Validation of Algorithms to Identify Human Immunodeficiency Virus Cases Using Administrative Data in Manitoba

by

Alexandrea J. Anderson

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Community Health Sciences
University of Manitoba
Winnipeg

Abstract

Introduction: Administrative data are valuable for describing Human Immunodeficiency Virus (HIV) cases and their health outcomes, but it is important to first validate these data to assess their accuracy. To date, most HIV and Acquired Immunodeficiency Syndrome (AIDS) algorithms were validated using USA Medicare/Medicaid data. Two Canadian studies (Ontario and British Columbia) validated HIV algorithms, but the prescription and laboratory data used varied from those available in Manitoba and the reference standard in Ontario and BC were not population-based. The objective of this study was to validate algorithms consisting of physician visit, hospitalization, and antiretroviral prescription data against positive confirmatory HIV laboratory tests to identify Manitobans living with HIV.

Methods: The validation cohort consisted of Manitobans with a valid Personal Health Identification Number and at least three years of continuous health coverage between 2007 and 2018. Positive confirmatory HIV tests from Cadham Provincial Laboratory were the reference standard. Fifteen algorithms requiring two or three years of data were evaluated. Seven measures of accuracy were calculated for each algorithm: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), Youden's J statistic, kappa statistic, and area under the receiver operating characteristic curve (AUC). Four sensitivity analyses were also completed.

Results: The validation cohort included 1,454,010 individuals, of which 1,589 were HIV cases and 1,452,421 were HIV non-cases. Algorithm sensitivity ranged from 81.1% to 96.5%. PPV ranged from 44.1% to 96.0%. Specificity and NPV were very high for all algorithms. Youden's J Statistic ranged from 0.81 to 0.96. Kappa ranged from 0.61 to 0.91. AUC ranged from 0.91 to 0.98. The sensitivity analyses produced similar results.

Conclusion: Different HIV algorithms performed best under different scenarios. One or more physician visits for HIV, one or more hospitalizations for HIV, or two or more antiretroviral prescriptions in two years was best to identify all possible HIV cases, without concern for the number of false positives. Six or more physician visits in two years was best to identify as many true positive HIV cases as possible with minimal false positives. Three or more physician visits in two years most accurately distinguished between HIV cases and HIV non-cases. This study demonstrates that administrative data in Manitoba can accurately identify people living with HIV who interact with the healthcare system.

Acknowledgements

The work presented in this thesis was conducted on Treaty 1 Territory but includes data from all Manitobans who reside on Treaty 1, 2, 3, 4 and 5 Territory and in communities who are signatories to Treaties 6 and 10. Manitoba encompasses the original lands of the Anishinaabeg, Cree, Oji-Cree, Dakota, and Dene peoples and is the homeland of the Métis Nation to which I, and my maternal ancestors, belong.

I would like to sincerely thank my supervisor, Dr. Alyson Mahar, and co-supervisor, Dr. Lisa Lix, for their guidance throughout my MSc program, for their infinite patience with the pace of my progress, and for their support of work/life balance during unprecedented times. They are true mentors and role models. I will always be grateful for what they have taught me.

I would also like to thank my committee members, Dr. Carla Loeppky and Dr. Paul Van Caeseele, for providing constructive feedback throughout the development of my thesis. My thesis would not be complete without their methodological and public health expertise. I owe Dr. Loeppky deep gratitude for taking a chance on me many years ago when I was an enthusiastic undergraduate student with no epidemiologic training whatsoever. My career would not be where it is today without her mentorship.

I appreciate the workspace and resources that were provided to me by the Manitoba Centre for Health Policy. I am thankful for the financial support I received from Dr. Mahar's start-up funds, the Research Manitoba Master's Studentship, the Visual and Automated Disease Analytics (VADA) Training Program Fellowship, and the Canadian Institutes of Health Research (CIHR) Frederick Banting and Charles Best Canada Graduate Scholarship and Tri-Council Top-Up Award, the Master's Award for Indigenous Students, and the Evelyn Shapiro Award for Health Service Research. I am also thankful for the travel awards I received which allowed me to gain valuable experience presenting my research at conferences including the Canadian Society of Epidemiology and Biostatistics Travel Award, the Faculty of Graduate Studies Travel Award, the CIHR and the Institute of Cancer Research's Institute Community Support Travel Award, and the University of Manitoba Graduate Student Association (UMGSA) Travel Award.

The results and conclusions are those of the authors and no official endorsement by the Manitoba Centre for Health Policy, Manitoba Health, or other data providers is intended or should be inferred.

Dedication

My dearest Alex, there are no words to describe how much your support over the past four years has meant to me. I couldn't have done this without you. Thank you for doing seemingly endless amounts of dishes, preparing meals, drying my tears, making me laugh, and taking me on adventures. Thank you. I love you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral therapy
AUC	Area under the receiver operating characteristic curve
BC	British Columbia
BCCDC	BC Centre for Disease Control
BCCfE	BC Centre for Excellence in HIV/AIDS
BCMoH	BC Ministry of Health
CI	Confidence interval
COVID-19	SARS-CoV-2 virus
CPL	Cadham Provincial Laboratory
DIN	Drug Identification Number
DPIN	Drug Program Information Network
H	Inpatient hospitalization related to HIV
HAART	Highly active antiretroviral therapy
HIPC	Health Information Privacy Committee
HIV	Human Immunodeficiency Virus
HREB	Health Research Ethics Board
ICD-9-CM	International Classification of Diseases, 9th Revision, with Clinical Modifications
ICD-10-CA	International Statistical Classification of Diseases and Related Health Problems, 10th Revision, with Canadian Enhancements
LHIV	Advancing Primary Healthcare for Persons Living with HIV in Canada
LIMS	Laboratory Information Management System
MCHP	Manitoba Centre for Health Policy
MHIR	Manitoba Health Insurance Registry
NPV	Negative predicted value
ODB	Ontario Drug Benefit
P	Physician visit related to HIV
PHAC	Public Health Agency of Canada
PHIN	Personal Health Identification Number
PPV	Positive predicted value
RHA	Regional health authority
Rx	Antiretroviral prescription
UNAIDS	Joint United Nations Programme on HIV/AIDS
USA	United States of America
WRHA	Winnipeg Regional Health Authority

Chapter 1 - Introduction

1.1 Background

Human Immunodeficiency Virus (HIV) and Acquired Immunodeficiency Syndrome (AIDS) are global public health concerns. HIV is a sexually transmitted and blood-borne infection; it can be transmitted by sexual contact, sharing injection drug use materials, needle-stick injuries at work, blood transfusions, organ transplants, and perinatally from mother to child (1). If left untreated, HIV will progress to AIDS within five to 10 years and death within three to five years of AIDS diagnosis. (2, 3). Those patients who are healthy at the time of HIV diagnosis and who follow their treatment program exactly as prescribed to achieve viral suppression are expected to have a near-normal life expectancy (4). In 2020, there were an estimated 1.5 million new HIV infections, a total of 37.7 million people living with HIV, and approximately 680,000 deaths from AIDS-related illness, worldwide. (5)

The Joint United Nations Programme on HIV/AIDS (UNAIDS) and the United Nations General Assembly strive to end the AIDS epidemic by 2030. (6-8) To achieve this, there must be rapid commitment to the collection and analysis of HIV/AIDS data to inform where interventions should be targeted. (6-8) Administrative databases are a valuable tool for describing characteristics of people living with HIV and their health outcomes. Given the low prevalence of HIV in the Canadian and Manitoban populations, the use of population-based administrative data allows for the comprehensive inclusion of HIV cases in research contexts. However, given that there are many factors affecting the accuracy of the diagnostic codes submitted for physician billings or inpatient hospitalizations, such as a lack of standardized training for all settings, heterogeneity in record keeping, and variation in physician billing practices, it is important to validate these data prior to using them to identify HIV cases for research (9)

Two Canadian studies, one in Ontario and one in British Columbia (BC), have validated algorithms to identify HIV cases using provincial, population-based, administrative databases, although these two studies have various limitations (9, 10). The Ontario algorithm was validated within a cohort at two primary care clinics where most HIV patients within Ontario were served. (9) Therefore, the algorithm selected for use in Ontario may not be generalizable to health centers or populations with HIV prevalence lower than 20%, or to people living with HIV who access health care less often. (9) The BC algorithm was validated within a cohort that only consisted of

HIV-positive individuals, so it was not possible to assess the specificity of the algorithm. The BC algorithm was also validated over a 15-year period, which may limit its utility for data that comprise a shorter time frame. In addition, there are differences in the availability of data across provinces, most notably regarding the availability of prescription drug and laboratory test data. These limitations make it important to assess the accuracy of the algorithms when applied in Manitoba.

In Manitoba, two potential sources for identifying HIV cases exist; these sources have been linked to the provincial administrative data repository for research: "Advancing Primary Healthcare for Persons Living with HIV in Canada" (LHIV)-Manitoba cohort, a prospective clinical cohort of people living with HIV in Manitoba, (11) and a database of diagnostic and confirmatory HIV tests performed by Cadham Provincial Laboratory (CPL). (12) These two sources of HIV case information have some limitations. LHIV-Manitoba only captures adult HIV-positive patients who have consented to participate, and deceased HIV patients. (11) Therefore, the cohort does not capture all HIV cases living in the province. The age, sex, and ethnicity distributions of cases captured in LHIV-Manitoba are known to vary significantly from those of all people diagnosed with HIV in Manitoba. (11) Because LHIV-Manitoba does not contain information on HIV non-cases, to compare outcomes for people living with HIV to people without HIV, it will be necessary to link to an additional data source that contains the outcomes for HIV non-cases. Regarding the CPL data, HIV tests completed prior to January 1, 2007 do not have a valid identifier so they cannot be linked to the other administrative datasets to perform longitudinal research on people whose only positive laboratory tests for HIV exist prior to 2007. (12) While the proportion of HIV cases with valid identifiers that can be linked to the other administrative data sources has increased since January 1, 2007, it has not reached 100%. (12) Therefore, a validated algorithm that uses administrative data sources such as physician billings, hospitalizations, and prescriptions can fill the gaps left by LHIV-Manitoba and the CPL testing data.

Given the limitations in existing Canadian algorithm studies, the LHIV-Manitoba clinical cohort of people living with HIV, and the CPL testing data, there is value in validating administrative data algorithms to comprehensively identify HIV cases and HIV non-cases living in Manitoba. The validated algorithms may be used to study progression of disease, comorbid illnesses, long-term outcomes, access to care and treatment, or any other HIV-related research with

the aim of improving the wellbeing of people living with HIV and contributing to the UNAIDS' and the United Nations General Assembly's goal of ending the AIDS epidemic by 2030. (6-8)

1.2 Purpose and Objective

The purpose of this study was to determine the best algorithm, or algorithms, for identifying HIV cases using administrative data in Manitoba. The objective of this study was to validate algorithms consisting of physician visit, hospitalization, and antiretroviral prescription data against positive (or non-negative, non-indeterminate) HIV laboratory tests (which includes tests that identified HIV antibody, HIV DNA, or the HIV virus itself, and tests for HIV drug resistance) to identify Manitobans living with HIV.

1.3 Thesis Organization

Chapter 2 provides a brief introduction to the epidemiology of HIV followed by a literature review of studies which have developed and/or validated HIV algorithms using administrative health data and a review of existing methods for identifying HIV cases in Manitoba. Chapter 3 describes the validation cohort, the application of the algorithms, and compares the performance of the algorithms. Chapter 4 presents the results of the HIV algorithm application and Chapter 5 discusses the results and implications of the study as well as opportunities for further research.

Chapter 2 - Literature Review

2.1 Introduction

This chapter provides the setting and context for this study, beginning with a description and brief epidemiology of HIV (Section 2.2). Section 2.3 provides a review of studies which have developed and/or validated HIV case definitions for administrative health data and a review of existing methods for identifying HIV cases in Manitoba. The final section summarizes the rationale for an HIV algorithm validation study in Manitoba (Section 2.4).

2.2 An Overview of HIV Infection

2.2.1 What is HIV?

HIV is a sexually transmitted and blood-borne infection caused by the human retroviruses HIV-1 and HIV-2 (13). HIV-1 is the most common type of HIV in North America, with HIV-2 found almost exclusively in West Africa (13). HIV may be diagnosed by laboratory tests which identify either the HIV antibody or the HIV virus itself (14). HIV is most often diagnosed by enzyme immunoassay (EIA) HIV antibody test (14, 15). Advancements in HIV testing, such as point-of-care tests, have provided access to HIV testing in remote locations and to transient populations who previously could not be connected with their results (14, 15).

HIV can be transmitted through various channels including: sexual contact; sharing injection drug use materials or using other non-sterile equipment to puncture the skin (e.g. when tattooing or piercing); occupational exposure; blood transfusions and tissue or organ transplantations occurring in Canada between 1978 and 1985 or in countries where screening of blood is not performed; and mother-to-child transmission during the perinatal period (1).

If untreated, HIV infection results in progressive destruction of CD4+ T lymphocytes (CD4 cells), leading to immunodeficiency, chronic inflammation, and vulnerability to opportunistic infections (13, 16). A patient with HIV is said to have progressed to AIDS once 1) they develop one or more indicator diseases (e.g. invasive cervical cancer, Kaposi's sarcoma, recurrent Salmonella septicemia), 2) they have a CD4 cell count of <200 cells/microliter, or 3) their CD4 cells make up < 14% of their total lymphocytes (16). Most patients with untreated HIV will develop AIDS within 5 to 10 years, and die within three to five years of their AIDS diagnosis (2, 3).

AIDS was first identified in 1981 based on an unusual increase in Kaposi sarcoma and *Pneumocystis* pneumonia cases and was believed to be caused by the lifestyle and behavioural factors of men who have sex with men (17). It was not until 1983 that HIV was identified as the causative agent of AIDS (17). The first treatment for HIV, introduced in 1987, was able to delay the onset of AIDS but did not completely stop progression of the disease (17).

Highly active antiretroviral therapy (HAART), introduced in 1996, marked the beginning of the current era where HIV has essentially become a manageable chronic disease (18). Today, HAART is known simply as antiretroviral therapy (ART). Adherence to ART is an effective method of preventing HIV transmission. People living with HIV who have a suppressed (undetectable) viral load are not able to transmit the infection to others (untransmissible) (19). Those patients who are healthy at the time of HIV diagnosis, and who follow their drug program exactly as prescribed to achieve viral suppression, are expected to have a near-normal life expectancy (4).

2.2.2 Descriptive Epidemiology

Globally, HIV and AIDS are major public health concerns (6, 20, 21) with the deadly AIDS pandemic still ongoing forty years after the first AIDS case was discovered. (22, 23) In 2015, the world leader in the AIDS movement, UNAIDS, released a global strategy outlining 10 targets to meet by 2020 in order to end the AIDS pandemic by 2030. (21) The first target is known as the HIV Cascade of Care (or Continuum of Care) and outlines the goal to have 90% of people living with HIV know their status, 90% of people who know their status to be adhered to ART treatment, and 90% of people on ART to have suppressed viral loads (<200 copies/mL). (21) Progress on this target is widely measured and reported.

2.2.2.1 Global Epidemiology

In 2020, UNAIDS estimated there were 37.7 million people living with HIV, worldwide. Of these, 84% knew their HIV status, leaving 6.1 million people unaware that they were living with HIV. (5) Among the people with HIV who knew their status, 87% were on ART. Among the people on ART, 90% were virally suppressed. Of all people living with HIV (aware and unaware of their status), 66% were virally suppressed. This means there were 12.4 million people worldwide able to transmit their HIV infection in 2020. (5)

There were an estimated 1.5 million new HIV infections in 2020. (5) The incidence of HIV infection varied across the world regions with 670,000 new infections in eastern and southern Africa (146 new cases per 100,000 population) compared with 67,000 new infections in western and central Europe and North America (7 new cases per 100,000 population). (5, 24) In 2019, HIV/AIDS was the ninth leading cause of death in low-income countries. (25) In 2020, approximately 680,000 people died of AIDS-related illness, worldwide. (5)

2.2.2.2 Canadian Epidemiology

In 2018, which is the most current Canadian HIV Cascade of Care data available at the time of writing, the Public Health Agency of Canada (PHAC) estimated there were 62,050 Canadians living with HIV/AIDS. (26) Of these, 87% were aware of their HIV status, leaving 8,300 Canadians (or 1 in 7 HIV cases) unaware of their HIV infection status. (26) Among the people with HIV who knew their status, 85% were on ART. Among the people on ART, 94% were virally suppressed. Of all people living with HIV (aware and unaware of their status), 70% were virally suppressed. This means there were approximately 18,700 people in Canada able to transmit their HIV infection in 2018. (26)

There were 1,639 people diagnosed with HIV (4.3 new HIV diagnoses per 100,000 population), in Canada, in 2020. (27) This represents a 21% decrease in cases from the 2,122 cases diagnosed in 2019, although this decrease may be due to a decreased testing performed because of the SARS-CoV-2 virus (COVID-19) pandemic. (27, 28) The number of cases diagnosed is different than the number of incident or new cases because not every incident case is diagnosed in the year of infection. Males had a greater HIV diagnosis rate than females in every year of data available since 1996 and accounted for approximately 7 in 10 newly diagnosed cases in 2020. (27, 28) Over half of new diagnoses were among people aged 20-39 years in 2020. (27) In 2020, HIV was ranked the thirty-second leading cause of death in Canada, with 135 HIV-related deaths reported (age-specific mortality rate of 0.4 deaths per 100,000 population). (29)

The HIV diagnosis rate varied by geographic region within Canada. Saskatchewan documented the highest rate with 15.7 new HIV diagnoses per 100,000 population in 2020, followed by Manitoba with 7.0 new HIV diagnoses per 100,000 population. (27) The lowest rates were in Atlantic Canada with 1.3 new diagnoses per 100,000 population. (27)

2.2.2.3 Manitoba Epidemiology

Fransoo et al. (30) determined that there were approximately 1,267 people living with HIV in Manitoba at the end of 2017. The Manitoba HIV Program reported that over 9 in 10 of the 115 HIV clients new to care in 2018 were linked to an HIV care provider and three-quarters of those clients were retained in care (31). Of clients retained in care, 96.3% were adhered to ART treatment at the time of audit in 2019 (31). Of those adhered to ART treatment, 89.7% achieved viral suppression (31). This means approximately 6 in 10 of the clients who entered the care of the Manitoba HIV Program in 2018 achieved viral suppression by 2019, although the proportion of clients who achieved viral suppression in this cohort was expected to increase over time (31).

Manitoba Health reported 117 new HIV diagnoses in 2020 (8.4 cases per 100,000 population). (32) This number is higher than the number of newly diagnosed Manitoban cases reported by PHAC in 2020 because it includes people with known HIV infection who migrated into the province, whereas PHAC only captures first time diagnoses. (27, 32) This is a slight decrease from the 119 cases diagnosed in 2019, but this decrease is likely due to 11,000 fewer individuals being screened for HIV in Manitoba during 2020, due to the COVID-19 pandemic. (32)

The HIV diagnosis rate varied by geographic region within Manitoba. In 2020, Winnipeg Regional Health Authority (RHA) accounted for most of the newly diagnosed cases (68% of cases, 10.1 cases per 100,000 population) although the Northern Health Region had the highest crude rate with 11.6 new cases per 100,000 population. Interlake-Eastern RHA had the lowest crude rate (3.7 cases per 100,000 population). (32)

2.3 Using Administrative Data to Identify HIV Cases

UNAIDS and the United Nations General Assembly strive to end the AIDS epidemic by 2030, but swift action to reduce inequalities faced by people living with HIV is required to meet the goal of eradication within the next decade. (6-8) One of the key commitments agreed to by the United Nations General Assembly on June 9, 2021 was to, “Strengthen and enhance the use of data, innovation, research and development, and science and technology to accelerate the end of AIDS.” (8) This commitment supports the collection and analysis of data to inform where inequalities are occurring so that targeted interventions can be planned. (8)

Much research that examines health outcomes for people living with HIV do so by using a clinical cohort of people diagnosed with HIV. (33) Clinical registries of this type may be limited in the type of people they capture (e.g., only patients from a small number of clinics). (9, 11, 33) Furthermore, establishing and sustaining the relationships required to maintain and continually update a clinical registry can be onerous. (33) The process for linking clinical registries to existing warehouses of administrative data to perform longitudinal research can be time consuming, delaying the time to research. (33) Then, unless agreements are established to feed the continual updates to the clinical registry into the administrative data warehouse, the list of HIV cases will become quickly outdated. For these reasons, there is a benefit to identifying HIV cases using administrative data for research purposes.

Administrative databases are a valuable tool to capture and analyze data pertaining to people living with HIV. Administrative health databases contain information related to the delivery of health services such as physician visits, hospital admissions and discharges, and prescription dispensations. (34) Administrative databases are not limited to health data; they may also contain a broad range of information related to education, housing, income, access to social supports, and more through linkage to surveys, clinical registries, and population registries. (34) While administrative data are collected for the purposes of service delivery and record keeping, they are increasingly used for secondary purposes such as research and surveillance. (34) The longitudinal nature of administrative data lends itself well to monitoring the health of populations over time. (34)

HIV is a chronic infection that requires multiple contacts with the healthcare system to diagnose, treat, and manage long-term, so a combination of data sources, types of healthcare encounters, and diagnostic codes may be used to reduce misclassification error. (9) The validation of administrative data for use in population-based research has been identified as a priority by an international consortium of health services researchers. (35)

Since 1987, a range of algorithms for identifying HIV/AIDS in administrative data have been developed and validated in North America (Table 2.1). Early algorithms relied on the presence of opportunistic infections to diagnose AIDS but, following the advent of ART in 1996 and the resulting reduction in opportunistic infections, the focus of the algorithms shifted to identifying HIV cases. The increased reliance on administrative health data to identify cases of HIV and AIDS coincides with when administrative data became more routinely available for use

in health research. (36, 37) The majority of administrative data algorithms were created using Medicare and Medicaid data in the United States of America (USA) (38-46). Two Canadian studies, one in Ontario and one in BC, have validated algorithms to identify HIV cases using provincial, population-based, administrative databases (9, 10).

In 2011, Antoniou et al. validated an algorithm for identifying HIV cases in Ontario, Canada using data held by ICES (9). Forty-eight algorithms consisting of a combination of physician billing information, hospitalization records, emergency department visits, and pharmacy records from the Ontario Drug Benefit (ODB) program, were tested. The algorithms were compared over two- and three-year periods against a reference standard of chart abstractions from a random sample of medical records from two primary care clinics in downtown Toronto where approximately 20% of patients were known to have HIV. HIV visits were identified using International Classification of Diseases, 9th Revision, with Clinical Modifications (ICD-9-CM) codes: 042, 043, and 044 and International Classification of Diseases, 10th Revision, with Canadian Enhancements (ICD-10-CA) codes: B20-B24.

The sensitivity, specificity, kappa statistic, and area under the receiver operating characteristic curve (AUC) were estimated over two- and three-year time frames for each of the 48 algorithms. For all algorithms, the kappa statistic was greater than 0.9 while the specificity (with the exception of algorithms consisting of a single physician billing claim) was greater than 99%. Among the eight algorithms with the best performance, specificity ranged from 99.4% (95% confidence interval [CI] 98.8%-99.7%) to 99.7% (95% CI 99.3% - 99.9%); sensitivity ranged from 90.7% (95% CI 87.7%-93.1%) to 93.4% (95% CI 90.8%-95.5%); kappa ranged from 0.93 (95% CI 0.91-0.95) to 0.94 (95% CI 0.91-0.96) and AUC ranged from 0.952 to 0.963 when applied over a two-year time frame. When applied over a three-year time frame, the sensitivity of each algorithm improved, ranging from 96.2% (95% CI 94.0%-97.7%) to 97.9% (95% CI 96.1%-99.0%), with minimal loss to specificity, ranging from 99.1% (95% CI 98.5%-99.5%) to 99.6% (95% CI 99.1%-99.8%), with small improvements seen in kappa and AUC, ranging from 0.96 (95% CI 0.95-0.98) to 0.97 (95% CI 0.95-0.98) and 0.978 to 0.985, respectively.

