared with NL data. “Meds” indicates taking 3 or more HIV medications. “All 3” indicates having an HIV keyword, HIV code and taking 3 or more HIV medications.

Figure 3: Accuracy and performance values of all HIV algorithms compared with Nine Circles Clinic data. “Meds” indicates taking 3 or more HIV medications. “All 3” indicates having an HIV keyword, HIV code and taking 3 or more HIV medications.

Figure 4: Accuracy and performance values of all HIV algorithms compared with a random sample from CPCSSN (Canadian data). “Meds” indicates taking 3 or more HIV medications. “All 3” indicates having an HIV keyword, HIV code and taking 3 or more HIV medications.

We determined that the presence of an HIV keyword in the EMR along with either an ICD code or taking 3 or more HIV medications concurrently was the optimal algorithm for predicting PLHIV as it had AUC of > 0.79 in all three datasets and Kappa > 0.75 in two of the three datasets (see Table 1). Comparing the AUC for this algorithm to the AUC of other algorithms in three datasets either indicated significantly better diagnostic capabilities of the algorithm ($p < 0.05$) or did not show any significant difference with other algorithms with AUC > 0.80 . From our three references, the sensitivity for this algorithm ranged from 88.9 to 97.9 and the specificity ranged from 63.5 to 100. This algorithm resulted in the identification of 4222 cases in the CPCSSN database.

Discussion

Based on our validation, we found that the optimal algorithm for determining PLHIV was the presence of an HIV keyword with either an HIV code or taking 3 or more HIV medications concurrently. While the use of multiple ICD codes is often useful for defining chronic conditions in administrative data used for secondary purposes, our study demonstrated a low sensitivity for exclusively using HIV ICD codes from a primary care sample.²⁰

Our findings are consistent with previous validation efforts for other conditions. For example, CPCSSN has previously validated case definitions for eight diseases which involve both ICD-9 codes and disease-specific medications.²¹

Canadian EMRs often allow for free text entry in the "health condition" or "problem list" entries. For this reason, it is important to include keywords that are free text to include more search results as a

248 computer algorithm would not recognize misspelled words. For example, we chose to search for
249 “human immun” instead of “human immunodeficiency virus” because when physicians would write out
250 the word, we would find other versions like “human immune virus”, “human immuno virus”, “human
251 immunodeficiency”. Shortening the search term allowed us to include all variations of the word.
252 Furthermore, it is also important to remove unrelated keywords. In the creation of our database.
253 Searching for the word “AIDS” for example resulted in the condition “AIDS” which was expected, but
254 also other unrelated keywords such as “hearing aids” and “NSAIDS”. The list of included keywords had
255 to be manually reviewed for this reason to see which keywords were actually unrelated to HIV. Even
256 after removing negative and unrelated keywords, there will still be cases which are not PLHIV that will
257 still be picked up as having a positive keyword. For example, the keywords for physician notes reading
258 “brother is HIV-positive” does not indicate the person likely has HIV themselves. As such, the presence
259 of keywords alone does not ensure that the algorithm has correctly identified a person living with HIV.

260

261 For both the Canadian sample and Nine Circles data, algorithms containing HIV keywords and/or ICD
262 codes were optimal. This was in contrast to the NL cohort where algorithms that included HIV
263 medications were optimal. However, some individuals did not appear in the NL database but seem very
264 likely to have HIV. For example, some individuals did not appear in the NL external reference database
265 for PLHIV but in the CPCSSN database had “Positive HIV Ab Serology” in their condition keywords,
266 “AIDS/HIV” in their encounter keywords, an ICD-9 code for HIV, and were taking 3 HIV medications. It
267 seems virtually impossible that these individuals would not have HIV, yet they were deemed negative
268 based on the NL external reference. These individuals may have tested positive for HIV in another
269 province and care is provided by a primary care provider in NL. For this reason, it seems that the
270 algorithm including individuals with an HIV keyword and either codes or 3 or more concurrent HIV
271 medications was the optimal algorithm.