The algorithm that required at least three physician billing claims within three consecutive years, with specificity of 99.6% (95% CI: 99.1%-99.8%), sensitivity of 96.2% (95% CI: 94.0%-97.7%), kappa statistic of 0.97 (95% CI: 0.95-0.98), and AUC of 0.979, was selected as the best algorithm of the 48 tested to identify individuals living with HIV using administrative data in

Ontario, by Antonioiu et al. (9) With 48 algorithms tested over two time periods, this study represents the most comprehensive effort to validate an administrative data algorithm for identifying HIV status to date. The greatest limitation of this study was that the reference standard data were sampled from two primary care clinics where most HIV patients within Ontario were served. This sample is not population based because it had a higher prevalence of HIV than the general population and only consisted of HIV patients who were engaged in care. Therefore, the selected algorithm may not be generalizable to health centers, or populations, with HIV prevalence lower than 20%, or to people living with HIV who access health care less often.

In 2013, Nosyk et al., (10) built on the work of Antoniou et al. (9) and validated an algorithm for identifying HIV cases in BC, Canada using data held by the BC Ministry of Health (BCMoH), the BC Centre for Disease Control (BCCDC), and the BC Centre for Excellence in HIV/AIDS (BCCfE) (10). Four algorithms were chosen from the 48 developed by Antoniou et al. (9) and applied over a 15-year study period.

Nosyk et al. (10) defined confirmed HIV cases as those individuals who had either a) an HIV-positive test in the BCCDC surveillance database or b) a viral load test with detectable viral load, a CD4 test, or an HIV-related antiretroviral dispensation in the BCCfE database. Unconfirmed HIV cases were those individuals who only had HIV-related visits in the BCMoH's Medical Services Plan Database or Discharge Abstract Database. HIV-related visits were identified in the BCMoH databases using ICD-9-CM codes: 042, 043, 044, V08, 795.71, and 795.8 and ICD-10-CA codes: B20-B24, R75, and Z21. Reference standard cases (referred to as gold standard cases by Nosyk et al. (10)) were confirmed HIV cases who had linkable records in each of the BCCfE, BCCDC, and BCMoH databases.

The sensitivity of the algorithms compared to the reference standard was estimated for the four algorithms. In addition, Nosyk et al. (10) developed five *a priori* hypotheses to assess the face validity of the algorithms by evaluating the HIV non-cases identified by each algorithm with regard to: 1) the proportion that were female, 2) all-cause mortality rate, 3) the rate of physician visits, 4) the rate of inpatient hospitalizations, and 5) the rate of prescription dispensation.

The algorithm that required at least three HIV-related physician visits or at least one HIV-related hospital admission was chosen as the best algorithm of the four tested by Nosyk et al. (10) This algorithm identified 17.6% of the unconfirmed HIV cases, had a sensitivity of 88%, and had satisfactory results for each of the five hypotheses. (10) For comparison, Antoniou et al. (9)

reported that their algorithm requiring at least three physician billing claims or at least one hospitalization within three consecutive years had sensitivity of 96.2% (95% CI: 94.0%-97.7%).

One limitation of the Nosyk et al. (10) study was that the validation cohort only consisted of HIV-positive individuals, so it was not possible to assess the specificity of the algorithm. A second limitation is that the algorithms were validated over a 15-year period, which may limit their utility for data that comprise a shorter time frame.

The selection of a "best" algorithm is subjective and should consider availability of local data and simplicity (fewer datasets and shorter time periods), in addition to algorithm performance. Antoniou et al. and Nosyk et al. chose different algorithms to identify HIV cases because they used different algorithm selection criteria. Antoniou et al., "selected the algorithm with the highest specificity while maximizing sensitivity over the shortest interval of time" (9). Nosyk et al. (10) did not calculate specificity due to not having HIV-negative cases within their reference standard so instead selected the best algorithm based on the number of unconfirmed HIV cases identified by the algorithm, the sensitivity of the algorithm when compared to the reference standard, and the results of the five *a priori* hypotheses.

The administrative data available in Manitoba differs from the data available in Ontario and BC, most notably regarding the availability of prescription drug and laboratory test data, which are critical to measuring HIV status. In Manitoba, the Drug Program Information Network (DPIN) captures prescriptions dispensed to all insured residents, with the exception of medications provided through hospital pharmacies, nursing stations, ward stock, and outpatient visits at CancerCare Manitoba (47). In Ontario, the ODB pharmaceutical data only includes prescriptions from a subset of the Ontario population (those over the age of 65, recipients of social assistance, and individuals who have high prescription drug costs relative to their income) (9) In BC, the BCCfE Drug Treatment Program captures information on all insured residents who receive ART for the treatment of HIV infection in BC (48). In Manitoba, CPL maintains a database of all HIV laboratory tests performed in the province that are available through the Manitoba Centre for Health Policy (MCHP) (49). In BC, the BCCDC surveillance database captures all individuals with a positive HIV laboratory test performed in the province but does not capture negative tests (10). However in Ontario, linkage to the Ontario Laboratory Information System data at ICES did not occur until 2016, so diagnostic and viral load HIV tests performed by the Public Health Ontario Laboratory were not available at the time Antoniou et al. (9) developed and validated their

algorithm (50, 51). These data differences, along with the limitations of the Ontario and BC HIV algorithm validation studies, makes it important to assess the accuracy of algorithms when applied in Manitoba.

The availability of an HIV administrative data algorithm validated for use in a specific population helps to increase HIV research within that population. For example, since validation in 2011, the Ontario HIV administrative data algorithm validated by Antoniou et al. (9) has been used to study a broad range of topics from HIV incidence, mortality, prevalence, and the HIV Cascade of Care to socioeconomic and sex-related disparities in hospital admission rates among patients with HIV, comorbidities among people living with HIV, cancer screening among people living with HIV, prenatal, postnatal, and neonatal care and adverse outcomes among women living with HIV, how healthcare is delivered to people living with HIV, continuity of HIV care during and after release from prison, and HIV drug resistance testing. (52-71)

2.4 Conclusion

Given the limitations of existing Canadian algorithms, the differences in the administrative data sources found in Manitoba, Ontario, and BC, the limitations of the LHIV-Manitoba clinical cohort, and the limitations of the CPL testing data, it is important to assess the accuracy of administrative data algorithms consisting of physician visit, hospitalization, and antiretroviral prescription data against positive confirmatory HIV laboratory tests to determine which of the algorithms may be used to comprehensively identify HIV cases and HIV non-cases living in Manitoba and fill in the gaps of the reference standard CPL HIV test data.

The validated algorithms may be used to study progression of disease, comorbid illnesses, long-term outcomes, access to HIV care and treatment, or any other HIV-related research with the aim of improving the wellbeing of people living with HIV and contributing to the UNAIDS' and the United Nations General Assembly's goal of ending the AIDS epidemic by 2030. (6-8) This study seeks to validate case-finding algorithms using Manitoba's administrative data to identify Manitobans living with HIV using the algorithms developed by Antoniou et al. (9)

Table 2.1 Studies that developed and/or validated algorithms to identify HIV/AIDS cases using administrative data

Year Published	Authors	Title	Disease	Location	Administrative Data Used in Algorithm	Time Period
1987	Keyes, M. et al. (72)	A Comparison of Two Methods of Identifying AIDS Cases in Medical Claims Analyses (Conference Paper)	AIDS	Unknown	Unknown	Unknown
1988	Andrews, R. et al. (38)	Acquired Immunodeficiency Syndrome in California's Medicaid Program, 1981-84	AIDS	California, United States of America (USA)	Medicaid claims	1981-1984
1991	Keyes, M. et al. (39)	A Methodology for Building an AIDS Research File Using Medicaid Claims and Administrative Data Bases	AIDS	California and New York, USA	Medicaid claims	California: Jan. 1982-Dec. 1986. New York: Oct. 1982-Sept. 1987
1993	Turner, B. et al. (73)	Benefits of Shared Care for Advanced HIV-Infection (Conference Paper)	AIDS	Unknown	Unknown	Unknown
1995	Fanning, T. et al. (40)	The Quality of Medicaid Data for HIV/AIDS Research: Examination of a Statewide Data Base	AIDS	New York, USA	Medicaid claims	1983-1990
1997	Thornton, C. et al. (41)	Methods for Identifying AIDS Cases in Medicare and Medicaid Claims Data	AIDS	USA	Medicaid and Medicare claims	California Medicaid: 1991-1992 Medicare: 1991-1993

Year Published	Authors	Title	Disease	Location	Administrative Data Used in Algorithm	Time Period
1998	Fasciano, N. et al. (42)	Profile of Medicare Beneficiaries with AIDS: Application of an AIDS Case finding Algorithm	AIDS	USA	Medicare claims	1991-1993
2001	Mrus, J. et al. (43)	Development of an HIV Research Database Using Medicaid Claims Data	HIV/AIDS	Ohio, USA	Medicaid claims	June 1997-Sept. 1998
2004	Walkup, J. et al. (46)	Sensitivity of an AIDS Case-Finding Algorithm: Who Are We Missing?	AIDS	New Jersey, USA	Medicaid claims	1992-1998
2006	Fultz, S. et al. (74)	Development and Verification of a "Virtual" Cohort Using the National VA Health Information System	HIV	USA	Electronic medical records	1998-2003
2011	Antoniou, T. et al. (9)	Validation of Case-Finding Algorithms Derived from Administrative Data for Identifying Adults Living with Human Immunodeficiency Virus Infection	HIV	Ontario, Canada	Physician billing claims, hospital abstracts, emergency department visits, and pharmacy records	2005/06-2007/08
2011	Chesnut, T. et al. (44)	An Expenditure Analysis of High-Cost Medicaid Recipients with HIV Disease in New York State	HIV	New York, USA	Medicaid claims	2003-2007

Year Published	Authors	Title	Disease	Location	Administrative Data Used in Algorithm	Time Period
2013	Nosyk, B. et al. (10)	Application and Validation of Case-Finding Algorithms for Identifying Individuals with Human Immunodeficiency Virus from Administrative Data in British Columbia, Canada	HIV	British Columbia, Canada	Physician billing claims and hospital abstracts	1995-2010
2014	Felsen, U. et al. (75)	Development of an Electronic Medical Record-Based Algorithm to Identify Patients with Unknown HIV Status	HIV	Montefiore Medical Center, New York, USA	Electronic medical records	2008-2012
2014	Goetz, M. et al. (76)	Development and Validation of an Algorithm to Identify Patients Newly Diagnosed with HIV Infection from Electronic Health Records	HIV	Los Angeles, Houston, Washington DC, and Atlanta, USA	Electronic medical records	2000-2012
2015	Leibowitz, A. et al. (45)	Identifying a Sample of HIV-Positive Beneficiaries from Medicaid Claims Data and Estimating their Treatment Costs.	HIV	California, USA	Medicaid and Medicare claims	2007
2018	Paul, D. et al. (77)	Development and Validation of an Electronic Medical Record (EMR)-Based Computed Phenotype of HIV-1 Infection	HIV	Durham and North Carolina, USA	Electronic medical records	2007-2011

Chapter 3 - Methods

3.1 Overview

This chapter begins with the study objective and hypothesis (Section 3.2), then describes the study design (Section 3.3), research setting, data sources, and data management (Section 3.4). The chapter then focuses on the methods to validate the HIV administrative data algorithms (Section 3.5). The chapter concludes by describing ethical considerations (Section 3.6).

3.2 Objectives and Hypotheses

The objective of this study was to validate algorithms consisting of physician visit, hospitalization, and antiretroviral prescription data against positive confirmatory HIV laboratory tests to identify Manitobans living with HIV. The review of the literature (Chapter 2) informed the hypothesis that an algorithm consisting of three physician visits in three years will have the highest accuracy for identifying people living with HIV.

3.3 Study Designs and Reporting Guidelines

A validation study design was used to validate an algorithm to identify Manitobans living with HIV using administrative health data. The validation study was reported according to Benchimol et al.'s (78) Checklist of Reporting Criteria for Studies Validating Health Administrative Data Algorithms.

3.4 Research Setting, Data Sources and Data Management

The study objective was achieved using data from the Manitoba Population Research Data Repository (the Repository) housed at MCHP, a population health data centre established in 1991. (37) The Repository is a comprehensive collection of routinely collected administrative, registry, survey, and clinical data about nearly all Manitoba residents. (36, 37) The coverage for some datasets extends to 1970 with new data sources becoming available over time. (37) The Repository data are de-identified but can be linked using a scrambled, Personal Health Identification Number (PHIN), allowing for an individual's interactions with the health system to be tracked over time. (33, 36, 79) Although data in the Repository were generated through the routine administration of programs and services, the quality of the data is closely monitored, providing a valuable source of population-level data for research purposes. (80, 81) All provincially insured health services

provided to Manitoba residents with a valid Manitoba Health card are captured in the Repository, making it a very rich and unique data source for conducting health services research. (37)

This study built on the ability to perform deterministic linkages between the databases in the Repository including CPL databases, DPIN, Hospital Abstracts, the Manitoba Health Insurance Registry (MHIR), Medical Claims/Medical Services, and Statistics Canada's Census data. (82)

Cadham Provincial Laboratory. The CPL provides public health laboratory services in Manitoba. Two CPL databases are housed at MCHP. The CPL Mainframe contains historical data from 1992/93 to July 2010. (83) CPL Mainframe data are not complete for 2009/10, the number of records decreases toward the end of the fiscal year. (83) The CPL Laboratory Information Management System (LIMS) is updated annually and contains data from August 31, 2009 to May 31, 2019. (49) The CPL databases include testing information related to microbiology, parasitology, screening, virology, newborn screening, and public health chemistry and quality assurance. (49, 83) The CPL LIMS contains information related to the patient, the sample(s) submitted, the test(s) performed, and the test result(s). (49) The CPL Mainframe database contains similar information to CPL LIMS. Almost all confirmatory HIV testing in Manitoba is performed by CPL. (15, 84) Prior to January 1, 2007, all HIV testing in Manitoba was performed non-nominally; (84) all HIV tests were ordered using a code known only to the ordering healthcare provider, so only the ordering healthcare provider could link the test results to the individual's health care record. (16) As a result, HIV laboratory tests performed prior to January 1, 2007 cannot be linked to other databases in the Repository via the individuals' scrambled PHIN. (12) After January 1, 2007, the number of HIV cases with valid identifiers that can be linked to the other administrative data sources has increased, but it has not reached 100%. (12)

Drug Program Information Network. DPIN is an annually-updated database capturing information on all prescriptions with a Drug Identification Number (DIN) dispensed from community pharmacies to Manitoba residents between 1995/96 and 2018/19, regardless of insurance coverage or final payer. (47) The DPIN database contains detailed information on each prescription dispensed including the DIN, the generic name, the brand name, the strength, the quantity dispensed, the cost, and the date of dispensation (85). The reason for receiving the prescription is not captured in the database and neither is information on prescriptions which were never filled. (47) Medications received through hospital pharmacies, nursing stations, ward stock, and outpatient visits at CancerCare Manitoba are also not captured in DPIN. (47)

Hospital Abstracts. The Hospital Abstracts database is updated annually and contains clinical information, including diagnosis and procedure codes, for patients discharged from Manitoba hospitals between 1970/71 and 2018/19. (86) These records include five-digit ICD-9-CM diagnosis codes (from January 1, 2000 to March 31, 2004) with up to 16 codes listed per record and five-digit ICD-10-CA diagnosis codes (from April 1, 2004 onward) with up to 25 codes listed per record. (86, 87)

Medical Claims/Medical Services. The Medical Claims/Medical Services database is updated annually and contains billing information for healthcare providers (e.g. physicians, nurse practitioners) who provided service in physician offices, hospitals, and outpatient departments between 1970/71 and 2018/19. (88) Each Medical Claims/Medical Services claim is associated with a single ICD-9-CM diagnosis code and a single tariff from January 1, 2000 to December 31, 2018. (88) For data prior to 2015/16, the diagnosis code is recorded at the three-digit level, after 2015/16 a five-digit diagnosis code is available for some records. (88)

Manitoba Health Insurance Registry. The MHIR is updated semi-annually and contains information on all individuals registered with the provincial health insurance plan including individual-level demographics, birth dates, and migration in/out of province. (89) The MHIR contains information on all individuals registered with Manitoba Health from 1970/71 to June 2019, allowing for longitudinal analyses. (33, 36, 89)

Canadian Census. Statistics Canada's Census data includes aggregate demographic information collected every five years from 1971 to 2016. (90) Demographics captured include the distribution of age, sex, marital status, employment, and income levels within small geographic areas known as dissemination areas. (90) Dissemination areas are currently the smallest geographic unit for which census data are released by Statistics Canada. (90) This means that health databases may be linked to the census data at the geographic-level, rather than the person-level.

3.5 Validating an HIV Administrative Data Algorithm

3.5.1 Construction of Validation Cohort

The validation cohort consisted of Manitobans with a valid PHIN and at least three years of continuous health coverage in the MHIR during the study observation period, which was from January 1, 2007 to December 31, 2018. HIV status was identified by linking the validation cohort with the CPL databases and searching for HIV laboratory tests for these individuals. HIV

laboratory tests in the CPL databases were considered the reference standard because, since 2007, if an HIV patient was connected to care, they should have had a nominal diagnostic or follow-up laboratory test performed that would confirm their HIV status. Patients with a positive (or non-negative, non-indeterminate) result on an HIV laboratory test (which includes tests that identified HIV antibody, HIV DNA, or the HIV virus itself, and tests for HIV drug resistance) in the CPL databases during the study observation period were classified as HIV cases. All other members of the validation cohort, including those individuals with negative or indeterminate HIV laboratory tests and those individuals who did not have an HIV CPL laboratory test, were classified as HIV non-cases. The method to identify HIV laboratory tests is described in Section 3.5.5.1. This cohort was used to validate both the two- and three-year algorithms, following the methods used by Lix et al. (91) for validating algorithms that require varying years of administrative data.

3.5.2 Algorithms Selected for Validation

Antoniou et al. (9) developed and validated 48 algorithms to identify individuals living with HIV using administrative health data held by ICES, in Ontario. A combination of physician billing information, inpatient hospitalization records, emergency department visits, pharmacy records, and health insurance registry data were used to construct the algorithms. The reference standard consisted of primary care chart review data from two downtown Toronto clinics. The number of physician billings, hospitalizations, emergency department visits, and/or prescriptions required per patient varied for each of the 48 algorithms. The sensitivity, specificity, kappa statistic, and area under the receiver operating characteristic curve (AUC) were calculated over two- and three-year time frames for each of the 48 algorithms. The algorithm that required at least three physician billing claims within three consecutive years, with specificity of 99.6% (95% CI: 99.1%-99.8%), sensitivity of 96.2% (95% CI: 94.0%-97.7%), kappa statistic of 0.97 (95% CI: 0.95-0.98), and AUC of 0.979, was selected as the best algorithm to identify individuals living with HIV using administrative data in Ontario.

Fifteen of the 48 algorithms developed and validated by Antoniou et al. (9) were chosen to validate in Manitoba. The excluded 33 algorithms required either HIV physician billing codes specific to Ontario or emergency department visit data. In Manitoba, emergency department visit data are not collected routinely for residents outside of Winnipeg. For each of the 15 selected algorithms, the number of 1) HIV-related physician billing claims from the Medical

Claims/Medical Services database (described in Section 0), 2) HIV-related inpatient hospitalizations from the Hospital Abstracts database (described in Section 0), and 3) ART prescriptions from the DPIN database (described in Section 3.5.4), varied. Each of the 15 algorithms were validated over both two- and three-year periods. The resulting 30 algorithms that were validated are described in Table 3.1. It should be noted that Algorithms 1 and 16 will produce the same results because the entire period is used to ascertain a single physician claim. Algorithms 7 and 22 will also produce the same results because the entire period is used to ascertain a single physician claim or a single hospitalization.

The combinations of physician billing claims, inpatient hospitalizations and prescriptions required by the HIV case ascertainment algorithms chosen for validation in Manitoba were unchanged from those originally developed and validated by Antoniou et al. (9), in 2011. That said, updates were made to the list of diagnosis codes used to identify HIV-related physician billings and inpatient hospitalizations, and to the list of DINs used to identify HIV ART prescriptions, as described in Sections 0 and 3.5.4.

Table 3.1 HIV case ascertainment algorithms validated in Manitoba

Algorithm #	Case Definition[a]
1	1 or more physician claims in 2 years
2	2 or more physician claims in 2 years
3	3 or more physician claims in 2 years
4	4 or more physician claims in 2 years
5	5 or more physician claims in 2 years
6	6 or more physician claims in 2 years
7	1 or more physician claims in 2 years or 1 or more hospital discharges in 2 years
8	2 or more physician claims in 2 years or 1 or more hospital discharges in 2 years
9	3 or more physician claims in 2 years or 1 or more hospital discharges in 2 years
10	1 or more physician claims in 2 years or 2 or more antiretroviral prescriptions (Rx) in 2 years
11	2 or more physician claims in 2 years or 2 or more Rx in 2 years
12	3 or more physician claims in 2 years or 2 or more Rx in 2 years

Algorithm #	Case Definition[a]
13	1 or more physician claims in 2 years or 1 or more hospital discharges in 2 years or 2 or more Rx in 2 years
14	2 or more physician claims in 2 years or 1 or more hospital discharges in 2 years or 2 or more Rx in 2 years
15	3 or more physician claims in 2 years or 1 or more hospital discharges in 2 years or 2 or more Rx in 2 years
16	1 or more physician claims in 3 years
17	2 or more physician claims in 3 years
18	3 or more physician claims in 3 years
19	4 or more physician claims in 3 years
20	5 or more physician claims in 3 years
21	6 or more physician claims in 3 years
22	1 or more physician claims in 3 years or 1 or more hospital discharges in 3 years
23	2 or more physician claims in 3 years or 1 or more hospital discharges in 3 years
24	3 or more physician claims in 3 years or 1 or more hospital discharges in 3 years
25	1 or more physician claims in 3 years or 2 or more antiretroviral Rx in 3 years
26	2 or more physician claims in 3 years or 2 or more Rx in 3 years
27	3 or more physician claims in 3 years or 2 or more Rx in 3 years
28	1 or more physician claims in 3 years or 1 or more hospital discharges in 3 years or 2 or more Rx in 3 years
29	2 or more physician claims or 1 or more hospital discharges in 3 years or 2 or more Rx in 3 years
30	3 or more physician claims in 3 years or 1 or more hospital discharges in 3 years or 2 or more Rx in 3 years

[a]Subset of algorithms developed and tested by Antoniou et al. (9)

3.5.3 Diagnosis Codes for Algorithm Validation

This study used the ICD diagnosis codes listed in Table 3.2 to identify HIV-related physician billing claims and inpatient hospitalizations in the Medical Claims/Medical Services and Hospital Abstracts databases, respectively. The ICD codes used by Antoniou et al. (9) (ICD-9-CM: 042-044; ICD-10-CA: B20-B24) formed the basis of this list. There were four additional ICD codes used by Nosyk et al. (10) (ICD-9-CM: 042-044, V08, 795.71; ICD-10-CA: B20-B24, R75, Z21). Finally, a review of ICD-9-CM and ICD-10-CA documentation revealed three additional HIV-related diagnosis codes (ICD-9-CM: 079.53; ICD-10-CA: F02.4, O98.7). (92-94)

All ICD-9-CM and ICD-10-CA codes in Table 3.2 were included when searching the Hospital Abstracts database. When searching the Medical Claims data, only ICD-9-CM codes 042-044 and V08 were included because, for most years of data, the diagnosis codes were recorded at the three-digit level and the truncation of 795.71 and 079.53 to 795 and 079, respectively, rendered the codes non-specific to HIV.

Table 3.2 HIV and related International Classification of Diseases (ICD) diagnosis codes

ICD Version	Code	Description
ICD-9-CM	042[a,b]	Human immunodeficiency virus (HIV) disease
ICD-9-CM	043[a,b]	HIV infection causing other specified conditions
ICD-9-CM	044[a,b]	Other HIV Infection
ICD-9-CM	V08[b]	Asymptomatic HIV infection status
ICD-9-CM	795.71[b]	Nonspecific serologic evidence of HIV (incl. non-conclusive HIV test finding in infants). Note: replaces ICD-9-CM code 795.8 as of October 1, 1994 (95, 96).
ICD-9-CM	079.53	HIV, Type 2
ICD-10-CA	B20[a,b]	HIV resulting in infectious and parasitic diseases
ICD-10-CA	B21[a,b]	HIV resulting in malignant neoplasms
ICD-10-CA	B22[a,b]	HIV resulting in other specified diseases
ICD-10-CA	B23[a,b]	HIV resulting in other conditions
ICD-10-CA	B24[a,b]	Unspecified HIV (includes AIDS and AIDS-related complex)

ICD Version	Code	Description
ICD-10-CA	R75[b]	Laboratory evidence of HIV (incl. non-conclusive HIV-test finding in infants)
ICD-10-CA	Z21[b]	Asymptomatic HIV infection status
ICD-10-CA	F02.4	Dementia in HIV disease
ICD-10-CA	O98.7	HIV disease complicating pregnancy, childbirth and the puerperium

ICD-9-CM: International Classification of Diseases, 9th Revision, with Clinical Modifications.
ICD-10-CA: International Statistical Classification of Diseases and Related Health Problems, 10th Revision, with Canadian Enhancements.
[a]ICD code used in algorithms developed and tested by Antoniou et al. (9)
[b]ICD codes used in algorithms tested by Nosyk et al. (10)

3.5.4 Drug Idenification Numbers for Algorithm Validation

This study used the DINs listed in Table 3.3 to identify HIV ART prescriptions in DPIN. The DINs used by Antoniou et al. (9) (denoted by boldface font in Table 3.3) formed the basis of this list. A pharmacist familiar with HIV ART updated the list to be current to 2018, adding 47 DINs. One DIN included by Antoniou et al. (9) was removed because it could not be found in the Health Canada Drug Product Database. A version of this DIN list was recently used to identify HIV ART in Manitoba. (97)

Table 3.3 HIV Antiretroviral Drug Identification Numbers (DINs)

Product Name	DINs
3TC	**02247825, 02192691, 02192683**
Abacavir	02480956, **02458381**, 02454513, **02450682, 02416662, 02416255, 02405776, 02399539, 02396769**
Agenerase	**02243543, 02243542, 02243541**
Apo-Zidovudine	**01946323**
Aptivus	**02273322**
Atripla	**02300699**
Biktarvy	**02478579**
Celsentri	**02299852, 02299844**
Combivir	**02239213**
Complera	**02374129**
Crixivan	**02229196, 02229161**
Darunavir	02487268, 02487141, 02486148, 02486121, 02486113, 02459426, 02437465, 02437457, 02437449, 02437430, 02437422
Delstrigo	02482592

Product Name	DINs
Descovy	02454424, 02454416
Dovato	02491753
Edurant	**02370603**
Efavirenz	**02389762, 02381524, 02418428**
Emtriva	02272091
Fortovase	**02239083**
Fuzeon	**02247725**
Genvoya	02449498
Heptovir	**02239194, 02239193**
Hivid	01990918, **01990896**
Imunovir	02240713
Incivek	02371553
Intelence	02396750, **02375931, 02306778**
Invirase	**02279320, 02216965**
Isentress	02392437, 02392429, 02301881
Isentress HD	02465337
Juluca	02375774
Kaletra	**02312301, 02285533, 02243644, 02243643**
Kivexa	**02269341**
Lamivudine	**02393239, 02369060, 02369052**
Lamivudine-Zidovudine	**02375540, 02387247, 02414414**
Nevirapine	**02387727, 02352893, 02318601, 02427931**
Norvir	**02357593**
Norvir - cancelled	**02241480, 02229145, 02229137**
Novo-AZT	01953877
Odefsey	02461463
Pifeltro	02481545
Prezcobix	**02426501**
Prezista	02416530, **02393050, 02369753, 02338432, 02324024, 02324016, 02284057**
Rescriptor	**02238348**
Retrovir (AZT)	**02238699, 01902660, 01902652,** 01902644
Reyataz	02294176, **02248611, 02248610**
Sebivo	02288389
Stribild	**02397137**
Sustiva	**02246045, 02239888, 02239887, 02239886**
Symtuza	02473720
Telzir	02261553, **02261545**
Tivicay	02461226, 02461218, **02414945**
Triumeq	**02430932**
Trizivir	**02244757**
Truvada	**02274906**
Tybost	02411423
Victrelis	02370816
Videx	01940635, 01940554, **01940546**, 01940538, 01940511

Product Name	DINs
Videx EC	**02244599, 02244598, 02244597, 02244596**
Viracept	**02248761, 02238618, 02238617**
Viramune - cancelled	**02238748**
Viramune XR - cancelled	**02367289**
Viread	**02247128**
Viteka	02411180, 02411172
Zerit	**02216116, 02216108, 02216094, 02216086**
Ziagen	**02240358, 02240357**

Notes: 1) The DINs used by Antoniou et al. (9) are in boldface font. 2) One DIN (09852875) used by Antoniou et al. (9) was not included in this study because it could not be found in the Health Canada Drug Product Database.

3.5.5 Study Variables

3.5.5.1 HIV Case Identification

Validation Cohort. A dichotomous HIV indicator (HIV case/HIV non-case) was created to denote whether an individual in the validation cohort had a positive (or non-negative, non-indeterminate) result on an HIV laboratory test (which includes tests that identified HIV antibody, HIV DNA, or the HIV virus itself, and tests for HIV drug resistance) in the study observation period. A positive result on an HIV laboratory test was identified using the same method as in two MCHP deliverables. (12, 30) For individuals with more than one positive HIV laboratory test in the study observation period, the earliest test was used to determine the date of HIV case identification in the study period.

Algorithm Case Definitions. One dichotomous HIV indicator for each of the 30 algorithms validated (e.g., Algorithm 1: HIV case/HIV non-case, Algorithm 2: HIV case/HIV non-case) was created to denote whether an individual in the validation cohort met the HIV algorithm criteria in the study observation period. To meet each algorithm criteria, an individual must have had the prescribed number and type (physician visit, hospital discharges, prescriptions) of records in the Medical Claims, Hospital Abstracts, or DPIN databases during the study period. For example, Algorithm 9 requires three physician visits or one hospitalization for HIV within two years to be considered an HIV case. This means that in the study observation period, an individual must have had at least three physician claims within a 780-day (i.e., two-year) period or at least one hospital discharge to be ascertained as an HIV case. All Medical Claims, Hospital Abstracts, and DPIN records over the study observation period were considered when applying the algorithms.

3.5.5.2 Covariates

Several covariates were used to describe the validation cohort. Age, sex, area of residence, area-level household income, the amount of time covered by health insurance, and the earliest year of HIV case identification were measured from the MHIR, Statistics Canada's Canadian Census, and the CPL databases. The index date at which age, sex, area of residence, and area-level household income were measured for HIV cases and HIV non-cases was defined as the start of insurance coverage (if insurance coverage started after January 1, 2007) or January 1, 2007. Year of HIV case identification was measured as the year of the first confirmatory HIV laboratory test and therefore only measured for HIV cases. These variables were used to describe the population under study. Table 3.4 presents a rationale for selecting these covariates, their sources, and their operational definitions.

3.5.6 Sensitivity Analyses

Sensitivity Analyses 1 and 2: To produce algorithm performance estimates that were comparable to the performance estimates produced by the other Canadian HIV algorithm studies, the 30 algorithms were applied using 1) only the ICD codes used by Antoniou et al. (9) and 2) only the ICD codes used by Nosyk et al. (10)

Sensitivity Analysis 3: To determine the performance of the algorithms among a population consisting of adults only, the validation cohort was restricted to those individuals who were at least 18 years old as of January 1, 2007.

Sensitivity Analysis 4: To validate the algorithms in a cohort that contained only true HIV positive and true HIV negative individuals, the cohort was limited to females who received a prenatal HIV screening test in the study observation period, and who were at least 15 years old at the time of the test. All pregnant females should receive an HIV screening test as part of their routine prenatal care and thus should have a positive or negative prenatal HIV test result in the CPL data. For females with more than one prenatal HIV screen in the study observation period, their earliest positive screen (HIV cases) or earliest negative screen (HIV non-cases) was included.

Table 3.4 Definitions and sources of variables to describe the validation cohort

Variable	Rationale	Type	Source	Operational Definition
Age, years	The accuracy of the HIV algorithms may vary by age. (78)	Continuous	MHIR	0 years – 116.2 years
Sex	The accuracy of the HIV algorithms may vary by sex. (78)	Categorical	MHIR	Male/Female
Area of residence	The accuracy of the HIV algorithms may vary across geographic areas. (78)	Categorical	MHIR	IERHA/NHR/ SH-SS/PMH/WHRA/ Public Trustee/Not Found

Variable	Rationale	Type	Source	Operational Definition
Area-level household income	Income is a social determinant of health known to influence both an individuals' healthcare experience and their health outcomes. (98) Area-level household income is also related to the geographic region in which an individual resides. (98) Thus, the accuracy of the HIV algorithms may vary with area-level household income. (78)	Categorical	MHIR (Postal Code), Census	Income Quintile: Q1/NF (Lowest/Not Found) Q2 Q3 Q4 Q5 (Highest)
Health insurance coverage, years	Individuals with fewer years of provincial health insurance coverage may have had fewer interactions with the healthcare system than individuals with more years of provincial health insurance coverage. Thus, algorithms validated in populations with many years of health insurance coverage may not perform as well in populations with less health insurance coverage.	Continuous	MHIR	3 years – 12 years.
Year of HIV case identification[b]	The number of cases identified in a given year will represent the number of newly diagnosed cases, plus the number of previously diagnosed cases undergoing follow-up testing for the first time during the time frame.	Categorical	CPL, CPL LIMS	2007-2018

MHIR, Manitoba Health Insurance Registry; MC/MS, Medical Claims/Medical Services; Census, Statistics Canada's Census Data, CPL, Cadham Provincial Laboratory database, CPL LIMS, Cadham Provincial Laboratory Laboratory Information Management System.
[b] For those with a positive result on a confirmatory HIV laboratory test only.

3.5.7 Statistical Analyses

All statistical analyses were completed using SAS® 9.4.

3.5.7.1 Description of Validation Cohort

Continuous variables were described using mean and standard deviation. Categorical variables were described using frequencies and percentages.

3.5.7.2 Algorithm Performance Estimates

Following the Checklist of Reporting Criteria for Studies Validating Health Administrative Data Algorithms, (78) a confusion matrix (Table 3.5) was created and used to calculate seven measures of accuracy for each of the 30 algorithms: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), Youden's J statistic, kappa statistic, and AUC. The definitions for each of these performance measures are given in Table 3.6. Sensitivity, specificity, PPV, and NPV are reported as percentages. The 95% confidence interval for each measure was calculated using a binomial probability distribution.

Table 3.5 Confusion matrix for HIV cases and non-cases

		Validation Cohort		
		HIV Case	**HIV Non-Case**	**Total**
Algorithm	**HIV Case**	TP	FP	TP + FP
	HIV Non-Case	FN	TN	FN + TN
	Total	TP + FN	FP + TN	TP + FP + FN + TN

Note: TP = true positives; FP = false positives; FN = false negatives; TN = true negatives.

Table 3.6 Definitions of algorithm performance measures

Measure	Definition (99)	Equation*
Sensitivity	The proportion of HIV cases in the validation cohort who were identified as HIV cases by the administrative data algorithms.	$Sensitivity = \frac{TP}{TP + FN}$
Specificity	The proportion of HIV non-cases in the validation cohort who were identified as HIV non-cases by the administrative data algorithms.	$Specificity = \frac{TN}{FP + TN}$
Positive Predictive Value (PPV)	The proportion of HIV cases identified by the administrative data algorithms that are also HIV cases in the validation cohort. In other words, the probability an individual is an HIV case when identified as an HIV case by an administrative data algorithm.	$PPV = \frac{TP}{TP + FP}$
Negative Predictive Value (NPV)	The proportion of HIV non-cases identified by the administrative data algorithms that are also HIV non-cases in the validation cohort. In other words, the probability an individual is an HIV non-case when identified as an HIV non-case by an administrative data algorithm.	$NPV = \frac{TN}{FN + TN}$
Youden's J Statistic	A summary measure of sensitivity and specificity that can be used to identify the optimal combination of	$J = Sensitivity + Specificity - 1$

Measure	Definition (99)	Equation*
	these two estimates. The optimal value of Youden's Index is 1.0.	
Kappa Statistic	A measure of accuracy that assesses the agreement between two sources while taking into account the accuracy that would be generated simply by chance. The closer to 1.0 the kappa (κ) statistic is, the better the algorithm is able to distinguish between HIV cases and HIV non-cases.	The interpretation of κ used in this study is (91, 100): • Poor agreement: $\kappa < 0.20$ • Fair agreement: $\kappa = 0.20$ to 0.39 • Moderate agreement: $\kappa = 0.40$ to 0.59 • Good agreement: $\kappa = 0.60$ to 0.79 • Very good agreement: $\kappa = 0.80$ to 1.00
Area Under the Receiver Operating Characteristics Curve (AUC)	A measure of accuracy that tells how well the model performs. The closer the *AUC* is to 1.0, the better the algorithm is able to discriminate between HIV cases and non-cases.	The interpretation of *AUC* used in this report is (101): • No discrimination (may as well flip a coin): $AUC = 0.5$ • Poor discrimination (not much better than coin toss): $0.5 < AUC < 0.7$ • Acceptable discrimination: $0.7 \leq AUC < 0.8$ • Excellent discrimination: $0.8 \leq AUC < 0.9$ • Outstanding discrimination: $AUC \geq 0.9$.

Note: TP = true positives; FP = false positives; FN = false negatives; TN = true negatives (refer to Table 3.5).
* Multiply the equations for sensitivity, specificity, PPV, and NPV by 100 to get the percentage.
The 95% confidence intervals for each measure were calculated using a binomial probability distribution.
Definition reference (99): Kuhn M, Johnson K. Applied Predictive Modelling. USA: Springer; 2016.
Equation references:

- (100) Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall; 1991.
- (101) Hosmer D, Lemeshow S, Sturdivant R. Applied Logistic Regression. 3 ed: John Wiley & Sons, Inc.; 2013.
- (91) Lix L, Yogendran M, Burchill C, Metge C, McKeen N, Moore D, et al. Defining and Validating Chronic Diseases: An Administrative Data Approach. Winnipeg, Manitoba; 2006

3.5.7.3 Provincial Prevalence Estimates

Unlike PPV and NPV, sensitivity and specificity are commonly thought of as functions of the test itself and unrelated to the prevalence of the disease being measured. (102-104) However, several studies have shown that changes in prevalence are associated with changes in sensitivity and specificity. (102-104) Factors that affect prevalence, such as how a disease is defined and measured, or an imperfect reference standard, also affect sensitivity and specificity. (102) In practice, it is difficult to identify and measure all possible mechanisms that may affect prevalence, sensitivity, and specificity. Instead, it has been recommended that researchers ensure the prevalence of the disease in their target population (in this case, the Manitoba population) is similar to the prevalence of disease in the population in which the algorithm was validated (in this case, the validation cohort), to ensure the algorithm will perform with reasonably similar accuracy. (102-104)

To this end, and to assist with comparisons across algorithms, crude provincial prevalence estimates of HIV were calculated for each of the 30 administrative data algorithms. Prevalence estimates were derived using a denominator of the Manitoba population continuously registered with the MHIR between January 1, 2016 and December 31, 2018. The numerators consisted of anyone from the denominator who met the algorithm case definition during the study observation period.

3.6 Ethical Considerations

Data held by the MCHP are protected under The Personal Health Information Act (105) and the Freedom of Information and Protection of Privacy Act (106). Prior to initiating this project, ethics approval was obtained from the University of Manitoba Health Research Ethics Board (HREB: HS23945). Special permissions were obtained from Manitoba Health as the trustee of the data required for this study prior to obtaining access to the data. Data access was granted by the Health Information Privacy Committee (HIPC: 2020/2021-11) on behalf of Manitoba Health. As per MCHP policies, all project data were stored and analyzed on a secure server that was accessed either on-site or through a secure, monitored remote access system. Groupings resulting in one to five observations were suppressed in accordance with MCHP publication policies to avoid the potential identification of individual patients.

Chapter 4 - Results

4.1 Introduction

This chapter provides the results of the study. Section 4.2 presents the validation cohort selection process and demographics. Section 4.3 details the algorithms' application to the various data sources and provides the demographics of the HIV cases identified by each algorithm. Section 4.4 presents how well the algorithms identified HIV cases relative to the reference standard and the HIV prevalence calculated by each algorithm. Section 4.5 summarizes the four sensitivity analyses.

4.2 Description of Validation Cohort

The validation cohort included 1,454,010 Manitobans with a valid PHIN and at least three years of continuous health insurance coverage in the MHIR during the study observation period, which was from January 1, 2007 to December 31, 2018 (Figure 4.1). Of these individuals, 1,589 (0.1%) were HIV cases, as defined by having at least one positive result on a confirmatory HIV laboratory test (which includes tests that identified HIV antibody, HIV DNA, or the HIV virus itself, and tests for HIV drug resistance) in the CPL databases during the study observation period and 1,452,421 (99.9%) were HIV non-cases.

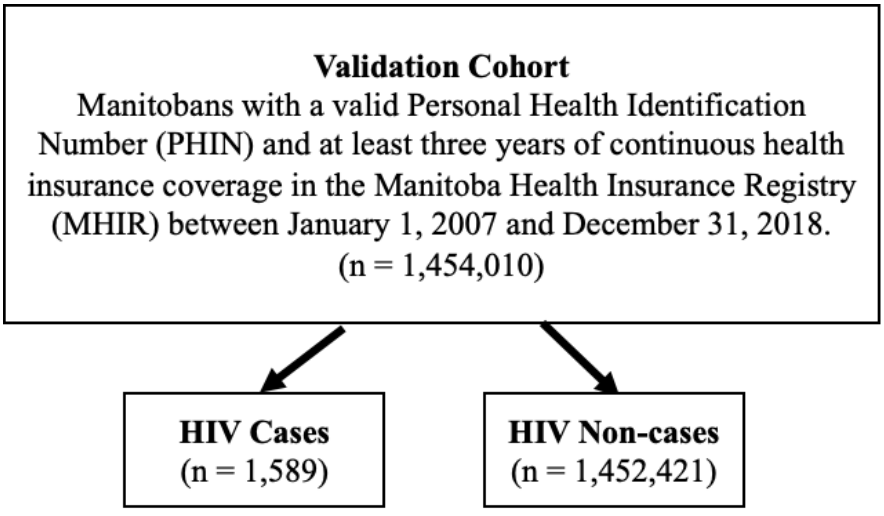

Figure 4.1. Construction of the validation cohort

Characteristics of the validation cohort, stratified by HIV cases and HIV non-cases are provided in

Table 4.1. The number of people who could not be assigned to an income quintile (e.g. Public Trustees) was small, so they were included with Income Quintile 1 (Q1).

The average HIV case was 37.5 years old as of the index date and 40.2 years old at the time of their earliest positive result on a confirmatory HIV laboratory test during the study observation period. Two-thirds of HIV cases were male. Most HIV cases lived in the Winnipeg Regional Health Authority (WRHA). Approximately one-half of HIV cases were assigned to the lowest income quintile, Q1, with a decreasing number of HIV cases as the quintiles increased. The number of HIV cases identified by year (where year was assigned based on the date of earliest positive result on a confirmatory HIV laboratory test in the study observation period) was higher in the first half of the study observation period because follow-up tests for prevalent HIV cases infected before 2007 were being identified. By the second half of the study observation period, the identification of HIV cases approximates the incidence of HIV in Manitoba in those years.

The average HIV non-case was 32.7 years old as of the index date; almost 10% of the HIV non-cases were less than one year of age as of their index date. More than one-half of HIV non-cases lived in WRHA. HIV non-cases were evenly distributed over all five income quintiles and by sex.

4.3 Algorithm Application

Thirty algorithms (Table 3.1) that used a combination of physician billing claims, inpatient hospitalizations, and ART prescriptions to identify HIV cases were tested. Six possible ICD-9-CM codes and nine possible ICD-10-CA codes for diagnosing HIV were identified in previous literature and by reviewing ICD-9-CM and ICD-10-CA documentation (Table 3.2). When searching the Medical Claims database, all included ICD-9-CM diagnosis codes (042-044 and V08) contributed to algorithm development. When searching the Hospital Abstracts database, two of the six included ICD-9-CM diagnosis codes (042.1 and 042.9) and six of the nine included ICD-10-CA diagnosis codes (B24, O98.701, O98.703, O98.704, R75, and Z21) contributed to algorithm development. When searching the DPIN database, 93 (Table 4.2) of the possible 142 DINs (Table 3.3) identified by pharmacists or previous literature, contributed to algorithm development.

Error! Reference source not found. to Figure 4.5 present Venn diagrams of the number of HIV cases identified by data source for all algorithms that required more than one data source.

The number of HIV cases identified by algorithms that required only physician visits can be found in Appendix A in Tables A.1-A.6 and Tables A.16-A.21.

Tables A.1-A.6 and Tables A.16-A.21 demonstrate that as the number of physician visits required by an algorithm increased, the number of HIV cases identified by the algorithm decreased, regardless of whether the algorithms were applied to a two- or a three-year period.