The study had several limitations. Our algorithms did not include laboratory tests as the CPCSSN database did not have laboratory test information for HIV. At this time, we could not link de-identified CPCSSN data with provincial laboratory data and thus cannot know with certainty if the individuals in the CPCSSN database are PLHIV. Comparing the cases using three references including two with confirmed cases can combat this problem, although one may say that Nine Circles is not a completely independent external reference. The Nine Circles clinic was already linked to CPCSSN and they used ICD codes for their case definition at the time of this study. Moreover, for the concurrent use of medications, we assumed that medications without a listed stop date were continued indefinitely. One may also suggest the use of regression model-based algorithm selection to support the algorithm selection method. Although use of highest diagnostic validity and the statistical significance among the areas under the dependent ROC curves is common in selection of algorithms, our study was limited to 10 algorithms therefore we were not able to use model-based algorithm selection.²²⁻²⁵

Conclusions

Using primary care electronic health data, we created an optimal algorithm for identifying PLHIV. This algorithm includes the presence of an HIV keyword (from physicians notes from billing, encounter diagnosis or health condition), along with either an ICD code or taking 3 or more HIV medications concurrently was the optimal algorithm for identifying PLHIV. We anticipate this will contribute to future research and the capacity for quality improvement in the care of PLHIV in the primary care setting.

294 **List of Abbreviations**

295 ART – antiretroviral therapy

296 CPCSSN - Canadian Primary Care Sentinel Surveillance Network

297 EMR - Electronic Medical Records

298 HIV - Human Immunodeficiency Virus

299 ICD - International Classification of Diseases

300 NL – Newfoundland and Labrador

301 PCP – Primary care provider

302 PLHIV – People living with the Human Immunodeficiency Virus

303

304 **Declarations**

305

306 **Ethics approval and consent to participate:** Ethics was obtained from the Health Research Ethics Board
307 of Newfoundland and Labrador.

308

309 **Consent for publication:** Not applicable.

310

311 **Availability of data and material:** The datasets used and/or analysed during the current study are
312 available from the corresponding author on reasonable request.

313

314 **Competing interests:** The authors declare that they have no competing interests.

315

316 **Funding:** Funding for this project was received from the Canadian Institutes for Health Research.

317

318 **Authors' contributions:** SA designed the study, SB and TC extracted the data, SB and SA analyzed and
319 interpreted the data. SB drafted the manuscript, and all authors read, provided feedback and approved
320 the final manuscript.

321

322

323

324

325

326

327

328

329 **References**

- 330 1. Deeks SG, Lewin SR, Havlir DV. The end of AIDS: HIV infection as a chronic disease. *The Lancet*.
331 2013;382(9903):1525-1533.
- 332 2. Kendall CE, Wong J, Taljaard M, et al. A cross-sectional, population-based study measuring
333 comorbidity among people living with HIV in ontario. *BMC Public Health*. 2014;14(1):161.
- 334 3. Johnston S, Kendall C, Hogel M, McLaren M, Liddy C. Measures of quality of care for people with HIV:
335 A scoping review of performance indicators for primary care. *PLoS one*. 2015;10(9):e0136757.
- 336 4. Tu K, Widdifield J, Young J, et al. Are family physicians comprehensively using electronic medical
337 records such that the data can be used for secondary purposes? A canadian perspective.
338 *BMC Med Inform Decis Mak*. 2015;15(1):67.
- 339 5. Williamson T, Green ME, Birtwhistle R, et al. Validating the 8 CPCSSN case definitions for chronic
340 disease surveillance in a primary care database of electronic health records. *Ann Fam Med*.
341 2014;12(4):367-372. doi: 10.1370/afm.1644
- 342 6. Kerr CA, Neeman N, Davis RB, et al. HIV quality of care assessment at an academic hospital: Outcomes
343 and lessons learned. *Am J Med Qual*. 2012;27(4):321-328.
- 344 7. Paul DW, Neely NB, Clement M, et al. Development and validation of an electronic medical record
345 (EMR)-based computed phenotype of HIV-1 infection. *JAMIA*. 2017;25(2):150-157.
- 346 8. Antoniou T, Zagorski B, Loutfy MR, Strike C, Glazier RH. Validation of case-finding algorithms derived
347 from administrative data for identifying adults living with human immunodeficiency virus infection. *PLoS*
348 *one*. 2011;6(6):e21748.