Table 4.1. Description of the validation cohort

Characteristic	HIV Case (N = 1,589)	HIV Non-Case (N = 1,452,421)
Age at index date, years	36.5 (12.6)	32.7 (23.3)
[minimum, maximum]	[0, 83.8]	[0, 116.2]
Sex		
Male	1,053 (66.3)	722,957 (49.8)
Female	536 (33.7)	729,464 (50.2)
Area of residence (RHA)		
IERHA	81 (5.1)	131,480 (9.1)
NHR	66 (4.2)	88,321 (6.1)
SH-SS	67 (4.2)	20,2427 (13.9)
PMH	67 (4.2)	18,9501 (13.0)
WRHA	1,296 (81.6)	835,718 (57.5)
Public Trustee/Not Found	12 (0.8)	4,974 (0.3)
Income quintile		
Q1-Lowest/Not Found	820 (51.6)	316,887 (21.8)
Q2	305 (19.2)	292,131 (20.1)
Q3	215 (13.5)	283,394 (19.5)
Q4	135 (8.5)	283,806 (19.5)
Q5-Highest	114 (7.2)	276,203 (19.0)
Health insurance coverage, years	10.3 ± 2.6	10.3 ± 2.8
Age at HIV case identification, years*	40.2 ± 12.2	-
Year of HIV case identification*		
2007	201 (12.6)	-
2008	159 (10.0)	-
2009	312 (19.6)	-
2010	238 (15.0)	-
2011	116 (7.3)	-
2012	92 (5.8)	-
2013	103 (6.5)	-
2014	77 (4.8)	-
2015	86 (5.4)	-
2016	63 (4.0)	-
2017	64 (4.0)	-
2018	78 (4.9)	-

Notes:
Data are mean ± standard deviation or n (%)
*calculated at the date of earliest positive result on a confirmatory HIV laboratory test in the study observation period, for HIV cases only.
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority

Table 4.2. HIV Antiretroviral Drug Identification Numbers (DINs) identified in the Drug Program Information Network (DPIN) database

Product Name	DINs
3TC	**02247825, 02192691, 02192683**
Abacavir	**02458381**, 02454513, **02450682, 02416662, 02416255, 02405776, 02399539, 02396769**
Apo-Zidovudine	**01946323**
Aptivus	**02273322**
Atripla	**02300699**
Biktarvy	**02478579**
Celsentri	**02299852, 02299844**
Combivir	**02239213**
Complera	**02374129**
Crixivan	**02229196**
Descovy	02454424
Edurant	**02370603**
Efavirenz	**02389762, 02381524, 02418428**
Fortovase	**02239083**
Fuzeon	**02247725**
Genvoya	02449498
Heptovir	**02239193**
Imunovir	02240713
Incivek	02371553
Intelence	**02375931, 02306778**
Invirase	**02279320, 02216965**
Isentress	02301881
Kaletra	**02312301, 02285533, 02243644, 02243643**
Kivexa	**02269341**
Lamivudine	**02393239, 02369060, 02369052**
Lamivudine-Zidovudine	**02375540, 02387247, 02414414**
Nevirapine	**02387727, 02352893, 02318601, 02427931**
Norvir	**02357593**
Norvir - cancelled	**02241480, 02229145, 02229137**
Odefsey	02461463
Prezcobix	**02426501**
Prezista	**02393050, 02338432, 02324024, 02324016, 02284057**
Rescriptor	**02238348**
Retrovir (AZT)	**01902660, 01902652,** 01902644
Reyataz	02294176, **02248611, 02248610**
Stribild	**02397137**
Sustiva	**02246045, 02239888, ~~02239887,~~ 02239886**
Telzir	**02261545**
Tivicay	**02414945**
Triumeq	**02430932**
Trizivir	**02244757**

Product Name	DINs
Truvada	**02274906**
Victrelis	02370816
Videx EC	**02244599, 02244598, 02244597**
Viracept	**02248761, 02238617**
Viramune - cancelled	**02238748**
Viramune XR - cancelled	**02367289**
Viread	**02247128**
Zerit	**02216116, 02216108, 02216094, 02216086**
Ziagen	**02240358, 02240357**

Notes: The DINs used by Antoniou et al. (9) are in boldface font.

More HIV cases were identified when searching within a three-year period compared to a two-year period. The difference in the number of HIV cases identified by varying the time frame ranged from zero cases when searching for one or more physician visits (Algorithm 1 and 16: 2,194 cases each) to 35 cases when searching for six or more physician visits (Algorithm 6: 1,343 cases vs. Algorithm 21: 1,378 cases).

Figure 4.2, Figure 4.3, Figure 4.4, and Figure 4.5 show that for algorithms where physician visits were combined with hospitalizations and/or prescriptions, the proportion of HIV cases identified solely by physician visits decreased as the number of physician visits required by an algorithm increased. In turn, the proportion of HIV cases identified by hospitalizations, prescriptions, or some combination thereof, increased. For example, in Algorithm 7 (Figure 4.2), 56.2% of HIV cases were identified solely by having one or more physician visit in 2 years, while in Algorithm 9 (Figure 4.2), only 40.7% of HIV cases were identified solely by having one or more physician visit in 2 years.

The number of HIV cases identified via one or more inpatient hospitalization was not influenced by the number of years searched by the algorithm, or any other data sources included in the algorithm. If an individual had at least one HIV hospitalization within two years then they also had at least one HIV hospitalization within three years. Any algorithm that included at least one inpatient hospitalization identified a total of 1,010 HIV cases via inpatient hospitalization.

The number of HIV cases identified via two or more HIV antiretroviral prescriptions was minimally influenced by the number of years searched by the algorithm. Searching an additional year of prescription data identified only nine additional HIV cases. Any two-year algorithm that included two or more HIV antiretroviral prescriptions identified a total of 2,661 HIV cases via

prescriptions while any three-year algorithm that included two or more HIV antiretroviral prescriptions identified a total of 2,670 HIV cases via prescriptions.

Tables A.1 to A.30 in Appendix A describe the characteristics of the HIV cases and HIV non-cases identified by the 30 algorithms. Across all algorithms, the false positive HIV cases tend to reflect the descriptive characteristics of HIV non-cases from the validation cohort whereas false negative cases tend to reflect the descriptive characteristics of the HIV cases from the validation cohort. Thus, for algorithms where false positives contribute a high proportion of the total HIV cases, the characteristics of the HIV cases shift toward those of HIV non-cases.

Of the HIV cases identified by the algorithms, the percent of which were false positives relative to the reference standard ranged from 4.0% (Algorithm 6) to 55.9% (Algorithm 28). The 12 algorithms which required two or more HIV antiretroviral prescriptions (Algorithms 10-15 and 25-30) produced the greatest number of false positive HIV cases relative to the reference standard, with the percent of false positives ranging from 46.8% (Algorithm 12: 1,323 false positives) to 55.9% (Algorithm 28: 1,943 false positives). Across all algorithms, as the number of physician visits required by an algorithm increased, the number of false positives identified decreased.

For all algorithms, the average age of false positive HIV cases was almost 10 years younger than the average age of true positive HIV cases. The sex distribution also varied between true and false positives with false positives having a more even balance of cases between males and females, while true positives were generally two-thirds male. All other descriptive characteristics followed relatively similar distributions, regardless of true or false positive. Since the HIV cases identified by the 12 algorithms that required two or more HIV antiretroviral prescriptions (Algorithms 10-15 and 25-30) included so many false positives, the age and sex characteristics of these HIV cases do not align with the age and sex characteristics of the reference standard (

Table 4.1).

Overall, the number of HIV cases identified by these algorithms ranged from 1,343 HIV cases (246 fewer cases than the reference standard) for Algorithm 6, which had the lowest number of false positives and required six or more physician visits in two years, to 3,477 HIV cases (1,888 more cases than the reference standard) for Algorithm 28, which had the highest number of false positives and required one or more physician visits, one or more hospitalizations, or one or more antiretroviral prescription in three years.

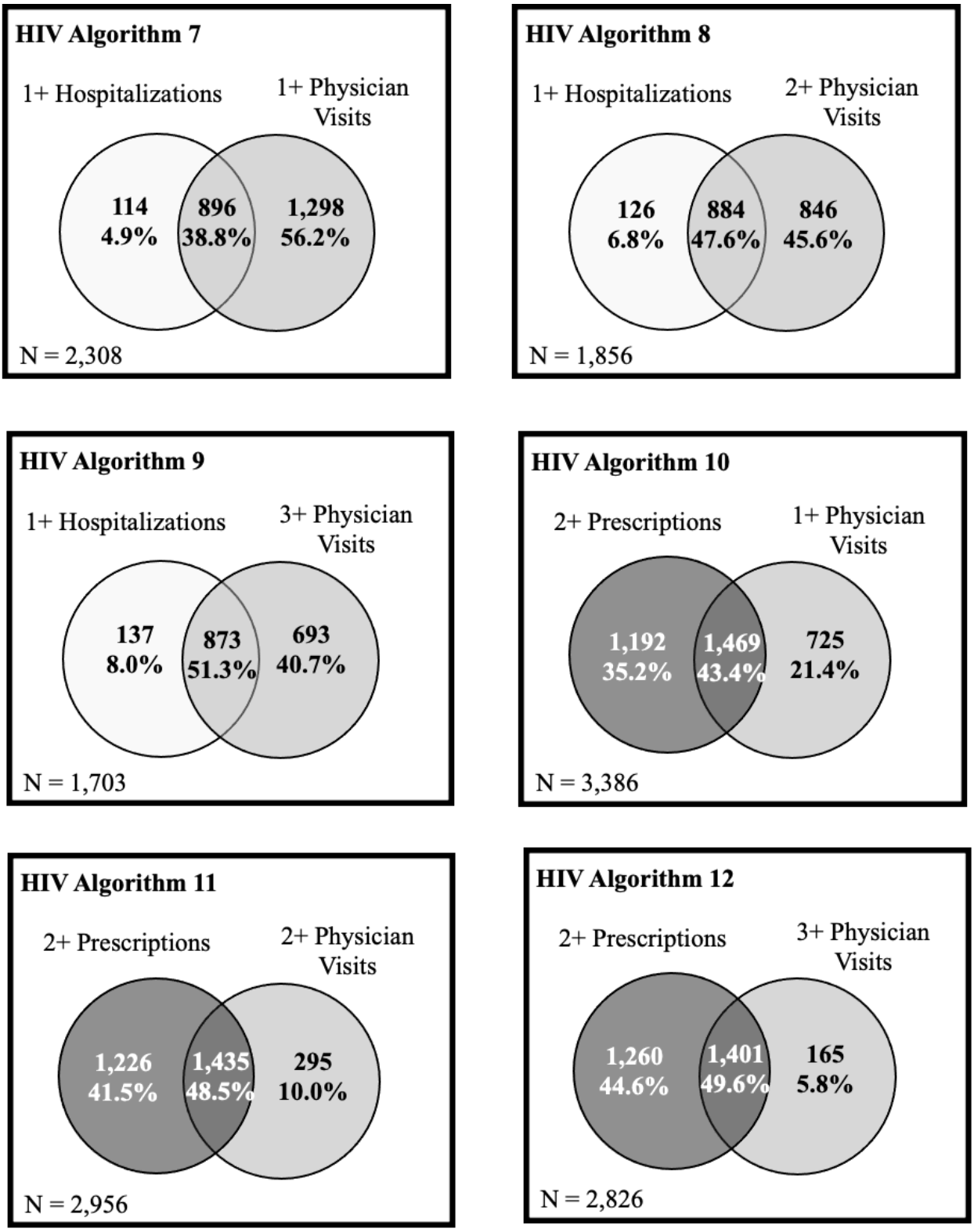

Figure 4.2 HIV cases identified by Algorithms 7-12

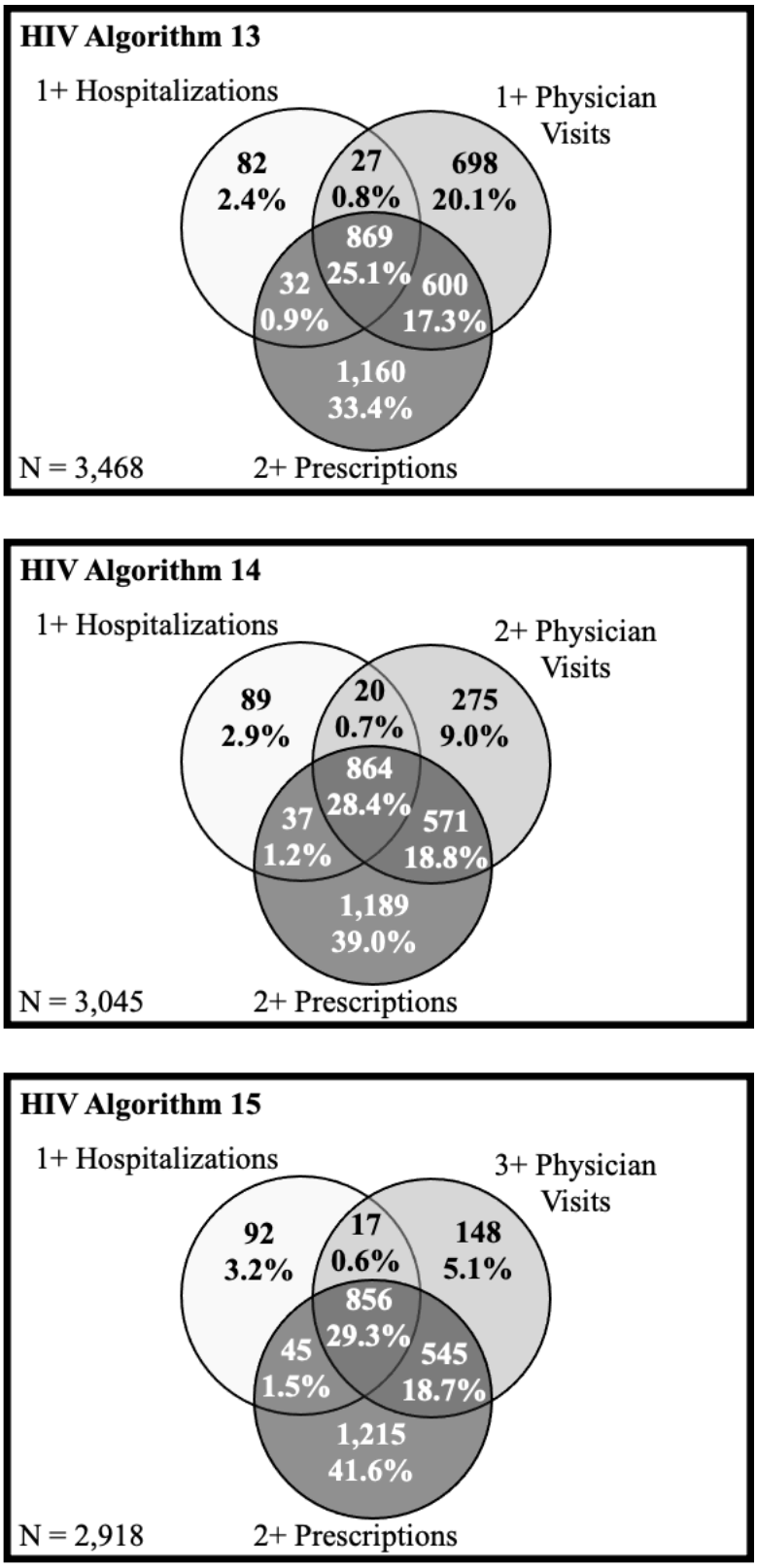

Figure 4.3 HIV cases identified by Algorithms 13-15

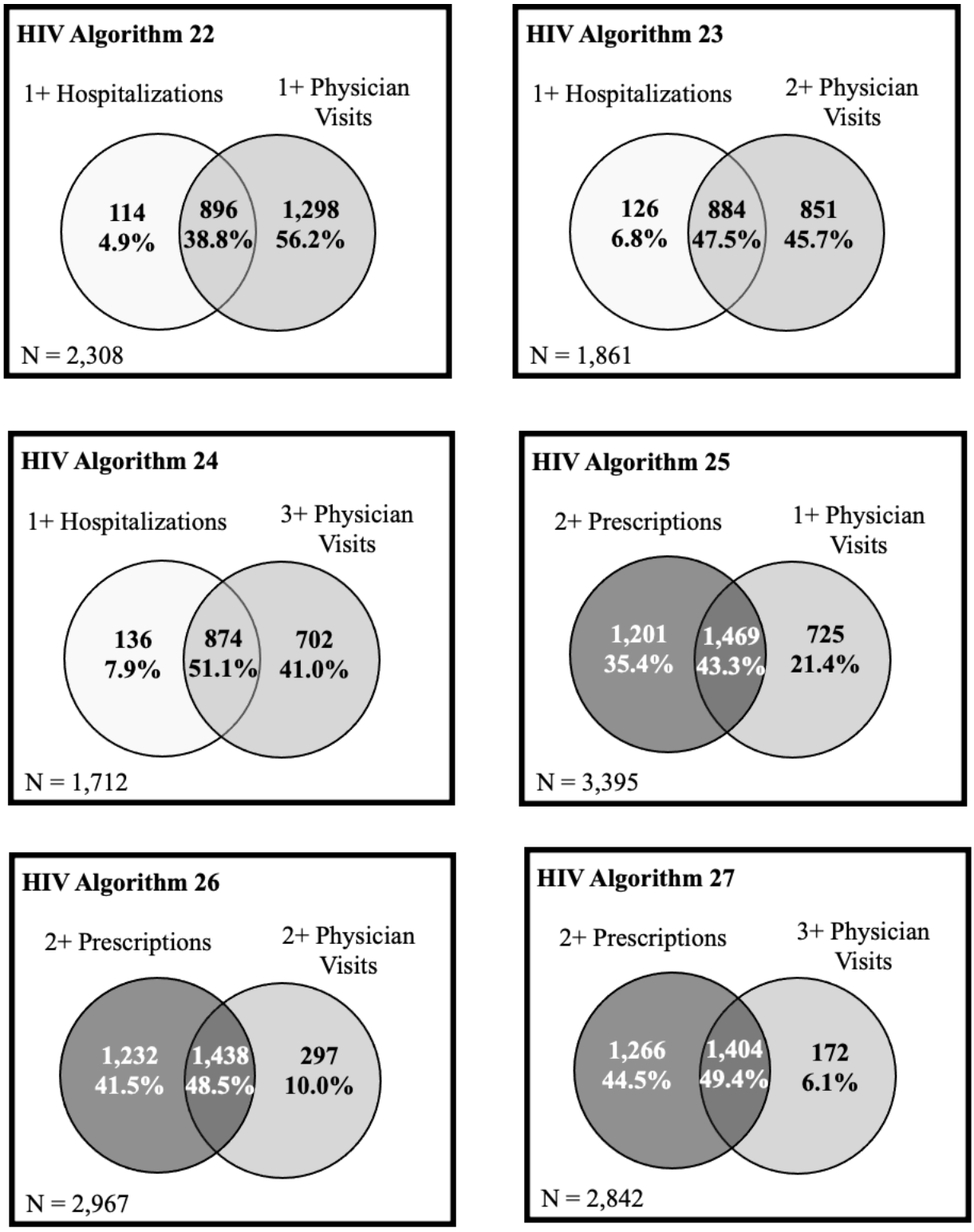

Figure 4.4 HIV cases identified by Algorithms 22-27

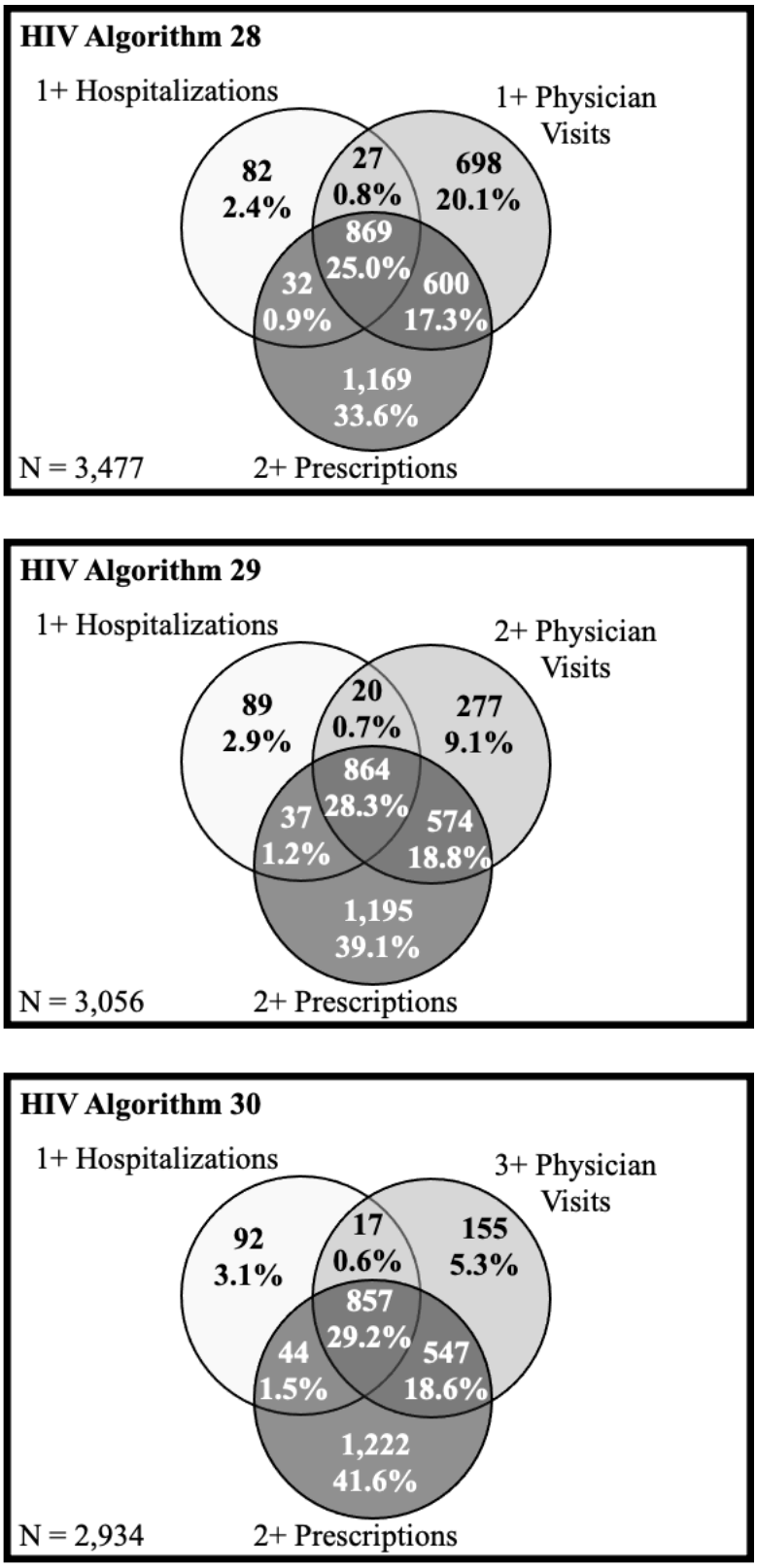

Figure 4.5 HIV cases identified by Algorithms 28-30

The number of HIV cases identified by Algorithm 18, which required three or more physician visits in three years, was closest to the number of HIV cases identified by the reference standard, with only 13 fewer cases identified. Given there were few false positives produced by this algorithm (142 false positives, accounting for 9.0% of the HIV cases), the descriptive characteristics of the HIV cases identified closely aligned with the descriptive characteristics of the reference standard. The average age as of the index date was 35.8 years for HIV cases identified by Algorithm 18 compared to 36.5 years for HIV cases identified by the reference standard. 64.7% of HIV cases identified by the algorithm were males and 81.5% were from Winnipeg, compared to 66.3% and 81.6% of HIV cases identified by the reference standard, respectively. Approximately one-half (51.3%) of HIV cases identified by the algorithm were from Q1, similar to the 51.6% of HIV cases identified by the reference standard.

4.4 Algorithm Performance Estimates

The sensitivity, specificity, and PPV for all algorithms are given in Table 4.3. Algorithm sensitivity ranged from 81.1% (95% CI 79.1%-83.0%) for Algorithm 6 to 96.5% (95% CI 95.5%-97.4%) for Algorithms 13 and 28. The PPV of the algorithms ranged from 44.1% (95% CI 42.5%-45.8%) for Algorithm 28 to 96.0% (95% CI 94.8%-97.0%) for Algorithm 6. All algorithms which required two or more antiretroviral prescriptions (Algorithms 10-15 and 25-30) had PPV $\leq$ 53.2%. Specificity and NPV were very high for all algorithms. Specificity ranged from 99.9% (95% CI 99.9%-99.9%) to 100.0% (95% CI 100.0%-100.0%) for all algorithms. Since many algorithms had the same specificity, specificity is represented by symbols in Table 4.3. NPV was 100.0% (95% CI 100.0%-100.0%) for all algorithms and so was not included in Table 4.3.

Algorithms with more data available, whether due to increased time frame or more data sources, identified more positive HIV cases, largely due to an increasing number of false positives. Thus, sensitivity was higher for all algorithms in the three-year time frame compared to the two-year time frame and combining physician visit data with hospitalization data, prescription data, or both data sources, increased the algorithm sensitivity compared to that of the algorithm with the same number of physician visits alone. As algorithm restrictions increased, such as requiring multiple physician visits, the number of positive HIV cases identified decreased. Algorithms 1-6 and 16-21 included physician visit data only; regardless of the time frame used, sensitivity

decreased as the number of required physician visits increased. Combining the data sources lessened the impact that increasing the number of physician visits had on decreasing sensitivity.

The Youden's J Statistic, kappa, and AUC for all algorithms are given in Table 4.4. Youden's J Statistic ranged from 0.81 (95% CI 0.79 - 0.83) to 0.96 (95% CI 0.96 - 0.97) across all algorithms. The kappa statistic ranged from 0.61 (95% CI 0.59-0.62) to 0.91 (95% CI 0.90-0.92), indicating good or very good agreement between all algorithms and the reference standard. The AUC ranged from 0.91 (95% CI 0.90-0.92) to 0.98 (95% CI 0.98-0.99) indicating outstanding discrimination between HIV cases and HIV non-cases for all algorithms. The time frame over which the algorithms were applied had little effect on the values of Youden's J Statistic, kappa, or AUC.

Youden's J Statistic, kappa, and AUC varied little regardless of whether the algorithms were applied over two years or three years. Increasing the number of physician visits required by an algorithm decreased Youden's J Statistic and AUC and increased kappa (until five physician visits). At five physician visits, kappa also began to decrease. Combining physician visit data with hospitalization data, prescription data, or both data sources, increased Youden's J Statistic and AUC, but decreased kappa, compared to that of the algorithm with the same number of physician visits alone.

Table 4.5 shows the algorithm-derived prevalence of HIV among Manitobans with continuous health insurance coverage between January 1, 2016 and December 31, 2018. The prevalence calculated using the algorithms ranged from 90.2 diagnosed HIV cases per 100,000 population (95% CI 85.1 - 95.6) for Algorithm 6 to 232.7 diagnosed HIV cases per 100,000 population (95% CI 224.4 - 241.4) for Algorithm 28. Prevalence calculated using algorithms that required antiretroviral prescription data was much greater than the prevalence calculated using algorithms that did not require prescription data. Excluding algorithms that required two or more antiretroviral prescriptions, the maximum prevalence was 151.9 diagnosed HIV cases per 100,000 population (95% CI 145.2 - 158.9) for Algorithms 7 and 22.

Table 4.3. Performance estimates (sensitivity, specificity, PPV) of algorithms identifying people living with HIV in administrative health data

Algorithm	True Positive	False Positive	False Negative	True Negative	Sensitivity	PPV
1. 1+ P, 2 years*	1,484	710	105	1,451,711	93.4% (92.1% - 94.6%)	67.6% (65.6% - 69.6%)
2. 2+ P, 2 years*	1,462	268	127	1,452,153	92.0% (90.6% - 93.3%)	84.5% (82.7% - 86.2%)
3. 3+ P, 2 years*	1,427	139	162	1,452,282	89.8% (88.2% - 91.3%)	91.1% (89.6% - 92.5%)
4. 4+ P, 2 years*	1,388	87	201	1,452,334	87.4% (85.6% - 89.0%)	94.1% (92.8% - 95.3%)
5. 5+ P, 2 years*	1,342	66	247	1,452,355	84.5% (82.6% - 86.2%)	95.3% (94.1% - 96.4%)
6. 6+ P, 2 years*	1,289	54	300	1,452,367	81.1% (79.1% - 83.0%)	96.0% (94.8% - 97.0%)
7. 1+ P, 1+ H, 2 years**	1,514	794	75	1,451,627	95.3% (94.1% - 96.3%)	65.6% (63.6% - 67.5%)
8. 2+ P, 1+ H, 2 years*	1,499	357	90	1,452,064	94.3% (93.1% - 95.4%)	80.8% (78.9% - 82.5%)
9. 3+ P, 1+ H, 2 years*	1,475	228	114	1,452,193	92.8% (91.4% - 94.1%)	86.6% (84.9% - 88.2%)
10. 1+ P, 2+ Rx, 2 years***	1,528	1,858	61	1,450,563	96.2% (95.1% - 97.1%)	45.1% (43.4% - 46.8%)
11. 2+ P, 2+ Rx, 2 years***	1,518	1,438	71	1,450,983	95.5% (94.4% - 96.5%)	51.4% (49.5% - 53.2%)
12. 3+ P, 2+ Rx, 2 years***	1,503	1,323	86	1,451,098	94.6% (93.4% - 95.7%)	53.2% (51.3% - 55.0%)
13. 1+ P, 1+ H, 2+ Rx, 2 years***	1,534	1,934	55	1,450,487	96.5% (95.5% - 97.4%)	44.2% (42.6% - 45.9%)
14. 2+ P, 1+ H, 2+ Rx, 2 years***	1,526	1519	63	1,450,902	96.0% (95.0% - 96.9%)	50.1% (48.3% - 51.9%)
15. 3+ P, 1+ H, 2+ Rx, 2 years***	1,514	1,404	75	1,451,017	95.3% (94.1% - 96.3%)	51.9% (50.0% - 53.7%)
16. 1+ P, 3 years*	1,484	710	105	1,451,711	93.4% (92.1% - 94.6%)	67.6% (65.6% - 69.6%)
17. 2+ P, 3 years*	1,465	270	124	1,452,151	92.2% (90.8% - 93.5%)	84.4% (82.7% - 86.1%)
18. 3+ P, 3 years*	1,434	142	155	1,452,279	90.3% (88.7% - 91.7%)	91.0% (89.5% - 92.4%)
19. 4+ P, 3 years*	1,402	90	187	1,452,331	88.2% (86.5% - 89.8%)	94.0% (92.6% - 95.1%)
20. 5+ P, 3 years*	1,363	68	226	1,452,353	85.8% (84.0% - 87.5%)	95.3% (94.0% - 96.3%)
21. 6+ P, 3 years*	1,321	57	268	1,452,364	83.1% (81.2% - 84.9%)	95.9% (94.7% - 96.9%)
22. 1+ P, 1+ H, 3 years**	1,514	794	75	1,451,627	95.3% (94.1% - 96.3%)	65.6% (63.6% - 67.5%)
23. 2+ P, 1+ H, 3 years*	1,502	359	87	1,452,062	94.5% (93.3% - 95.6%)	80.7% (78.8% - 82.5%)
24. 3+ P, 1+ H, 3 years*	1,481	231	108	1,452,190	93.2% (91.9% - 94.4%)	86.5% (84.8% - 88.1%)
25. 1+ P, 2+ Rx, 3 years***	1,528	1,867	61	1,450,554	96.2% (95.1% - 97.1%)	45.0% (43.3% - 46.7%)
26. 2+ P, 2+ Rx, 3 years***	1,518	1,449	71	1,450,972	95.5% (94.4% - 96.5%)	51.2% (49.4% - 53.0%)
27. 3+ P, 2+ Rx, 3 years***	1,507	1,335	82	1,451,086	94.8% (93.6% - 95.9%)	53.0% (51.2% - 54.9%)

Algorithm	True Positive	False Positive	False Negative	True Negative	Sensitivity	PPV
28. 1+ P, 1+ H, 2+ Rx, 3 years***	1,534	1,943	55	1,450,478	96.5% (95.5% - 97.4%)	44.1% (42.5% - 45.8%)
29. 2+ P, 1+ H, 2+ Rx, 3 years***	1,526	1,530	63	1,450,891	96.0% (95.0% - 96.9%)	49.9% (48.2% - 51.7%)
30. 3+ P, 1+ H, 2+ Rx, 3 years***	1,518	1,416	71	1,451,005	95.5% (94.4% - 96.5%)	51.7% (49.9% - 53.6%)

* Specificity of 100.0% (100.0% - 100.0%). ** Specificity of 100.0% (99.9% - 100.0%). *** Specificity of 99.9% (99.9% - 99.9%).
P = physician visit; H = hospitalization; Rx = prescription; PPV = positive predictive value.
Negative predictive value was 100.0% (100.0% - 100.0%) for all algorithms.

Table 4.4. Performance estimates (Youden's J Statistic, kappa, AUC) of algorithms identifying people living with HIV in administrative health data

Algorithm	Youden's J	Kappa	AUC
1. 1+ P, 2 years	0.93 (0.92 - 0.95)	0.78 (0.77 - 0.80)	0.97 (0.96 - 0.97)
2. 2+ P, 2 years	0.92 (0.91 - 0.93)	0.88 (0.87 - 0.89)	0.96 (0.95 - 0.97)
3. 3+ P, 2 years	0.90 (0.88 - 0.91)	0.90 (0.89 - 0.92)	0.95 (0.94 - 0.96)
4. 4+ P, 2 years	0.87 (0.86 - 0.89)	0.91 (0.9 - 0.92)	0.94 (0.93 - 0.94)
5. 5+ P, 2 years	0.84 (0.83 - 0.86)	0.90 (0.88 - 0.91)	0.92 (0.91 - 0.93)
6. 6+ P, 2 years	0.81 (0.79 - 0.83)	0.88 (0.87 - 0.89)	0.91 (0.9 - 0.92)
7. 1+ P, 1+ H, 2 years	0.95 (0.94 - 0.96)	0.78 (0.76 - 0.79)	0.98 (0.97 - 0.98)
8. 2+ P, 1+ H, 2 years	0.94 (0.93 - 0.95)	0.87 (0.86 - 0.88)	0.97 (0.97 - 0.98)
9. 3+ P, 1+ H, 2 years	0.93 (0.92 - 0.94)	0.90 (0.89 - 0.91)	0.96 (0.96 - 0.97)
10. 1+ P, 2+ Rx, 2 years	0.96 (0.95 - 0.97)	0.61 (0.60 - 0.63)	0.98 (0.98 - 0.98)
11. 2+ P, 2+ Rx, 2 years	0.95 (0.94 - 0.96)	0.67 (0.65 - 0.68)	0.98 (0.97 - 0.98)
12. 3+ P, 2+ Rx, 2 years	0.94 (0.93 - 0.96)	0.68 (0.66 - 0.70)	0.97 (0.97 - 0.98)
13. 1+ P, 1+ H, 2+ Rx, 2 years	0.96 (0.96 - 0.97)	0.61 (0.59 - 0.62)	0.98 (0.98 - 0.99)
14. 2+ P, 1+ H, 2+ Rx, 2 years	0.96 (0.95 - 0.97)	0.66 (0.64 - 0.67)	0.98 (0.97 - 0.98)
15. 3+ P, 1+ H, 2+ Rx, 2 years	0.95 (0.94 - 0.96)	0.67 (0.66 - 0.69)	0.98 (0.97 - 0.98)
16. 1+ P, 3 years	0.93 (0.92 - 0.95)	0.78 (0.77 - 0.8)	0.97 (0.96 - 0.97)
17. 2+ P, 3 years	0.92 (0.91 - 0.93)	0.88 (0.87 - 0.89)	0.96 (0.95 - 0.97)
18. 3+ P, 3 years	0.90 (0.89 - 0.92)	0.91 (0.90 - 0.92)	0.95 (0.94 - 0.96)
19. 4+ P, 3 years	0.88 (0.87 - 0.90)	0.91 (0.90 - 0.92)	0.94 (0.93 - 0.95)
20. 5+ P, 3 years	0.86 (0.84 - 0.87)	0.90 (0.89 - 0.91)	0.93 (0.92 - 0.94)
21. 6+ P, 3 years	0.83 (0.81 - 0.85)	0.89 (0.88 - 0.90)	0.92 (0.91 - 0.92)
22. 1+ P, 1+ H, 3 years	0.95 (0.94 - 0.96)	0.78 (0.76 - 0.79)	0.98 (0.97 - 0.98)
23. 2+ P, 1+ H, 3 years	0.95 (0.93 - 0.96)	0.87 (0.86 - 0.88)	0.97 (0.97 - 0.98)
24. 3+ P, 1+ H, 3 years	0.93 (0.92 - 0.94)	0.90 (0.89 - 0.91)	0.97 (0.96 - 0.97)
25. 1+ P, 2+ Rx, 3 years	0.96 (0.95 - 0.97)	0.61 (0.60 - 0.63)	0.98 (0.98 - 0.98)
26. 2+ P, 2+ Rx, 3 years	0.95 (0.94 - 0.96)	0.67 (0.65 - 0.68)	0.98 (0.97 - 0.98)
27. 3+ P, 2+ Rx, 3 years	0.95 (0.94 - 0.96)	0.68 (0.66 - 0.70)	0.97 (0.97 - 0.98)
28. 1+ P, 1+ H, 2+ Rx, 3 years	0.96 (0.96 - 0.97)	0.61 (0.59 - 0.62)	0.98 (0.98 - 0.99)
29. 2+ P, 1+ H, 2+ Rx, 3 years	0.96 (0.95 - 0.97)	0.66 (0.64 - 0.67)	0.98 (0.97 - 0.98)
30. 3+ P, 1+ H, 2+ Rx, 3 years	0.95 (0.94 - 0.96)	0.67 (0.65 - 0.69)	0.98 (0.97 - 0.98)

P = physician visit; H = hospitalization; Rx = prescription; AUC = area under the receiver operating characteristic curve.

Table 4.5. Crude prevalence of HIV per 100,000 Manitobans with continuous health insurance coverage between January 1, 2016 and December 31, 2018, by algorithm

Algorithm	Number of HIV Cases	Prevalence (95% CI)
1. 1+ P, 2 years	1,836	147.6 (141 - 154.5)
2. 2+ P, 2 years	1,439	115.7 (109.9 - 121.8)
3. 3+ P, 2 years	1,304	104.8 (99.3 - 110.7)
4. 4+ P, 2 years	1,230	98.9 (93.5 - 104.6)
5. 5+ P, 2 years	1,175	94.5 (89.2 - 100)
6. 6+ P, 2 years	1,122	90.2 (85.1 - 95.6)
7. 1+ P, 1+ H, 2 years	1,889	151.9 (145.2 - 158.9)
8. 2+ P, 1+ H, 2 years	1,497	120.3 (114.4 - 126.6)
9. 3+ P, 1+ H, 2 years	1,367	109.9 (104.2 - 115.9)
10. 1+ P, 2+ Rx, 2 years	2,837	228.1 (219.8 - 236.6)
11. 2+ P, 2+ Rx, 2 years	2,465	198.2 (190.5 - 206.1)
12. 3+ P, 2+ Rx, 2 years	2,354	189.2 (181.7 - 197.0)
13. 1+ P, 1+ H, 2+ Rx, 2 years	2,886	232.0 (223.7 - 240.6)
14. 2+ P, 1+ H, 2+ Rx, 2 years	2,517	202.3 (194.6 - 210.4)
15. 3+ P, 1+ H, 2+ Rx, 2 years	2,408	193.6 (186.0 - 201.5)
16. 1+ P, 3 years	1,836	147.6 (141.0 - 154.5)
17. 2+ P, 3 years	1,442	115.9 (110.1 - 122.1)
18. 3+ P, 3 years	1,313	105.6 (100.0 - 111.4)
19. 4+ P, 3 years	1,246	100.2 (94.8 - 105.9)
20. 5+ P, 3 years	1,196	96.1 (90.8 - 101.8)
21. 6+ P, 3 years	1,155	92.9 (87.6 - 98.4)
22. 1+ P, 1+ H, 3 years	1,889	151.9 (145.2 - 158.9)
23. 2+ P, 1+ H, 3 years	1,500	120.6 (114.6 - 126.8)
24. 3+ P, 1+ H, 3 years	1,375	110.5 (104.8 - 116.5)
25. 1+ P, 2+ Rx, 3 years	2,846	228.8 (220.5 - 237.4)
26. 2+ P, 2+ Rx, 3 years	2,475	199.0 (191.3 - 207)
27. 3+ P, 2+ Rx, 3 years	2,370	190.5 (183 - 198.4)
28. 1+ P, 1+ H, 2+ Rx, 3 years	2,895	232.7 (224.4 - 241.4)
29. 2+ P, 1+ H, 2+ Rx, 3 years	2,527	203.1 (195.4 - 211.2)
30. 3+ P, 1+ H, 2+ Rx, 3 years	2,424	194.9 (187.3 - 202.8)

Note: The population denominator for all algorithms consisted of the 1,243,939 Manitobans with continuous health insurance coverage between January 1, 2016 and December 31, 2018.

4.5 Sensitivity Analyses

4.5.1 Sensitivity Analysis 1

To produce algorithm performance estimates that were comparable to the performance estimates produced by the other Canadian HIV algorithm studies, the 30 HIV administrative data algorithms were applied using only the ICD codes from Antoniou et al. (9) (ICD-9-CM: 042-044; ICD-10-CA: B20-B24). This sensitivity analysis used the same validation cohort as the main analysis and therefore has the same flow chart depicting the construction of the validation cohort (Figure 4.1) and the same table describing the validation cohort (Table 4.1), so these figures were not presented again.

The algorithm performance estimates and prevalence for this sensitivity analysis were very similar to those of the main analysis, to the extent that the algorithm selected as best would be the same regardless of the methodology that was used. Therefore, the addition of ICD-9-CM codes V08, 795.51 and ICD-10-CA codes R75 and Z71 in the main analysis had minimal added value.

The sensitivity, specificity, and PPV for all algorithms are given in Table 4.6. Algorithm sensitivity ranged from 80.9% (95% CI 78.9% - 82.8%) for Algorithm 6 to 96.4% (95% CI 95.3% - 97.2%) for Algorithms 13 and 28. The PPV of the algorithms ranged from 45.3% (95% CI 43.6% - 47.0%) for Algorithm 28 to 96.1% (95% CI 94.9% - 97.1%) for Algorithm 6. All algorithms which required two or more antiretroviral prescriptions (Algorithms 10-15 and 25-30) had PPV $\leq$ 53.3%. Specificity and NPV were very high for all algorithms. Specificity ranged from 99.9% (95% CI 99.9% - 99.9%) to 100.0% (95% CI 100.0% - 100.0%) for all algorithms. Since many algorithms had the same specificity, specificity is represented by symbols in Table 4.6. NPV was 100.0% (95% CI 100.0%-100.0%) for all algorithms and so was not included in Table 4.6.

The Youden's J Statistic, kappa, and AUC for all algorithms are given in Table 4.7. Youden's J Statistic ranged from 0.81 (95% CI 0.79 - 0.83) to 0.96 (95% CI 0.95 - 0.97) across all algorithms. The kappa statistic ranged from 0.62 (95% CI 0.60 - 0.63) to 0.91 (95% CI 0.90 - 0.92), indicating good, or very good, agreement between all algorithms and the reference standard. The AUC ranged from 0.90 (95% CI 0.89 - 0.91) to 0.98 (95% CI 0.98 - 0.99) indicating outstanding discrimination between HIV cases and HIV non-cases for all algorithms.

Table 4.6. Performance estimates (sensitivity, specificity, PPV) of algorithms identifying people living with HIV in administrative health data, Sensitivity Analysis 1

Algorithm	True Positive	False Positive	False Negative	True Negative	Sensitivity	PPV
1. 1+ P, 2 years*	1,482	682	107	1,451,739	93.3% (91.9% - 94.5%)	68.5% (66.5% - 70.4%)
2. 2+ P, 2 years*	1,460	259	129	1,452,162	91.9% (90.4% - 93.2%)	84.9% (83.2% - 86.6%)
3. 3+ P, 2 years*	1,425	134	164	1,452,287	89.7% (88.1% - 91.1%)	91.4% (89.9% - 92.8%)
4. 4+ P, 2 years*	1,385	85	204	1,452,336	87.2% (85.4% - 88.8%)	94.2% (92.9% - 95.4%)
5. 5+ P, 2 years*	1,338	64	251	1,452,357	84.2% (82.3% - 86.0%)	95.4% (94.2% - 96.5%)
6. 6+ P, 2 years*	1,285	52	304	1,452,369	80.9% (78.9% - 82.8%)	96.1% (94.9% - 97.1%)
7. 1+ P, 1+ H, 2 years*	1,493	694	96	1,451,727	94.0% (92.7% - 95.1%)	68.3% (66.3% - 70.2%)
8. 2+ P, 1+ H, 2 years*	1,472	273	117	1,452,148	92.6% (91.2% - 93.9%)	84.4% (82.6% - 86.0%)
9. 3+ P, 1+ H, 2 years*	1,441	148	148	1,452,273	90.7% (89.2% - 92.1%)	90.7% (89.2% - 92.1%)
10. 1+ P, 2+ Rx, 2 years*	1,528	1,830	61	1,450,591	96.2% (95.1% - 97.1%)	45.5% (43.8% - 47.2%)
11. 2+ P, 2+ Rx, 2 years***	1,517	1,429	72	1,450,992	95.5% (94.3% - 96.4%)	51.5% (49.7% - 53.3%)
12. 3+ P, 2+ Rx, 2 years***	1,503	1,318	86	1,451,103	94.6% (93.4% - 95.7%)	53.3% (51.4% - 55.1%)
13. 1+ P, 1+ H, 2+ Rx, 2 years***	1,531	1,841	58	1,450,580	96.4% (95.3% - 97.2%)	45.4% (43.7% - 47.1%)
14. 2+ P, 1+ H, 2+ Rx, 2 years***	1,520	1,442	69	1,450,979	95.7% (94.5% - 96.6%)	51.3% (49.5% - 53.1%)
15. 3+ P, 1+ H, 2+ Rx, 2 years***	1,507	1,331	82	1,451,090	94.8% (93.6% - 95.9%)	53.1% (51.2% - 55.0%
16. 1+ P, 3 years*	1,482	682	107	1,451,739	93.3% (91.9% - 94.5%)	68.5% (66.5% - 70.4%)
17. 2+ P, 3 years*	1,462	261	127	1,452,160	92.0% (90.6% - 93.3%)	84.9% (83.1% - 86.5%)
18. 3+ P, 3 years*	1,432	137	157	1,452,284	90.1% (88.6% - 91.5%)	91.3% (89.8% - 92.6%)
19. 4+ P, 3 years*	1,400	88	189	1,452,333	88.1% (86.4% - 89.7%)	94.1% (92.8% - 95.2%)
20. 5+ P, 3 years*	1,360	66	229	1,452,355	85.6% (83.8% - 87.3%)	95.4% (94.2% - 96.4%)
21. 6+ P, 3 years*	1,319	55	270	1,452,366	83.0% (81.1% - 84.8%)	96.0% (94.8% - 97.0%)
22. 1+ P, 1+ H, 3 years*	1,493	694	96	1,451,727	94.0% (92.7% - 95.1%)	68.3% (66.3% - 70.2%)

Algorithm	True Positive	False Positive	False Negative	True Negative	Sensitivity	PPV
23. 2+ P, 1+ H, 3 years*	1,474	275	115	1,452,146	92.8% (91.4% - 94.0%)	84.3% (82.5% - 86.0%)
24. 3+ P, 1+ H, 3 years*	1,448	151	141	1,452,270	91.1% (89.6% - 92.5%)	90.6% (89.0% - 92.0%)
25. 1+ P, 2+ Rx, 3 years***	1,528	1,839	61	1,450,582	96.2% (95.1% - 97.1%)	45.4% (43.7% - 47.1%)
26. 2+ P, 2+ Rx, 3 years***	1,517	1,440	72	1,450,981	95.5% (94.3% - 96.4%)	51.3% (49.5% - 53.1%)
27. 3+ P, 2+ Rx, 3 years***	1,507	1,330	82	1,451,091	94.8% (93.6% - 95.9%)	53.1% (51.3% - 55.0%)
28. 1+ P, 1+ H, 2+ Rx, 3 years***	1,531	1,850	58	1,450,571	96.4% (95.3% - 97.2%)	45.3% (43.6% - 47.0%)
29. 2+ P, 1+ H, 2+ Rx, 3 years***	1,520	1,453	69	1,450,968	95.7% (94.5% - 96.6%)	51.1% (49.3% - 52.9%)
30. 3+ P, 1+ H, 2+ Rx, 3 years***	1,511	1,343	78	1,451,078	95.1% (93.9% - 96.1%)	52.9% (51.1% - 54.8%)

* Specificity of 100.0% (100.0% - 100.0%). ** Specificity of 100.0% (99.9% - 100.0%). *** Specificity of 99.9% (99.9% - 99.9%).
P = physician visit; H = hospitalization; Rx = prescription; PPV = positive predictive value.
Negative predictive value was 100.0% (100.0% - 100.0%) for all algorithms.