349 9. Canadian Primary Care Sentinel Surveillance Network. Canadian primary care sentinel surveillance
350 network. cpssn.ca (2016). Accessed 6 Nov 2019.

351 10. Asghari S, Boyd S, Knight J, et al. Developing a comprehensive database with sensitive health
352 information: A profile of people living with HIV in newfoundland and labrador, canada. IJPS. 2019;In
353 Press.

354 11. U.S. Department of Health and Human Services. Guidelines for the use of antiretroviral agents in
355 adults and adolescents living with HIV. [https://aidsinfo.nih.gov/guidelines/html/1/adult-and-adolescent-](https://aidsinfo.nih.gov/guidelines/html/1/adult-and-adolescent-arv/25/hbv-hiv)
356 [arv/25/hbv-hiv](https://aidsinfo.nih.gov/guidelines/html/1/adult-and-adolescent-arv/25/hbv-hiv) (2017). Accessed 17 Jan 2019.

357 12. Liddy C. Advancing primary healthcare for persons living with HIV in canada. 2013-2018.

358 13. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated
359 receiver operating characteristic curves: A nonparametric approach. Biometrics. 1988;44(3):837-845.

360 14. Gur D, Bandos AI, Rockette HE. Comparing areas under receiver operating characteristic curves:
361 Potential impact of the “last” experimentally measured operating point. Radiology. 2008;247(1):12-15.

362 15. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic
363 curves derived from the same cases. Radiology. 1983;148(3):839-843. doi:
364 10.1148/radiology.148.3.6878708

365 16. Cleves MA. From the help desk: Comparing areas under receiver operating characteristic curves from
366 two or more probit or logit models. The Stata Journal. 2002;2(3):301-313.

367 17. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol. 1993;46(5):423-429.

- 368 18. McHugh ML. Interrater reliability: The kappa statistic. *Biochemia medica: Biochemia medica*.
369 2012;22(3):276-282.
- 370 19. Fagerland MW, Lydersen S, Laake P. The McNemar test for binary matched-pairs data: Mid-p and
371 asymptotic are better than exact conditional. *BMC Med Res Methodol*. 2013;13(1):91.
- 372 20. Kosowan L, Wicklow B, Queenan J, Yeung R, Amed S, Singer A. Enhancing health surveillance:
373 Validation of a novel electronic medical Records–Based definition of cases of pediatric type 1 and type 2
374 diabetes mellitus. *Can J Diabetes*. 2019.
- 375 21. CPCSSN Team. Case definitions: Canadian primary care sentinel surveillance network (CPCSSN),
376 version 2. <http://cpcssn.ca/wp-content/uploads/2019/07/CPCSSN-Case-Definitions-v2.pdf> (2019).
377 Accessed 6 Nov 2019.
- 378 22. Kroeker K, Widdifield J, Muthukumarana S, Jiang D, Lix LM. Model-based methods for case
379 definitions from administrative health data: Application to rheumatoid arthritis. *BMJ Open*.
380 2017;7(6):e016173-2017-016173. doi: 10.1136/bmjopen-2017-016173
- 381 23. Chubak J, Pocobelli G, Weiss NS. Tradeoffs between accuracy measures for electronic health care
382 data algorithms. *J Clin Epidemiol*. 2012;65(3):343-349. e2.
- 383 24. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of
384 reporting guidelines for assessing the quality of validation studies of health administrative data. *J Clin*
385 *Epidemiol*. 2011;64(8):821-829.

25. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;3(1):25.

26. Nuesoft Technologies I. ICD-10 code lookup. <https://icdcodelookup.com/>. Accessed 17 Jan 2019.

27. Ministry of Health and Long-Term Care Primary Health Care Branch. Primary care physician compensation and complex patients. 2014.

Appendix:

Table 1: List of EMR codes included in the patient population

Code	Source	Description
042-044	ICD-9	Human immunodeficiency virus [HIV] disease ²⁶
V08	ICD-9	Asymptomatic human immunodeficiency virus [HIV] infection status ²⁶
795.71	ICD-9	Nonspecific serologic evidence of human immunodeficiency virus [HIV] ²⁶
B20-B24	ICD-10	Human immunodeficiency virus [HIV] disease ²⁶
R75	ICD-10	Inconclusive laboratory evidence of human immunodeficiency virus [HIV] ²⁶
Z21	ICD-10	Asymptomatic human immunodeficiency virus [HIV] infection status ²⁶
K022	Service code	HIV Primary Care ²⁷

Table 2: List of HIV keywords included in the patient population

"human immun"
"HIV"
"H.I.V."
"AIDS"
"A.I.D.S."
"acquired immun"
"antiretroviral"
"CD4"
"B5701"

Table 3: ATC Codes of HIV medications included in the patient population

J05AE01	J05AF03	J05AG05	J05AR15
J05AE02	J05AF04	J05AR01	J05AR17
J05AE03	J05AF05	J05AR02	J05AR18
J05AE04	J05AF06	J05AR03	J05AR19
J05AE05	J05AF07	J05AR04	J05AX07
J05AE07	J05AF09	J05AR06	J05AX08
J05AE08	J05AF13	J05AR08	J05AX09
J05AE09	J05AG01	J05AR09	J05AX11
J05AE10	J05AG02	J05AR10	J05AX12
J05AF01	J05AG03	J05AR13	V03AX03
J05AF02	J05AG04	J05AR14	

419 Table 4: Negative HIV keyword exclusions

Keyword	Number of keyword incidences
"HIV neg"	505
"No HIV"	118
"HIV -ve"	114
"PrEP"	41
"neg HIV"	39
"prophylaxis"	35
"negative HIV"	33
"HIV (-)"	19
"HIV test neg"	13
"HIV non reactive"	9
"HIV - neg"	8
"HIV ab neg"	7
"HIV testing neg"	6
"neg for HIV"	5
"HIV PEP"	4
"HIV test false"	3
"false + HIV"	3
"false positive HIV"	3
"HIV nonreactive"	2
"HIV-ve"	2
"HIV false"	1
"HIV(-)"	1

420

421

422

423

424

425

426

427

428

429 Table 5: Keyword exclusions unrelated to HIV

Keyword	Number of keyword incidences
"bilat"	688
"NSAIDS"	388
"hive"	169
"hear"	161
"SNHL"	113
"Sensorineural"	97
"deaf"	79
"Presbycusis"	34
"bandaids"	24
"audiology"	20
"noise induced"	19
"both ears"	18
"uses [XXX] aids"	18
"aids recommended"	17
"uses aids"	11
"Meniere"	11
"sleep aids"	9
"sleeping aids"	8
"AIDS X 2"	8
"rolaids"	8
"needs aids"	7
"mobility aids"	4
"AIDS X2"	4
"mobility aids"	4
"maids"	4
"walking aids"	3
"adaptive aids"	3
"wearing aids"	3