Table 4.7. Performance estimates (Youden's J Statistic, kappa, AUC) of algorithms identifying people living with HIV in administrative health data, Sensitivity Analysis 1

Algorithm	Youden's J	Kappa	AUC
1. 1+ P, 2 years	0.93 (0.92 - 0.94)	0.79 (0.78 - 0.8)	0.97 (0.96 - 0.97)
2. 2+ P, 2 years	0.92 (0.91 - 0.93)	0.88 (0.87 - 0.89)	0.96 (0.95 - 0.97)
3. 3+ P, 2 years	0.90 (0.88 - 0.91)	0.91 (0.89 - 0.92)	0.95 (0.94 - 0.96)
4. 4+ P, 2 years	0.87 (0.86 - 0.89)	0.91 (0.89 - 0.92)	0.94 (0.93 - 0.94)
5. 5+ P, 2 years	0.84 (0.82 - 0.86)	0.89 (0.88 - 0.91)	0.92 (0.91 - 0.93)
6. 6+ P, 2 years	0.81 (0.79 - 0.83)	0.88 (0.87 - 0.89)	0.9 (0.89 - 0.91)
7. 1+ P, 1+ H, 2 years	0.94 (0.93 - 0.95)	0.79 (0.78 - 0.8)	0.97 (0.96 - 0.98)
8. 2+ P, 1+ H, 2 years	0.93 (0.91 - 0.94)	0.88 (0.87 - 0.89)	0.96 (0.96 - 0.97)
9. 3+ P, 1+ H, 2 years	0.91 (0.89 - 0.92)	0.91 (0.9 - 0.92)	0.95 (0.95 - 0.96)
10. 1+ P, 2+ Rx, 2 years	0.96 (0.95 - 0.97)	0.62 (0.6 - 0.63)	0.98 (0.98 - 0.98)
11. 2+ P, 2+ Rx, 2 years	0.95 (0.94 - 0.96)	0.67 (0.65 - 0.68)	0.98 (0.97 - 0.98)
12. 3+ P, 2+ Rx, 2 years	0.94 (0.93 - 0.96)	0.68 (0.67 - 0.7)	0.97 (0.97 - 0.98)
13. 1+ P, 1+ H, 2+ Rx, 2 years	0.96 (0.95 - 0.97)	0.62 (0.6 - 0.63)	0.98 (0.98 - 0.99)
14. 2+ P, 1+ H, 2+ Rx, 2 years	0.96 (0.95 - 0.97)	0.67 (0.65 - 0.68)	0.98 (0.97 - 0.98)
15. 3+ P, 1+ H, 2+ Rx, 2 years	0.95 (0.94 - 0.96)	0.68 (0.66 - 0.7)	0.97 (0.97 - 0.98)
16. 1+ P, 3 years	0.93 (0.92 - 0.94)	0.79 (0.78 - 0.8)	0.97 (0.96 - 0.97)
17. 2+ P, 3 years	0.92 (0.91 - 0.93)	0.88 (0.87 - 0.89)	0.96 (0.95 - 0.97)
18. 3+ P, 3 years	0.90 (0.89 - 0.92)	0.91 (0.9 - 0.92)	0.95 (0.94 - 0.96)
19. 4+ P, 3 years	0.88 (0.87 - 0.90)	0.91 (0.9 - 0.92)	0.94 (0.93 - 0.95)
20. 5+ P, 3 years	0.86 (0.84 - 0.87)	0.9 (0.89 - 0.91)	0.93 (0.92 - 0.94)
21. 6+ P, 3 years	0.83 (0.81 - 0.85)	0.89 (0.88 - 0.9)	0.92 (0.91 - 0.92)
22. 1+ P, 1+ H, 3 years	0.94 (0.93 - 0.95)	0.79 (0.78 - 0.8)	0.97 (0.96 - 0.98)
23. 2+ P, 1+ H, 3 years	0.93 (0.91 - 0.94)	0.88 (0.87 - 0.89)	0.96 (0.96 - 0.97)
24. 3+ P, 1+ H, 3 years	0.91 (0.90 - 0.93)	0.91 (0.9 - 0.92)	0.96 (0.95 - 0.96)
25. 1+ P, 2+ Rx, 3 years	0.96 (0.95 - 0.97)	0.62 (0.6 - 0.63)	0.98 (0.98 - 0.98)
26. 2+ P, 2+ Rx, 3 years	0.95 (0.94 - 0.96)	0.67 (0.65 - 0.68)	0.98 (0.97 - 0.98)
27. 3+ P, 2+ Rx, 3 years	0.95 (0.94 - 0.96)	0.68 (0.66 - 0.7)	0.97 (0.97 - 0.98)
28. 1+ P, 1+ H, 2+ Rx, 3 years	0.96 (0.95 - 0.97)	0.62 (0.6 - 0.63)	0.98 (0.98 - 0.99)
29. 2+ P, 1+ H, 2+ Rx, 3 years	0.96 (0.95 - 0.97)	0.67 (0.65 - 0.68)	0.98 (0.97 - 0.98)
30. 3+ P, 1+ H, 2+ Rx, 3 years	0.95 (0.94 - 0.96)	0.68 (0.66 - 0.7)	0.98 (0.97 - 0.98)

P = physician visit; H = hospitalization; Rx = prescription; AUC = area under the receiver operating characteristic curve.

Table 4.8 shows the algorithm derived prevalence of HIV among Manitobans with continuous health insurance coverage between January 1, 2016 and December 31, 2018. The prevalence calculated using the algorithms ranged from 89.8 people diagnosed with HIV per 100,000 population (95% CI 84.7 - 95.2) for Algorithm 6 to 227.3 people diagnosed with HIV per 100,000 population (95% CI 219.0 - 235.8) for Algorithm 28. Prevalence calculated using algorithms that required antiretroviral prescription data was much greater than the prevalence calculated using algorithms that did not require prescription data. Excluding those algorithms which required two or more antiretroviral prescriptions, the maximum prevalence was 146.1 people diagnosed with HIV per 100,000 population (95% CI 139.6 – 153.0) for Algorithms 7 and 22.

Table 4.8. Crude prevalence of HIV per 100,000 Manitobans with continuous health insurance coverage between January 1, 2016 and December 31, 2018, by algorithm, Sensitivity Analysis 1

Algorithm	Number of HIV Cases	Prevalence (95% CI)
1. 1+ P, 2 years	1,810	145.5 (139.0 - 152.4)
2. 2+ P, 2 years	1,431	115.0 (109.2 - 121.2)
3. 3+ P, 2 years	1,299	104.4 (98.9 - 110.3)
4. 4+ P, 2 years	1,227	98.6 (93.3 - 104.3)
5. 5+ P, 2 years	1,170	94.1 (88.8 - 99.6)
6. 6+ P, 2 years	1,117	89.8 (84.7 - 95.2)
7. 1+ P, 1+ H, 2 years	1,818	146.1 (139.6 - 153)
8. 2+ P, 1+ H, 2 years	1,441	115.8 (110.0 - 122.0)
9. 3+ P, 1+ H, 2 years	1,309	105.2 (99.7 - 111.1)
10. 1+ P, 2+ Rx, 2 years	2,811	226.0 (217.8 - 234.5)
11. 2+ P, 2+ Rx, 2 years	2,457	197.5 (189.9 - 205.5)
12. 3+ P, 2+ Rx, 2 years	2,350	188.9 (181.4 - 196.7)
13. 1+ P, 1+ H, 2+ Rx, 2 years	2,818	226.5 (218.3 - 235.1)
14. 2+ P, 1+ H, 2+ Rx, 2 years	2,466	198.2 (190.6 - 206.2)
15. 3+ P, 1+ H, 2+ Rx, 2 years	2,359	189.6 (182.1 - 197.4)
16. 1+ P, 3 years	1,810	145.5 (139.0 - 152.4)
17. 2+ P, 3 years	1,434	115.3 (109.5 - 121.4)
18. 3+ P, 3 years	1,308	105.1 (99.6 - 111.0)
19. 4+ P, 3 years	1,244	100.0 (94.6 - 105.7)
20. 5+ P, 3 years	1,192	95.8 (90.5 - 101.4)
21. 6+ P, 3 years	1,152	92.6 (87.4 - 98.1)

Algorithm	Number of HIV Cases	Prevalence (95% CI)
22. 1+ P, 1+ H, 3 years	1,818	146.1 (139.6 - 153.0)
23. 2+ P, 1+ H, 3 years	1,444	116.1 (110.2 - 122.2)
24. 3+ P, 1+ H, 3 years	1,318	106.0 (100.4 - 111.8)
25. 1+ P, 2+ Rx, 3 years	2,820	226.7 (218.5 - 235.2)
26. 2+ P, 2+ Rx, 3 years	2,467	198.3 (190.6 - 206.3)
27. 3+ P, 2+ Rx, 3 years	2,366	190.2 (182.7 - 198)
28. 1+ P, 1+ H, 2+ Rx, 3 years	2,827	227.3 (219.0 - 235.8)
29. 2+ P, 1+ H, 2+ Rx, 3 years	2,476	199.0 (191.4 - 207.0)
30. 3+ P, 1+ H, 2+ Rx, 3 years	2,375	190.9 (183.4 - 198.8)

P = physician visit; H = hospitalization; Rx = prescription
Note: The population denominator for all algorithms consisted of the 1,243,939 Manitobans with continuous health insurance coverage between January 1, 2016 and December 31, 2018.

4.5.2 Sensitivity Analysis 2

To produce algorithm performance estimates that were comparable to the performance estimates produced by the other Canadian HIV algorithm studies, the 30 HIV administrative data algorithms were applied using only the ICD codes used by Nosyk et al. (10) (ICD-9-CM: 042-044, V08, 795.71; ICD-10-CA: B20-B24, R75, Z21). This sensitivity analysis used the same validation cohort as the main analysis and therefore has the same flow chart depicting the construction of the validation cohort (Figure 4.1) and the same table describing the validation cohort (Table 4.1), so these were not presented again.

The algorithm performance estimates and prevalence for this sensitivity analysis were nearly identical to those of the main analysis, to the extent that the algorithm selected as best would be the same regardless of which methodology was used. Therefore, the addition of ICD-9-CM code 079.53 and ICD-10-CA codes F02.4 and O98.7 in the main analysis had minimal added value.

The sensitivity, specificity, and PPV for all algorithms are given in Table 4.9. Algorithm sensitivity ranged from 81.1% (95% CI 79.1% - 83.0%) for Algorithm 6 to 96.5% (95% CI 95.5% - 97.4%) for Algorithms 13 and 28. The PPV of the algorithms ranged from 44.1% (95% CI 42.5% - 45.8%) for Algorithm 28 to 96.0% (95% CI 94.8% - 97.0%) for Algorithm 6. All algorithms which required two or more antiretroviral prescriptions (Algorithms 10-15 and 25-30) had PPV ≤ 53.2%. Specificity and NPV were very high for all algorithms. Specificity ranged from 99.9% (95% CI 99.9% - 99.9%) to 100.0% (95% CI 100.0% - 100.0%) for all algorithms.

Table 4.9. Performance estimates (sensitivity, specificity, PPV) of algorithms identifying people living with HIV in administrative health data, Sensitivity Analysis 2

Algorithm	True Positive	False Positive	False Negative	True Negative	Sensitivity	PPV
1. 1+ P, 2 years*	1,484	710	105	1,451,711	93.4% (92.1% - 94.6%)	67.6% (65.6% - 69.6%)
2. 2+ P, 2 years*	1,462	268	127	1,452,153	92.0% (90.6% - 93.3%)	84.5% (82.7% - 86.2%)
3. 3+ P, 2 years*	1,427	139	162	1,452,282	89.8% (88.2% - 91.3%)	91.1% (89.6% - 92.5%)
4. 4+ P, 2 years*	1,388	87	201	1,452,334	87.4% (85.6% - 89.0%)	94.1% (92.8% - 95.3%)
5. 5+ P, 2 years*	1,342	66	247	1,452,355	84.5% (82.6% - 86.2%)	95.3% (94.1% - 96.4%)
6. 6+ P, 2 years*	1,289	54	300	1,452,367	81.1% (79.1% - 83.0%)	96.0% (94.8% - 97.0%)
7. 1+ P, 1+ H, 2 years**	1,514	793	75	1,451,628	95.3% (94.1% - 96.3%)	65.6% (63.7% - 67.6%)
8. 2+ P, 1+ H, 2 years*	1,499	356	90	1,452,065	94.3% (93.1% - 95.4%)	80.8% (78.9% - 82.6%)
9. 3+ P, 1+ H, 2 years*	1,475	227	114	1,452,194	92.8% (91.4% - 94.1%)	86.7% (85.0% - 88.2%)
10. 1+ P, 2+ Rx, 2 year***s	1,528	1,858	61	1,450,563	96.2% (95.1% - 97.1%)	45.1% (43.4% - 46.8%)
11. 2+ P, 2+ Rx, 2 years***	1,518	1,438	71	1,450,983	95.5% (94.4% - 96.5%)	51.4% (49.5% - 53.2%)
12. 3+ P, 2+ Rx, 2 years***	1,503	1,323	86	1,451,098	94.6% (93.4% - 95.7%)	53.2% (51.3% - 55.0%)
13. 1+ P, 1+ H, 2+ Rx, 2 years***	1,534	1,933	55	1,450,488	96.5% (95.5% - 97.4%)	44.3% (42.6% - 45.9%)
14. 2+ P, 1+ H, 2+ Rx, 2 years***	1,526	1,518	63	1,450,903	96.0% (95.0% - 96.9%)	50.1% (48.3% - 51.9%)
15. 3+ P, 1+ H, 2+ Rx, 2 years***	1,514	1,403	75	1,451,018	95.3% (94.1% - 96.3%)	51.9% (50.1% - 53.7%)
16. 1+ P, 3 years*	1,484	710	105	1,451,711	93.4% (92.1% - 94.6%)	67.6% (65.6% - 69.6%)
17. 2+ P, 3 years*	1,465	270	124	1,452,151	92.2% (90.8% - 93.5%)	84.4% (82.7% - 86.1%)
18. 3+ P, 3 years*	1,434	142	155	1,452,279	90.3% (88.7% - 91.7%)	91.0% (89.5% - 92.4%)
19. 4+ P, 3 years*	1,402	90	187	1,452,331	88.2% (86.5% - 89.8%)	94.0% (92.6% - 95.1%)
20. 5+ P, 3 years*	1,363	68	226	1,452,353	85.8% (84.0% - 87.5%)	95.3% (94.0% - 96.3%)

Algorithm	True Positive	False Positive	False Negative	True Negative	Sensitivity	PPV
21. 6+ P, 3 years*	1,321	57	268	1,452,364	83.1% (81.2% - 84.9%)	95.9% (94.7% - 96.9%)
22. 1+ P, 1+ H, 3 years**	1,514	793	75	1,451,628	95.3% (94.1% - 96.3%)	65.6% (63.7% - 67.6%)
23. 2+ P, 1+ H, 3 years*	1,502	358	87	1,452,063	94.5% (93.3% - 95.6%)	80.8% (78.9% - 82.5%)
24. 3+ P, 1+ H, 3 years*	1,481	230	108	1,452,191	93.2% (91.9% - 94.4%)	86.6% (84.9% - 88.1%)
25. 1+ P, 2+ Rx, 3 years***	1,528	1,867	61	1,450,554	96.2% (95.1% - 97.1%)	45.0% (43.3% - 46.7%)
26. 2+ P, 2+ Rx, 3 years***	1,518	1,449	71	1,450,972	95.5% (94.4% - 96.5%)	51.2% (49.4% - 53.0%)
27. 3+ P, 2+ Rx, 3 years***	1,507	1,335	82	1,451,086	94.8% (93.6% - 95.9%)	53.0% (51.2% - 54.9%)
28. 1+ P, 1+ H, 2+ Rx, 3 years***	1,534	1,942	55	1,450,479	96.5% (95.5% - 97.4%)	44.1% (42.5% - 45.8%)
29. 2+ P, 1+ H, 2+ Rx, 3 years***	1,526	1,529	63	1,450,892	96.0% (95.0% - 96.9%)	50.0% (48.2% - 51.7%)
30. 3+ P, 1+ H, 2+ Rx, 3 years***	1,518	1,415	71	1,451,006	95.5% (94.4% - 96.5%)	51.8% (49.9% - 53.6%)

* Specificity of 100.0% (100.0% - 100.0%). ** Specificity of 100.0% (99.9% - 100.0%). *** Specificity of 99.9% (99.9% - 99.9%).
P = physician visit; H = hospitalization; Rx = prescription; PPV = positive predictive value.
Negative predictive value was 100.0% (100.0% - 100.0%) for all algorithms.

Since many algorithms had the same specificity, specificity is represented by symbols in Table 4.9. NPV was 100.0% (95% CI 100.0% - 100.0%) for all algorithms and so was not included in Table 4.9.

The Youden's J Statistic, kappa, and AUC for all algorithms are given in Table 4.10. Youden's J Statistic ranged from 0.81 (95% CI 0.79 - 0.83) to 0.96 (95% CI 0.96 - 0.97) across all algorithms. The kappa statistic ranged from 0.61 (95% CI 0.59 - 0.62) to 0.91 (95% CI 0.90 - 0.92), indicating good, or very good, agreement between all algorithms and the reference standard. The AUC ranged from 0.91 (95% CI 0.90 - 0.92) to 0.98 (95% CI 0.98 - 0.99) indicating outstanding discrimination between HIV cases and HIV non-cases for all algorithms.

Table 4.10. Performance estimates (Youden's J Statistic, kappa, AUC) of algorithms identifying people living with HIV in administrative health data, Sensitivity Analysis 2

Algorithm	Youden's J	Kappa	AUC
1. 1+ P, 2 years	0.93 (0.92 - 0.95)	0.78 (0.77 - 0.80)	0.97 (0.96 - 0.97)
2. 2+ P, 2 years	0.92 (0.91 - 0.93)	0.88 (0.87 - 0.89)	0.96 (0.95 - 0.97)
3. 3+ P, 2 years	0.90 (0.88 - 0.91)	0.90 (0.89 - 0.92)	0.95 (0.94 - 0.96)
4. 4+ P, 2 years	0.87 (0.86 - 0.89)	0.91 (0.90 - 0.92)	0.94 (0.93 - 0.94)
5. 5+ P, 2 years	0.84 (0.83 - 0.86)	0.90 (0.88 - 0.91)	0.92 (0.91 - 0.93)
6. 6+ P, 2 years	0.81 (0.79 - 0.83)	0.88 (0.87 - 0.89)	0.91 (0.90 - 0.92)
7. 1+ P, 1+ H, 2 years	0.95 (0.94 - 0.96)	0.78 (0.76 - 0.79)	0.98 (0.97 - 0.98)
8. 2+ P, 1+ H, 2 years	0.94 (0.93 - 0.95)	0.87 (0.86 - 0.88)	0.97 (0.97 - 0.98)
9. 3+ P, 1+ H, 2 years	0.93 (0.92 - 0.94)	0.90 (0.89 - 0.91)	0.96 (0.96 - 0.97)
10. 1+ P, 2+ Rx, 2 years	0.96 (0.95 - 0.97)	0.61 (0.60 - 0.63)	0.98 (0.98 - 0.98)
11. 2+ P, 2+ Rx, 2 years	0.95 (0.94 - 0.96)	0.67 (0.65 - 0.68)	0.98 (0.97 - 0.98)
12. 3+ P, 2+ Rx, 2 years	0.94 (0.93 - 0.96)	0.68 (0.66 - 0.70)	0.97 (0.97 - 0.98)
13. 1+ P, 1+ H, 2+ Rx, 2 years	0.96 (0.96 - 0.97)	0.61 (0.59 - 0.62)	0.98 (0.98 - 0.99)
14. 2+ P, 1+ H, 2+ Rx, 2 years	0.96 (0.95 - 0.97)	0.66 (0.64 - 0.67)	0.98 (0.97 - 0.98)
15. 3+ P, 1+ H, 2+ Rx, 2 years	0.95 (0.94 - 0.96)	0.67 (0.66 - 0.69)	0.98 (0.97 - 0.98)
16. 1+ P, 3 years	0.93 (0.92 - 0.95)	0.78 (0.77 - 0.80)	0.97 (0.96 - 0.97)
17. 2+ P, 3 years	0.92 (0.91 - 0.93)	0.88 (0.87 - 0.89)	0.96 (0.95 - 0.97)
18. 3+ P, 3 years	0.90 (0.89 - 0.92)	0.91 (0.90 - 0.92)	0.95 (0.94 - 0.96)
19. 4+ P, 3 years	0.88 (0.87 - 0.90)	0.91 (0.90 - 0.92)	0.94 (0.93 - 0.95)
20. 5+ P, 3 years	0.86 (0.84 - 0.87)	0.90 (0.89 - 0.91)	0.93 (0.92 - 0.94)
21. 6+ P, 3 years	0.83 (0.81 - 0.85)	0.89 (0.88 - 0.90)	0.92 (0.91 - 0.92)

Algorithm	Youden's J	Kappa	AUC
22. 1+ P, 1+ H, 3 years	0.95 (0.94 - 0.96)	0.78 (0.76 - 0.79)	0.98 (0.97 - 0.98)
23. 2+ P, 1+ H, 3 years	0.95 (0.93 - 0.96)	0.87 (0.86 - 0.88)	0.97 (0.97 - 0.98)
24. 3+ P, 1+ H, 3 years	0.93 (0.92 - 0.94)	0.9 (0.89 - 0.91)	0.97 (0.96 - 0.97)
25. 1+ P, 2+ Rx, 3 years	0.96 (0.95 - 0.97)	0.61 (0.60 - 0.63)	0.98 (0.98 - 0.98)
26. 2+ P, 2+ Rx, 3 years	0.95 (0.94 - 0.96)	0.67 (0.65 - 0.68)	0.98 (0.97 - 0.98)
27. 3+ P, 2+ Rx, 3 years	0.95 (0.94 - 0.96)	0.68 (0.66 - 0.70)	0.97 (0.97 - 0.98)
28. 1+ P, 1+ H, 2+ Rx, 3 years	0.96 (0.96 - 0.97)	0.61 (0.59 - 0.62)	0.98 (0.98 - 0.99)
29. 2+ P, 1+ H, 2+ Rx, 3 years	0.96 (0.95 - 0.97)	0.66 (0.64 - 0.67)	0.98 (0.97 - 0.98)
30. 3+ P, 1+ H, 2+ Rx, 3 years	0.95 (0.94 - 0.96)	0.67 (0.66 - 0.69)	0.98 (0.97 - 0.98)

P = physician visit; H = hospitalization; Rx = prescription; AUC = area under the receiver operating characteristic curve.

Table 4.11 shows the algorithm derived prevalence of HIV among Manitobans with continuous health insurance coverage between January 1, 2016 and December 31, 2018. The prevalence calculated using the algorithms ranged from 90.2 people diagnosed with HIV per 100% population (95% CI 85.1 - 95.6) for Algorithm 6 to 232.6 people diagnosed with HIV per 100,000 population (95% CI 224.3 - 241.3) for Algorithm 28. Prevalence calculated using algorithms that required antiretroviral prescription data was much greater than the prevalence calculated using algorithms that did not require prescription data. Excluding those algorithms which required two or more antiretroviral prescriptions, the maximum prevalence was 151.8 people diagnosed with HIV per 100,000 population (95% CI 145.1 – 158.8) for Algorithms 7 and 22.

Table 4.11. Crude prevalence of HIV per 100,000 Manitobans with continuous health insurance coverage between January 1, 2016 and December 31, 2018, by algorithm, Sensitivity Analysis 2

Algorithm	Number of HIV Cases	Prevalence (95% CI)
1. 1+ P, 2 years	1,836	147.6 (141.0 - 154.5)
2. 2+ P, 2 years	1,439	115.7 (109.9 - 121.8)
3. 3+ P, 2 years	1,304	104.8 (99.3 - 110.7)
4. 4+ P, 2 years	1,230	98.9 (93.5 - 104.6)
5. 5+ P, 2 years	1,175	94.5 (89.2 - 100.0)
6. 6+ P, 2 years	1,122	90.2 (85.1 - 95.6)
7. 1+ P, 1+ H, 2 years	1,888	151.8 (145.1 - 158.8)
8. 2+ P, 1+ H, 2 years	1,496	120.3 (114.3 - 126.5)

Algorithm	Number of HIV Cases	Prevalence (95% CI)
9. 3+ P, 1+ H, 2 years	1,366	109.8 (104.1 - 115.8)
10. 1+ P, 2+ Rx, 2 years	2,837	228.1 (219.8 - 236.6)
11. 2+ P, 2+ Rx, 2 years	2,465	198.2 (190.5 - 206.1)
12. 3+ P, 2+ Rx, 2 years	2,354	189.2 (181.7 - 197.0)
13. 1+ P, 1+ H, 2+ Rx, 2 years	2,885	231.9 (223.6 - 240.5)
14. 2+ P, 1+ H, 2+ Rx, 2 years	2,516	202.3 (194.5 - 210.3)
15. 3+ P, 1+ H, 2+ Rx, 2 years	2,407	193.5 (185.9 - 201.4)
16. 1+ P, 3 years	1,836	147.6 (141.0 - 154.5)
17. 2+ P, 3 years	1,442	115.9 (110.1 - 122.1)
18. 3+ P, 3 years	1,313	105.6 (100.0 - 111.4)
19. 4+ P, 3 years	1,246	100.2 (94.8 - 105.9)
20. 5+ P, 3 years	1,196	96.1 (90.8 - 101.8)
21. 6+ P, 3 years	1,155	92.9 (87.6 - 98.4)
22. 1+ P, 1+ H, 3 years	1,888	151.8 (145.1 - 158.8)
23. 2+ P, 1+ H, 3 years	1,499	120.5 (114.6 - 126.8)
24. 3+ P, 1+ H, 3 years	1,374	110.5 (104.8 - 116.5)
25. 1+ P, 2+ Rx, 3 years	2,846	228.8 (220.5 - 237.4)
26. 2+ P, 2+ Rx, 3 years	2,475	199.0 (191.3 - 207.0)
27. 3+ P, 2+ Rx, 3 years	2,370	190.5 (183.0 - 198.4)
28. 1+ P, 1+ H, 2+ Rx, 3 years	2,894	232.6 (224.3 - 241.3)
29. 2+ P, 1+ H, 2+ Rx, 3 years	2,526	203.1 (195.3 - 211.1)
30. 3+ P, 1+ H, 2+ Rx, 3 years	2,423	194.8 (187.2 - 202.7)

P = physician visit; H = hospitalization; Rx = prescription.
Note: The population denominator for all algorithms consisted of the 1,243,939 Manitobans with continuous health insurance coverage between January 1, 2016 and December 31, 2018.

4.5.3 Sensitivity Analysis 3

To validate the algorithms in a population consisting of adults only, this sensitivity analysis restricted the validation cohort to those individuals who were at least 18 years old as of January 1, 2007. This excluded 487,503 individuals from the cohort (Figure 4.6). The validation cohort used for this sensitivity analysis consisted of a total of 966,507 individuals, 1,452 (0.2%) were HIV cases and 965,055 (99.8%) were HIV non-cases.

The descriptive characteristics of the validation cohort used in this sensitivity analysis were similar to the descriptive characteristics of the validation cohort used in the main analysis, with a slightly higher average age due to the exclusion of children. All performance estimates for this sensitivity analysis had the same, or higher, values when compared with the main analysis, except

for specificity which was marginally lower. The slightly lower specificity was due to the smaller number of individuals included in the validation cohort (966,507 people for this sensitivity analysis compared to 1,454,010 people in the main analysis).

The results of this sensitivity analysis suggest that the algorithms perform marginally better when applied to only adults. This is likely because very few cases of HIV were diagnosed in people under 18 years of age so by removing their records from the analysis, the misclassification of HIV non-cases as HIV cases was reduced.

Characteristics to describe the validation cohort using in this sensitivity analysis, stratified by HIV cases and HIV non-cases are provided in Table 4.12. The index date at which age, sex, area of residence, and area-level household income were measured for HIV cases and

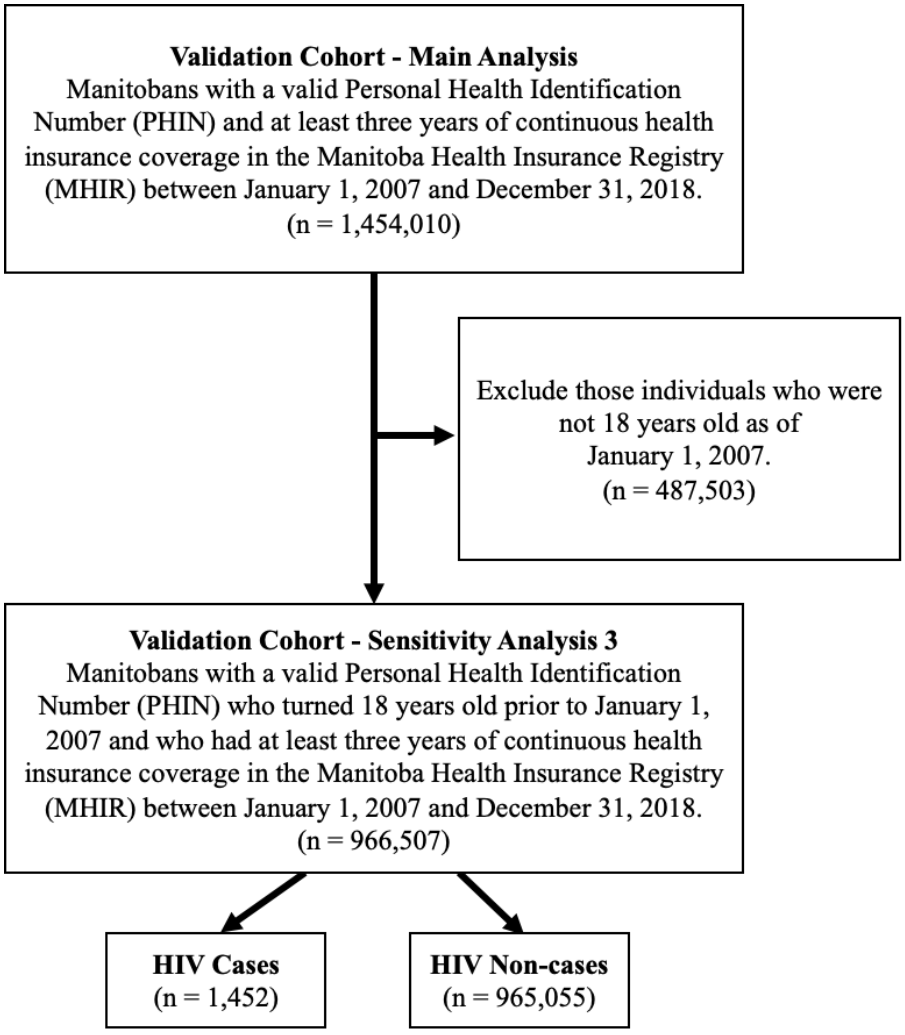

Figure 4.6. Construction of the validation cohort, Sensitivity Analysis 3

HIV non-cases was defined as the start of insurance coverage (if insurance coverage started after January 1, 2007) or January 1, 2007. The number of people who could not be assigned to an income quintile (e.g. Public Trustees) was small, so they were included with Q1.

The sensitivity, specificity, and PPV for all algorithms are given in Table 4.13. Algorithm sensitivity ranged from 82.8% (95% CI 80.7% - 84.7%) for Algorithm 6 to 97.5% (95% CI 96.5% - 98.2%) for Algorithms 13 and 28. The PPV of the algorithms ranged from 51.8% (95% CI 49.9% - 53.7%) for Algorithm 28 to 97.8% (95% CI 96.8% - 98.6%) for Algorithm 6. All algorithms which required two or more antiretroviral prescriptions and only one or more physician visit(Algorithms 10, 13, 25, 28) had PPV ≤ 53.1%. Specificity and NPV were very high for all algorithms. Specificity ranged from 99.9% (99.9% - 99.9%) to 100.0% (95% CI 99.9% - 100.0%) for all algorithms. Since many algorithms had the same specificity, specificity is represented by symbols in Table 4.13. NPV was 100.0% (95% CI 100.0% - 100.0%) for all algorithms and so was not included in Table 4.13.

The Youden's J Statistic, kappa, and AUC for all algorithms are given in Table 4.14. Youden's J Statistic ranged from 0.83 (95% CI 0.81 - 0.85) to 0.97 (95% CI 0.97 - 0.98) across all algorithms. The kappa statistic ranged from 0.68 (95% CI 0.66 - 0.69) to 0.93 (95% CI 0.92 - 0.94), indicating good, or very good, agreement between all algorithms and the reference standard. The AUC ranged from 0.91 (95% CI 0.90 - 0.92) to 0.99 (95% CI 0.98 - 0.99) indicating outstanding discrimination between HIV cases and HIV non-cases for all algorithms.

Table 4.15 shows the algorithm derived prevalence of HIV among Manitobans with continuous health insurance coverage between January 1, 2016 and December 31, 2018. The prevalence calculated using the algorithms ranged from 127.3 people diagnosed with HIV per 100% population (95% CI 119.7 – 135.4) for Algorithm 6 to 275.1 people diagnosed with HIV per 100,000 population (95% CI 263.8 – 286.8) for Algorithm 28. Prevalence calculated using algorithms that required antiretroviral prescription data was much greater than the prevalence calculated using algorithms that did not require prescription data. Excluding those algorithms which required two or more antiretroviral prescriptions, the maximum prevalence was 193.0 people diagnosed with HIV per 100,000 population (95% CI 183.6 - 202.9) for Algorithms 7 and 22.

Table 4.12. Description of the validation cohort, Sensitivity Analysis 3

Characteristic	HIV Case (N = 1,452)	HIV Non-Case (N = 965,055)
Age at index date, years	37.9 ± 11.1	45.0 ± 17.6
[minimum, maximum]	[18, 83.8]	[18, 109.4]
Sex		
Male	987 (68)	472,985 (49)
Female	465 (32)	492,070 (51)
Area of residence (RHA)		
IERHA	68 (4.7)	89,267 (9.2)
NHR	53 (3.7)	46,102 (4.8)
SH-SS	63 (4.3)	123,009 (12.7)
PMH	58 (4.0)	128,221 (13.3)
WRHA	1,203 (82.9)	575,927 (59.7)
Public Trustee/Not Found	7 (0.5)	2,529 (0.3)
Income quintile		
Q1-Lowest/Not Found	751 (51.7)	198,169 (20.5)
Q2	274 (18.9)	194,902 (20.2)
Q3	203 (14.0)	193,008 (20.0)
Q4	121 (8.3)	191,517 (19.8)
Q5-Highest	103 (7.1)	187,459 (19.4)
Health insurance coverage, years	10.3 ± 2.6	10.6 ± 2.6
Age at HIV case identification, years*	42.1 ± 10.9	-
Year of HIV case identification*		
2007	201 (13.8)	-
2008	152 (10.5)	-
2009	297 (20.5)	-
2010	228 (15.7)	-
2011	101 (7.0)	-
2012	86 (5.9)	-
2013	92 (6.3)	-
2014	67 (4.6)	-
2015	73 (5.0)	-
2016	54 (3.7)	-
2017	50 (3.4)	-
2018	51 (3.5)	-

Notes:
Data are mean ± standard deviation or n (%)
*calculated at the date of earliest positive result on a confirmatory HIV laboratory test in the study observation period, for HIV cases only.
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table 4.13. Performance estimates (sensitivity, specificity, PPV) of algorithms identifying people living with HIV in administrative health data, Sensitivity Analysis 3

Algorithm	True Positive	False Positive	False Negative	True Negative	Sensitivity	PPV
1. 1+ P, 2 years*	1,367	464	85	964,591	94.2% (92.8% - 95.3%)	74.7% (72.6% - 76.6%)
2. 2+ P, 2 years*	1,349	173	103	964,882	92.9% (91.5% - 94.2%)	88.6% (86.9% - 90.2%)
3. 3+ P, 2 years*	1,320	76	132	964,979	90.9% (89.3% - 92.3%)	94.6% (93.2% - 95.7%)
4. 4+ P, 2 years*	1,289	44	163	965,011	88.8% (87.0% - 90.4%)	96.7% (95.6% - 97.6%)
5. 5+ P, 2 years*	1,249	35	203	965,020	86.0% (84.1% - 87.8%)	97.3% (96.2% - 98.1%)
6. 6+ P, 2 years*	1,202	27	250	965,028	82.8% (80.7% - 84.7%)	97.8% (96.8% - 98.6%)
7. 1+ P, 1+ H, 2 years****	1,397	540	55	964,515	96.2% (95.1% - 97.1%)	72.1% (70.1% - 74.1%)
8. 2+ P, 1+ H, 2 years*	1,386	254	66	964,801	95.5% (94.3% - 96.5%)	84.5% (82.7% - 86.2%)
9. 3+ P, 1+ H, 2 years*	1,368	157	84	964,898	94.2% (92.9% - 95.4%)	89.7% (88.1% - 91.2%)
10. 1+ P, 2+ Rx, 2 year*s	1,409	1,243	43	963,812	97.0% (96.0% - 97.9%)	53.1% (51.2% - 55.0%)
11. 2+ P, 2+ Rx, 2 years***	1,400	971	52	964,084	96.4% (95.3% - 97.3%)	59.1% (57.0% - 61.0%)
12. 3+ P, 2+ Rx, 2 years***	1,388	886	64	964,169	95.6% (94.4% - 96.6%)	61.0% (59.0% - 63.1%)
13. 1+ P, 1+ H, 2+ Rx, 2 years***	1,415	1,311	37	963,744	97.5% (96.5% - 98.2%)	51.9% (50.0% - 53.8%)

Algorithm	True Positive	False Positive	False Negative	True Negative	Sensitivity	PPV
14. 2+ P, 1+ H, 2+ Rx, 2 years***	1,408	1,044	44	964,011	97.0% (96.0% - 97.8%)	57.4% (55.4% - 59.4%)
15. 3+ P, 1+ H, 2+ Rx, 2 years***	1,399	959	53	964,096	96.4% (95.3% - 97.3%)	59.3% (57.3% - 61.3%)
16. 1+ P, 3 years*	1,367	464	85	964,591	94.2% (92.8% - 95.3%)	74.7% (72.6% - 76.6%)
17. 2+ P, 3 years*	1,352	174	100	964,881	93.1% (91.7% - 94.4%)	88.6% (86.9% - 90.2%)
18. 3+ P, 3 years*	1,327	77	125	964,978	91.4% (89.8% - 92.8%)	94.5% (93.2% - 95.7%)
19. 4+ P, 3 years*	1,301	45	151	965,010	89.6% (87.9% - 91.1%)	96.7% (95.6% - 97.6%)
20. 5+ P, 3 years*	1,269	36	183	965,019	87.4% (85.6% - 89.1%)	97.2% (96.2% - 98.1%)
21. 6+ P, 3 years*	1,232	29	220	965,026	84.9% (82.9% - 86.7%)	97.7% (96.7% - 98.5%)
22. 1+ P, 1+ H, 3 years****	1,397	540	55	964,515	96.2% (95.1% - 97.1%)	72.1% (70.1% - 74.1%)
23. 2+ P, 1+ H, 3 years*	1,389	255	63	964,800	95.7% (94.5% - 96.7%)	84.5% (82.7% - 86.2%)
24. 3+ P, 1+ H, 3 years*	1,374	158	78	964,897	94.6% (93.3% - 95.7%)	89.7% (88.1% - 91.2%)
25. 1+ P, 2+ Rx, 3 years***	1,409	1,248	43	963,807	97.0% (96.0% - 97.9%)	53.0% (51.1% - 54.9%)
26. 2+ P, 2+ Rx, 3 years***	1,400	977	52	964,078	96.4% (95.3% - 97.3%)	58.9% (56.9% - 60.9%)
27. 3+ P, 2+ Rx, 3 years***	1,392	892	60	964,163	95.9% (94.7% - 96.8%)	61.0% (58.9% - 63.0%)
28. 1+ P, 1+ H, 2+ Rx, 3 years***	1,415	1,316	37	963,739	97.5% (96.5% - 98.2%)	51.8% (49.9% - 53.7%)

Algorithm	True Positive	False Positive	False Negative	True Negative	Sensitivity	PPV
29. 2+ P, 1+ H, 2+ Rx, 3 years***	1,408	1,050	44	964,005	97.0% (96.0% - 98.0%)	57.3% (55.3% - 59.3%)
30. 3+ P, 1+ H, 2+ Rx, 3 years***	1,403	965	49	964,090	96.6% (95.6% - 97.5%)	59.3% (57.2% - 61.2%)

* Specificity of 100.0% (100.0% - 100.0%). ** Specificity of 100.0% (99.9% - 100.0%). *** Specificity of 99.9% (99.9% - 99.9%). **** Specificity of 99.9% (99.9%-100.0%)
P = physician visit; H = hospitalization; Rx = prescription; PPV = positive predictive value.
Negative predictive value was 100.0% (100.0% - 100.0%) for all algorithms.

Table 4.14. Performance estimates (Youden's J Statistic, kappa, AUC) of algorithms identifying people living with HIV in administrative health data, Sensitivity Analysis 3

Algorithm	Youden's J	Kappa	AUC
1. 1+ P, 2 years	0.94 (0.93 - 0.95)	0.83 (0.82 - 0.85)	0.97 (0.96 - 0.98)
2. 2+ P, 2 years	0.93 (0.92 - 0.94)	0.91 (0.90 - 0.92)	0.96 (0.96 - 0.97)
3. 3+ P, 2 years	0.91 (0.89 - 0.92)	0.93 (0.92 - 0.94)	0.95 (0.95 - 0.96)
4. 4+ P, 2 years	0.89 (0.87 - 0.90)	0.93 (0.92 - 0.94)	0.94 (0.94 - 0.95)
5. 5+ P, 2 years	0.86 (0.84 - 0.88)	0.91 (0.90 - 0.92)	0.93 (0.92 - 0.94)
6. 6+ P, 2 years	0.83 (0.81 - 0.85)	0.90 (0.88 - 0.91)	0.91 (0.90 - 0.92)
7. 1+ P, 1+ H, 2 years	0.96 (0.95 - 0.97)	0.82 (0.81 - 0.84)	0.98 (0.98 - 0.99)
8. 2+ P, 1+ H, 2 years	0.95 (0.94 - 0.96)	0.90 (0.89 - 0.91)	0.98 (0.97 - 0.98)
9. 3+ P, 1+ H, 2 years	0.94 (0.93 - 0.95)	0.92 (0.91 - 0.93)	0.97 (0.97 - 0.98)
10. 1+ P, 2+ Rx, 2 years	0.97 (0.96 - 0.98)	0.69 (0.67 - 0.70)	0.98 (0.98 - 0.99)
11. 2+ P, 2+ Rx, 2 years	0.96 (0.95 - 0.97)	0.73 (0.72 - 0.75)	0.98 (0.98 - 0.99)
12. 3+ P, 2+ Rx, 2 years	0.96 (0.94 - 0.97)	0.74 (0.73 - 0.76)	0.98 (0.97 - 0.98)
13. 1+ P, 1+ H, 2+ Rx, 2 years	0.97 (0.97 - 0.98)	0.68 (0.66 - 0.69)	0.99 (0.98 - 0.99)
14. 2+ P, 1+ H, 2+ Rx, 2 years	0.97 (0.96 - 0.98)	0.72 (0.70 - 0.74)	0.98 (0.98 - 0.99)
15. 3+ P, 1+ H, 2+ Rx, 2 years	0.96 (0.95 - 0.97)	0.73 (0.72 - 0.75)	0.98 (0.98 - 0.99)
16. 1+ P, 3 years	0.94 (0.93 - 0.95)	0.83 (0.82 - 0.85)	0.97 (0.96 - 0.98)
17. 2+ P, 3 years	0.93 (0.92 - 0.94)	0.91 (0.90 - 0.92)	0.97 (0.96 - 0.97)
18. 3+ P, 3 years	0.91 (0.90 - 0.93)	0.93 (0.92 - 0.94)	0.96 (0.95 - 0.96)
19. 4+ P, 3 years	0.90 (0.88 - 0.91)	0.93 (0.92 - 0.94)	0.95 (0.94 - 0.96)
20. 5+ P, 3 years	0.87 (0.86 - 0.89)	0.92 (0.91 - 0.93)	0.94 (0.93 - 0.95)
21. 6+ P, 3 years	0.85 (0.83 - 0.87)	0.91 (0.90 - 0.92)	0.92 (0.92 - 0.93)
22. 1+ P, 1+ H, 3 years	0.96 (0.95 - 0.97)	0.82 (0.81 - 0.84)	0.98 (0.98 - 0.99)
23. 2+ P, 1+ H, 3 years	0.96 (0.95 - 0.97)	0.90 (0.89 - 0.91)	0.98 (0.97 - 0.98)
24. 3+ P, 1+ H, 3 years	0.95 (0.93 - 0.96)	0.92 (0.91 - 0.93)	0.97 (0.97 - 0.98)
25. 1+ P, 2+ Rx, 3 years	0.97 (0.96 - 0.98)	0.69 (0.67 - 0.7)	0.98 (0.98 - 0.99)
26. 2+ P, 2+ Rx, 3 years	0.96 (0.95 - 0.97)	0.73 (0.71 - 0.75)	0.98 (0.98 - 0.99)
27. 3+ P, 2+ Rx, 3 years	0.96 (0.95 - 0.97)	0.74 (0.73 - 0.76)	0.98 (0.97 - 0.98)
28. 1+ P, 1+ H, 2+ Rx, 3 years	0.97 (0.97 - 0.98)	0.68 (0.66 - 0.69)	0.99 (0.98 - 0.99)
29. 2+ P, 1+ H, 2+ Rx, 3 years	0.97 (0.96 - 0.98)	0.72 (0.70 - 0.74)	0.98 (0.98 - 0.99)
30. 3+ P, 1+ H, 2+ Rx, 3 years	0.97 (0.96 - 0.97)	0.73 (0.72 - 0.75)	0.98 (0.98 - 0.99)

P = physician visit; H = hospitalization; Rx = prescription; AUC = area under the receiver operating characteristic curve.