430

431

432

433

434

435

436 Table 6: Summary Statistics of NL

	Number of Cases (%)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Kappa (SE)	ROC (CI)
ICD Code	41 (49.4%)	90.0 (68.3-98.8)	63.5 (50.4-75.3)	43.9 (28.5-60.3)	95.2 (83.8-99.4)	0.3938 (0.0945)	0.7675 (0.6772- 0.8577)
Keywords	70 (84.4%)	95.0 (75.1-99.9)	19.0 (10.2-30.9)	27.1 (17.2-39.1)	92.3 (64.0-99.8)	0.0758 (0.0503)	0.5702 (0.5010- 0.6394)
3 or More HIV Medications	17 (20.5%)	75.0 (50.9-91.3)	96.8 (89.0-99.6)	88.2 (63.6-98.5)	92.4 (83.2-97.5)	0.7570 (0.1092)	0.8591 (0.7594- 0.9589)
ICD Code + Keywords	41 (49.4%)	90.0 (68.3-98.8)	63.5 (50.4-75.3)	43.9 (28.5-60.3)	95.2 (83.8-99.4)	0.3938 (0.0945)	0.7675 (0.6772- 0.8577)
ICD Code + 3 or More HIV Medications	14 (16.9%)	65.0 (40.8-84.6)	98.4 (91.5-100)	92.9 (66.1-99.8)	89.9 (80.2-95.8)	0.7065 (0.1071)	0.8171 (0.7087- 0.9254)
Keywords+ 3 or More HIV Medications	15 (18.1%)	70.0 (45.7-88.1)	98.4 (91.5-100)	93.3 (68.1-99.8)	91.2 (81.8-96.7)	0.7479 (0.1080)	0.8421 (0.7379- 0.9463)
ICD Code + Keywords + 3 or More HIV Medications	14 (16.9%)	65.0 (40.8-84.6)	98.4 (91.5-100)	92.9 (66.1-99.8)	89.9 (80.2-95.8)	0.7065 (0.1071)	0.8171 (0.7087- 0.9254)
ICD Code + (Keywords OR 3 or More HIV Medications)	41 (49.4%)	90.0 (68.3-98.8)	63.5 (50.4-75.3)	43.9 (28.5-60.3)	95.2 (83.8-99.4)	0.3938 (0.0945)	0.7675 (0.6772- 0.8579)
Keywords + (ICD Code OR 3 or More HIV Medications)	42 (50.6%)	95.0 (75.1-99.9)	63.5 (50.4-75.3)	45.2 (29.8-61.3)	97.6 (87.1-99.9)	0.4253 (0.0933)	0.7925 (0.7151- 0.8699)
3 or More HIV Medications + (ICD Code OR Keywords)	15 (18.1%)	70.0 (45.7-88.1)	98.4 (91.5-100)	93.3 (68.1-99.8)	91.2 (81.8-96.7)	0.7479 (0.1080)	0.8421 (0.7379- 0.9463)

437

438

439

440

441 Table 7: Summary statistics of Nine Circles Clinic

	Number of Cases (%)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Kappa (SE)	ROC (CI)
ICD Code	753 (95.1%)	96.5 (94.9- 97.7)	44.4 (25.5- 64.7)	98.0 (96.7- 98.9)	30.8 (17.0- 47.6)	0.3369 (0.0349)	0.7046 (0.6089- 0.8003)
Keywords	752 (94.9%)	98.3 (97.1- 99.1)	100 (87.2- 100)	100 (99.5- 100)	67.5 (50.9- 81.4)	0.7977 (0.0348)	0.9915 (0.9869- 0.9961)
3 or More HIV Medications	565 (71.3%)	73.9 (70.6- 76.9)	100 (87.2- 100)	100 (99.3- 100)	11.9 (8.0- 16.8)	0.1615 (0.0194)	0.8693 (0.8537- 0.8849)
ICD Code + Keywords	738 (93.2%)	96.5 (94.9- 97.7)	100 (87.2- 100)	100 (99.5- 100)	50.0 (36.1- 63.9)	0.6508 (0.0333)	0.9824 (0.9758- 0.9889)
ICD Code + 3 or More HIV Medications	541 (68.3%)	70.7 (67.4- 73.9)	100 (87.2- 100)	100 (99.3- 100)	10.8 (7.2- 15.3)	0.1414 (0.0182)	0.8536 (0.8375- 0.8697)
Keywords+ 3 or More HIV Medications	552 (69.7%)	72.2 (68.8- 75.3)	100 (87.2- 100)	100 (99.3- 100)	11.3 (7.5- 15.9)	0.1502 (0.0187)	0.8608 (0.8449- 0.8767)
ICD Code + Keywords + 3 or More HIV Medications	541 (68.3%)	70.7 (67.4- 73.9)	100 (87.2- 100)	100 (99.3- 100)	10.8 (7.2- 15.3)	0.1414 (0.0182)	0.8536 (0.8375- 0.8697)
ICD Code + (Keywords OR 3 or More HIV Medications)	738 (93.2%)	96.5 (94.9- 97.7)	100 (87.2- 100)	100 (99.5- 100)	50.0 (36.1- 63.9)	0.6508 (0.0333)	0.9824 (0.9758- 0.9889)
Keywords + (ICD Code OR 3 or More HIV Medications)	749 (94.6%)	97.9 (96.6- 98.8)	100 (87.2- 100)	100 (99.5- 100)	62.8 (46.7- 77.0)	0.7614 (0.0345)	0.9895 (0.98445- 0.9946)
3 or More HIV Medications + (ICD Code OR Keywords)	552 (69.7%)	72.2 (68.8- 75.3)	100 (87.2- 100)	100 (99.3- 100)	11.3 (7.5- 15.9)	0.1502 (0.0187)	0.8608 (0.8449- 0.8767)

442

443

444

445 Table 8: Summary statistics of Canadian sample

	Number of Cases (%)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Kappa (SE)	ROC (CI)
ICD Code	66 (66%)	88.9 (78.4-95.4)	73.0 (55.9-86.2)	84.8 (73.9-92.5)	79.4 (62.1-91.3)	0.6291 (0.0998)	0.8093 (0.7269- 0.8917)
Keywords	75 (75%)	93.7 (84.5-98.2)	56.8 (39.5-72.9)	78.7 (67.7-87.3)	84.0 (63.9-95.5)	0.5402 (0.0961)	0.8093 (0.7269- 0.8917)
3 or More HIV Medications	37 (37%)	58.7 (45.6-71.0)	100 (90.5-100)	100 (90.5-100)	58.7 (45.6-71.0)	0.5129 (0.0873)	0.7937 (0.7324- 0.8549)
ICD Code + Keywords	55 (55%)	87.3 (76.5-94.4)	100 (90.5-100)	100 (93.5-100)	82.2 (67.9-92.0)	0.8357 (0.0986)	0.9365 (0.8951- 0.9780)
ICD Code + 3 or More HIV Medications	33 (33%)	52.4 (39.4-65.1)	100 (90.5-100)	100 (89.4-100)	55.2 (42.6-67.4)	0.4487 (0.0834)	0.7619 (0.6998- 0.8241)
Keywords+ 3 or More HIV Medications	33 (33%)	52.4 (39.4-65.1)	100 (90.5-100)	100 (89.4-100)	55.2 (42.6-67.4)	0.4487 (0.0834)	0.7619 (0.6998- 0.8241)
ICD Code + Keywords + 3 or More HIV Medications	32 (32%)	50.8 (37.9-63.6)	100 (90.5-100)	100 (89.1-100)	54.4 (41.9-66.5)	0.4331 (0.0824)	0.7540 (0.6918- 0.8162)
ICD Code + (Keywords OR 3 or More HIV Medications)	56 (56%)	88.9 (78.4-95.4)	100 (90.5-100)	100 (93.6-100)	84.1 (69.9-93.4)	0.8555 (0.0990)	0.9444 (0.9053- 0.9836)
Keywords + (ICD Code OR 3 or More HIV Medications)	56 (56%)	88.9 (78.4-95.4)	100 (90.5-100)	100 (93.6-100)	84.1 (69.9-93.4)	0.8555 (0.0990)	0.9444 (0.9053- 0.9836)
3 or More HIV Medications + (ICD Code OR Keywords)	34 (34%)	54.0 (40.9-66.6)	100 (90.5-100)	100 (89.7-100)	56.1 (43.3-68.3)	0.4645 (0.0845)	0.7698 (0.7078- 0.8319)

446

Figures

*Note: Individuals could be in more than one group

Figure 1

Flow chart showing the individuals in CPCSSN with HIV codes, keywords and medications.

Figure 2

Accuracy and performance values of all HIV algorithms compared with NL data. “Meds” indicates taking 3 or more HIV medications. “All 3” indicates having an HIV keyword, HIV code and taking 3 or more HIV medications.

Figure 3

Accuracy and performance values of all HIV algorithms compared with Nine Circles Clinic data. “Meds” indicates taking 3 or more HIV medications. “All 3” indicates having an HIV keyword, HIV code and taking 3 or more HIV medications.

Figure 4

Accuracy and performance values of all HIV algorithms compared with a random sample from CPCSSN (Canadian data). “Meds” indicates taking 3 or more HIV medications. “All 3” indicates having an HIV keyword, HIV code and taking 3 or more HIV medications.