Table 4.15. Crude prevalence of HIV per 100,000 Manitobans ages 18+ years with continuous health insurance coverage between January 1, 2016 and December 31, 2018, by algorithm, Sensitivity Analysis 3

Algorithm	Number of HIV Cases	Prevalence (95% CI)
1. 1+ P, 2 years	1,493	187.1 (177.9 - 196.8)
2. 2+ P, 2 years	1,244	155.9 (147.5 - 164.8)
3. 3+ P, 2 years	1,146	143.6 (135.5 - 152.2)
4. 4+ P, 2 years	1,098	137.6 (129.7 - 146.0)
5. 5+ P, 2 years	1,061	133.0 (125.2 - 141.2)
6. 6+ P, 2 years	1,016	127.3 (119.7 - 135.4)
7. 1+ P, 1+ H, 2 years	1,540	193.0 (183.6 - 202.9)
8. 2+ P, 1+ H, 2 years	1,296	162.4 (153.8 - 171.5)
9. 3+ P, 1+ H, 2 years	1,203	150.8 (142.5 - 159.5)
10. 1+ P, 2+ Rx, 2 years	2,147	269.1 (257.9 - 280.7)
11. 2+ P, 2+ Rx, 2 years	1,918	240.4 (229.8 - 251.4)
12. 3+ P, 2+ Rx, 2 years	1,839	230.5 (220.2 - 241.2)
13. 1+ P, 1+ H, 2+ Rx, 2 years	2,190	274.5 (263.2 - 286.2)
14. 2+ P, 1+ H, 2+ Rx, 2 years	1,964	246.1 (235.5 - 257.3)
15. 3+ P, 1+ H, 2+ Rx, 2 years	1,887	236.5 (226.0 - 247.4)
16. 1+ P, 3 years	1,493	187.1 (177.9 - 196.8)
17. 2+ P, 3 years	1,246	156.2 (147.7 - 165.1)
18. 3+ P, 3 years	1,153	144.5 (136.4 - 153.1)
19. 4+ P, 3 years	1,111	139.2 (131.3 - 147.7)
20. 5+ P, 3 years	1,080	135.3 (127.5 - 143.7)
21. 6+ P, 3 years	1,046	131.1 (123.4 - 139.3)
22. 1+ P, 1+ H, 3 years	1,540	193.0 (183.6 - 202.9)
23. 2+ P, 1+ H, 3 years	1,298	162.7 (154.1 - 171.8)
24. 3+ P, 1+ H, 3 years	1,209	151.5 (143.2 - 160.3)
25. 1+ P, 2+ Rx, 3 years	2,152	269.7 (258.5 - 281.3)
26. 2+ P, 2+ Rx, 3 years	1,923	241.0 (230.5 - 252.0)
27. 3+ P, 2+ Rx, 3 years	1,849	231.7 (221.4 - 242.5)
28. 1+ P, 1+ H, 2+ Rx, 3 years	2,195	275.1 (263.8 - 286.8)
29. 2+ P, 1+ H, 2+ Rx, 3 years	1,969	246.8 (236.1 - 257.9)
30. 3+ P, 1+ H, 2+ Rx, 3 years	1,897	237.7 (227.3 - 248.7)

P = physician visit; H = hospitalization; Rx = prescription.

Note: The population denominator for all algorithms consisted of the 797,949 Manitobans who were at least 18 years old as of January 1, 2007 with continuous health insurance coverage between January 1, 2016 and December 31, 2018.

4.5.4 Sensitivity Analysis 4

To validate the algorithms in a cohort that contained only true HIV positive and true HIV negative individuals, this sensitivity analysis restricted the validation cohort to females who received a prenatal HIV screening test in the study observation period, and who were at least 15 years old at the time of the test. This excluded 1,375,809 individuals from the cohort (Figure 4.7). The validation cohort used for this sensitivity analysis consisted of a total of 78,201 individuals, 100 (0.1%) were HIV cases and 78,101 (99.9%) were HIV non-cases.

The characteristics of the validation cohort used in this sensitivity analysis were like the characteristics of the validation cohort used in the main analysis, with a slightly lower average age reflecting the average age of pregnant people in Manitoba. Compared to the main analysis, the algorithms in this sensitivity analysis had higher sensitivity, lower PPV, similar specificity and NPV, higher Youden's J Statistic, lower kappa, and higher AUC.

This sensitivity analysis was restricted to individuals with known HIV status, so we can more reliably use PPV to rule out algorithms which do not perform well. This sensitivity analysis can be used to remove 16 algorithms with PPV $\leq$ 49.8% from consideration: all those algorithms that required antiretroviral prescriptions and those that required only one physician visit (Algorithms 1, 7, 10-16, 22, 25-30). The remaining 14 algorithms have PPV $\geq$ 59.3%.

Characteristics to describe the validation cohort for this sensitivity analysis, stratified by HIV cases and HIV non-cases are provided in Table 4.16. The index date at which age, sex, area of residence, and area-level household income were measured for HIV cases and HIV non-cases was defined as the start of insurance coverage (if insurance coverage started after January 1, 2007) or January 1, 2007. The number of people who could not be assigned to an income quintile (e.g. Public Trustees) was small, so they were included with Q1. The number of people assigned to income quintile Q5 was also small, so they were included with Q4.

The average HIV case was 23.9 years old as of the index date and 29.4 years old at the time of their earliest positive HIV prenatal screen during the study observation period. Most HIV cases lived in the WRHA. Approximately one-half of HIV cases were assigned to the lowest income quintile, Q1, with a decreasing number of HIV cases as the quintiles increased.

The average HIV non-case was 23.2 years old as of the index date and 28.1 years old at the time of their earliest negative HIV prenatal screen during the study observation period. More than

one-half of HIV non-cases lived in WRHA. The distribution of HIV non-cases was slightly skewed toward the lower income quintiles.

The sensitivity, specificity, and PPV for all algorithms are given in Table 4.17. Algorithm sensitivity ranged from 90.0% (95% CI 82.4% - 95.1%) for Algorithm 6 to 100.0% (95% CI 96.4% - 100.0%) for multiple algorithms. The PPV of the algorithms ranged from 26.1% (95% CI 21.8% - 30.8%) for Algorithm 28 to 77.5% (95% CI 69.0% - 84.6%) for Algorithm 20. All algorithms which required two or more antiretroviral prescriptions and those that required only one or more physician visit (Algorithms 1, 7, 10-16, 22, 25-30) had PPV ≤ 49.8%. Specificity and NPV were very high for all algorithms. Specificity ranged from 99.6% (95% CI 99.6% - 99.7%) to 100.0% (95% CI 100.0% - 100.0%) for all algorithms. Since many algorithms had the same specificity, specificity is represented by symbols in Table 4.17. NPV was 100.0% (95% CI 100.0%-100.0%) for all algorithms and so was not included in Table 4.17.

The Youden's J Statistic, kappa, and AUC for all algorithms are given in Table 4.18. Youden's J Statistic ranged from 0.90 (95% CI 0.84 - 0.96) to 1.00 (95% CI 1.00 - 1.00) across all algorithms. The kappa statistic ranged from 0.41 (95% CI 0.36 - 0.47) to 0.85 (95% CI 0.79 - 0.90), indicating moderate to very good, agreement between all algorithms and the reference standard. The AUC ranged from 0.95 (95% CI 0.92 - 0.98) to 1.00 (95% CI 1.00 - 1.00) indicating outstanding discrimination between HIV cases and HIV non-cases for all algorithms.

Table 4.19 shows the algorithm derived prevalence of HIV among Manitobans females who received a prenatal HIV screening test at 15+ years of age in the study observation period and who had continuous health insurance coverage between January 1, 2016 and December 31, 2018. The prevalence calculated using the algorithms ranged from 145.5 females diagnosed with HIV per 100,000 population (95% CI 120.4 – 175.9) for Algorithm 6 to 481.4 females diagnosed with HIV per 100,000 population (95% CI 433.7 – 534.2) for Algorithm 28. Prevalence calculated using algorithms that either required antiretroviral prescription data or only required one or more physician visits was much greater than the prevalence calculated using the other 14 algorithms. Excluding those algorithms that either required antiretroviral prescription data or only required one or more physician visits, the maximum prevalence was 209.4 females diagnosed with HIV per 100,000 population (95% CI 178.8– 245.2) for Algorithms 8 and 23.

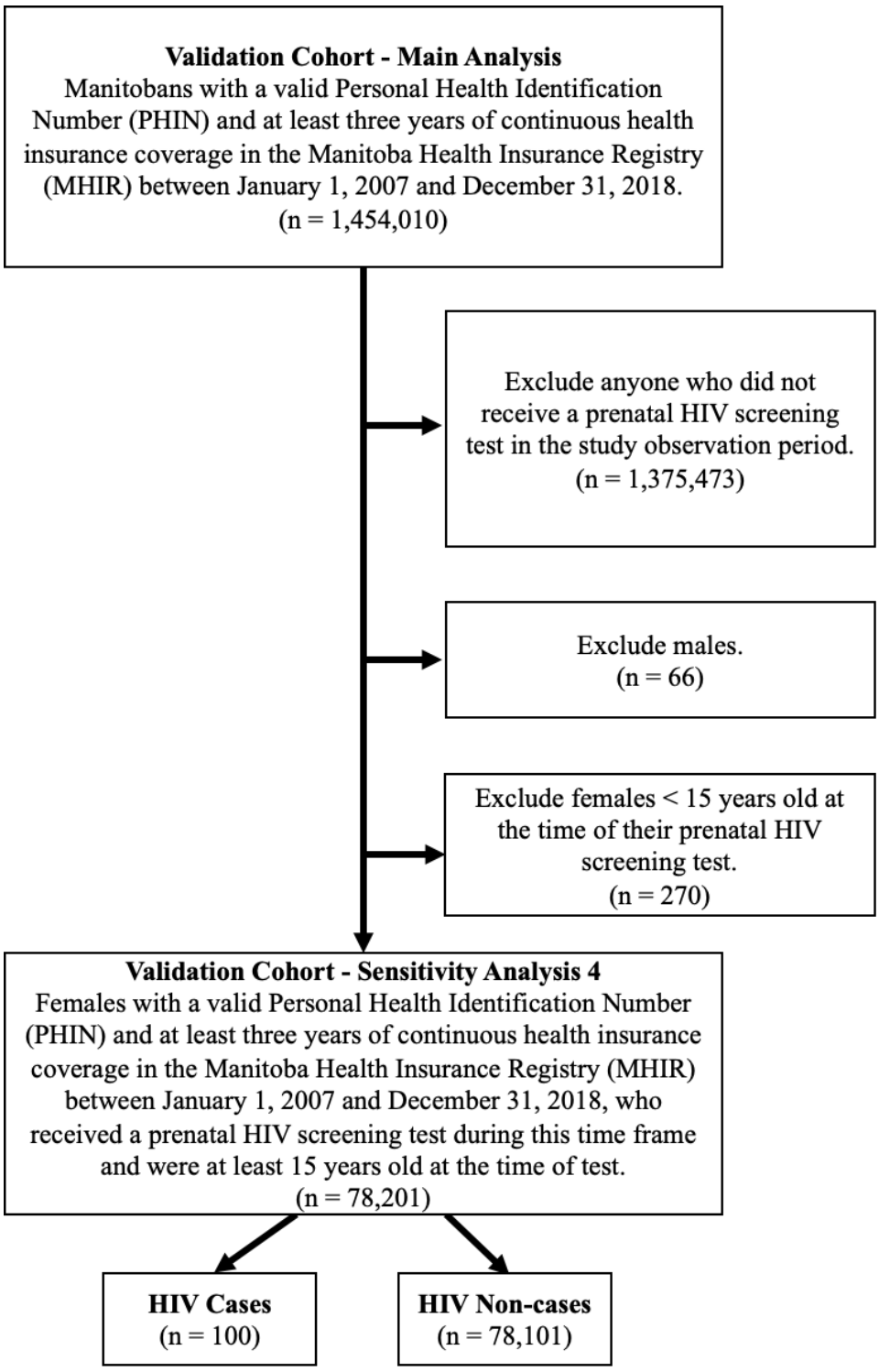

Figure 4.7. Construction of the validation cohort, Sensitivity Analysis 4

Table 4.16. Description of the validation cohort, Sensitivity Analysis 4

Characteristic	HIV Case (N = 100)	HIV Non-Case (N = 78,101)
Age at index date, years	23.9 ± 7.2	23.2 ± 7.1
[minimum, maximum]	[11.2, 46.5]	[3.3, 82.4]
Area of residence (RHA)		
IERHA	7 (7)	6,587 (8.4)
NHR	8 (8)	8,422 (10.8)
SH-SS	s	s
PMH	s	s
WRHA	77 (77)	40,885 (52.3)
Public Trustee/Not Found	s	s
Income quintile		
Q1-Lowest/Not Found	55 (55)	22,560 (28.9)
Q2	22 (22)	17,168 (22)
Q3	10 (10)	14,299 (18.3)
Q4/Q5-Highest	13 (13)	24,074 (30.8)
Health insurance coverage, years	10.5 ± 2.4	10.9 ± 2.3
Age at prenatal HIV screening test*, years	29.4 ± 6.6	28.1 ± 6.2
[minimum, maximum]	[16.3, 53.0]	[15.0, 85.2]
Year of prenatal HIV screening test*		
2007	0 (0)	4,247 (5.4)
2008	s	s
2009	6 (6)	7,776 (10)
2010	13 (13)	6,817 (8.7)
2011	13 (13)	6,494 (8.3)
2012	13 (13)	7,025 (9)
2013	6 (6)	7,936 (10.2)
2014	13 (13)	7,670 (9.8)
2015	11 (11)	7,324 (9.4)
2016	13 (13)	6,332 (8.1)
2017	s	s
2018	7 (7)	4,004 (5.1)

Notes:
Data are mean ± standard deviation or n (%)
*the earliest positive HIV prenatal screen (HIV cases) or earliest negative HIV prenatal screen (HIV non-cases).
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table 4.17. Performance estimates (sensitivity, specificity, PPV) of algorithms identifying people living with HIV in administrative health data, Sensitivity Analysis 4

Algorithm	True Positive	False Positive	False Negative	True Negative	Sensitivity	Specificity	PPV
1. 1+ P, 2 years	100	102	0	78,269	100.0% (96.4% - 100.0%)	99.9% (99.8% - 99.9%)	49.5% (42.4% - 56.6%)
2. 2+ P, 2 years	99	s	s	78,312	99.0% (94.6% - 100.0%)	99.9% (99.9% - 99.9%)	62.7% (54.6% - 70.2%)
3. 3+ P, 2 years	99	s	s	78,334	99.0% (94.6% - 100.0%)	100.0% (99.9% - 100.0%)	72.8% (64.5% - 80.1%)
4. 4+ P, 2 years	95	s	s	78,340	95.0% (88.7% - 98.4%)	100.0% (99.9% - 100.0%)	75.4% (66.9% - 82.6%)
5. 5+ P, 2 years	92	27	8	78,344	92.0% (84.8% - 96.5%)	100.0% (100.0% - 100.0%)	77.3% (68.7% - 84.5%)
6. 6+ P, 2 years	90	27	10	78,344	90.0% (82.4% - 95.1%)	100.0% (100.0% - 100.0%)	76.9% (68.2% - 84.2%)
7. 1+ P, 1+ H, 2 years	100	110	0	78,261	100.0% (96.4% - 100.0%)	99.9% (99.8% - 99.9%)	47.6% (40.7% - 54.6%)
8. 2+ P, 1+ H, 2 years	99	s	s	78,303	99.0% (94.6% - 100.0%)	99.9% (99.9% - 99.9%)	59.3% (51.4% - 66.8%)
9. 3+ P, 1+ H, 2 years	99	s	s	78,324	99.0% (94.6% - 100.0%)	99.9% (99.9% - 100.0%)	67.8% (59.6% - 75.3%)
10. 1+ P, 2+ Rx, 2 years	100	272	0	78,099	100.0% (96.4% - 100.0%)	99.7% (99.6% - 99.7%)	26.9% (22.4% - 31.7%)
11. 2+ P, 2+ Rx, 2 years	s	237	s	78,134	99.0% (94.6% - 100.0%)	99.7% (99.7% - 99.7%)	29.5% (24.6% - 34.7%)
12. 3+ P, 2+ Rx, 2 years	s	219	s	78,152	99.0% (94.6% - 100.0%)	99.7% (99.7% - 99.8%)	31.1% (26.1% - 36.5%)
13. 1+ P, 1+ H, 2+ Rx, 2 years	100	280	0	78,091	100.0% (96.4% - 100.0%)	99.6% (99.6% - 99.7%)	26.3% (22.0% - 31.1%)
14. 2+ P, 1+ H, 2+ Rx, 2 years	s	246	s	78,125	99.0% (94.6% - 100.0%)	99.7% (99.6% - 99.7%)	28.7% (24.0% - 33.8%)
15. 3+ P, 1+ H, 2+ Rx, 2 years	s	228	s	78,143	99.0% (94.6% - 100.0%)	99.7% (99.7% - 99.8%)	30.3% (25.3% - 35.6%)

Algorithm	True Positive	False Positive	False Negative	True Negative	Sensitivity	Specificity	PPV
16. 1+ P, 3 years	100	102	0	78,269	100.0% (96.4% - 100.0%)	99.9% (99.8% - 99.9%)	49.5% (42.4% - 56.6%)
17. 2+ P, 3 years	99	s	s	78,312	99.0% (94.6% - 100.0%)	99.9% (99.9% - 99.9%)	62.7% (54.6% - 70.2%)
18. 3+ P, 3 years	99	s	s	78,334	99.0% (94.6% - 100.0%)	100.0% (99.9% - 100.0%)	72.8% (64.5% - 80.1%)
19. 4+ P, 3 years	97	s	s	78,338	97.0% (91.5% - 99.4%)	100.0% (99.9% - 100.0%)	74.6% (66.2% - 81.8%)
20. 5+ P, 3 years	93	27	7	78,344	93.0% (86.1% - 97.1%)	100.0% (100.0% - 100.0%)	77.5% (69.0% - 84.6%)
21. 6+ P, 3 years	91	27	9	78,344	91.0% (83.6% - 95.8%)	100.0% (100.0% - 100.0%)	77.1% (68.5% - 84.4%)
22. 1+ P, 1+ H, 3 years	100	110	0	78,261	100.0% (96.4% - 100.0%)	99.9% (99.8% - 99.9%)	47.6% (40.7% - 54.6%)
23. 2+ P, 1+ H, 3 years	99	s	s	78,303	99.0% (94.6% - 100.0%)	99.9% (99.9% - 99.9%)	59.3% (51.4% - 66.8%)
24. 3+ P, 1+ H, 3 years	99	s	s	78,324	99.0% (94.6% - 100.0%)	99.9% (99.9% - 100.0%)	67.8% (59.6% - 75.3%)
25. 1+ P, 2+ Rx, 3 years	100	275	0	78,096	100.0% (96.4% - 100.0%)	99.7% (99.6% - 99.7%)	26.7% (22.3% - 31.5%)
26. 2+ P, 2+ Rx, 3 years	s	240	s	78,131	99.0% (94.6% - 100.0%)	99.7% (99.7% - 99.7%)	29.2% (24.4% - 34.4%)
27. 3+ P, 2+ Rx, 3 years	s	222	s	78,149	99.0% (94.6% - 100.0%)	99.7% (99.7% - 99.8%)	30.8% (25.8% - 36.2%)
28. 1+ P, 1+ H, 2+ Rx, 3 years	100	283	0	78,088	100.0% (96.4% - 100.0%)	99.6% (99.6% - 99.7%)	26.1% (21.8% - 30.8%)
29. 2+ P, 1+ H, 2+ Rx, 3 years	s	249	s	78,122	99.0% (94.6% - 100.0%)	99.7% (99.6% - 99.7%)	28.5% (23.8% - 33.5%)
30. 3+ P, 1+ H, 2+ Rx, 3 years	s	231	s	78,140	99.0% (94.6% - 100.0%)	99.7% (99.7% - 99.7%)	30.0% (25.1% - 35.3%)

P = physician visit; H = hospitalization; Rx = prescription; PPV = positive predictive value.
Negative predictive value was 100.0% (100.0% - 100.0%) for all algorithms.

Table 4.18. Performance estimates (Youden's J Statistic, kappa, AUC) of algorithms identifying people living with HIV in administrative health data, Sensitivity Analysis 4

Algorithm	Youden's J	Kappa	AUC
1. 1+ P, 2 years	1.00 (1.00 - 1.00)	0.66 (0.60 - 0.73)	1.00 (1.00 - 1.00)
2. 2+ P, 3 years	0.99 (0.97 - 1.00)	0.77 (0.71 - 0.83)	0.99 (0.98 - 1.00)
3. 3+ P, 3 years	0.99 (0.97 - 1.00)	0.84 (0.79 - 0.89)	0.99 (0.98 - 1.00)
4. 4+ P, 4 years	0.95 (0.91 - 0.99)	0.84 (0.79 - 0.89)	0.97 (0.95 - 1.00)
5. 5+ P, 5 years	0.92 (0.87 - 0.97)	0.84 (0.79 - 0.89)	0.96 (0.93 - 0.99)
6. 6+ P, 6 years	0.90 (0.84 - 0.96)	0.83 (0.78 - 0.88)	0.95 (0.92 - 0.98)
7. 1+ P, 1+ H, 7 years	1.00 (1.00 - 1.00)	0.65 (0.58 - 0.71)	1.00 (1.00 - 1.00)
8. 2+ P, 1+ H, 8 years	0.99 (0.97 - 1.00)	0.74 (0.69 - 0.80)	0.99 (0.98 - 1.00)
9. 3+ P, 1+ H, 9 years	0.99 (0.97 - 1.00)	0.80 (0.75 - 0.86)	0.99 (0.98 - 1.00)
10. 1+ P, 2+ Rx, 2 years	1.00 (1.00 - 1.00)	0.42 (0.37 - 0.48)	1.00 (1.00 - 1.00)
11. 2+ P, 2+ Rx, 2 years	0.99 (0.97 - 1.00)	0.46 (0.4 0- 0.51)	0.99 (0.98 - 1.00)
12. 3+ P, 2+ Rx, 2 years	0.99 (0.97 - 1.00)	0.47 (0.42 - 0.53)	0.99 (0.98 - 1.00)
13. 1+ P, 1+ H, 2+ Rx, 2 years	1.00 (1.00 - 1.00)	0.42 (0.36 - 0.47)	1.00 (1.00 - 1.00)
14. 2+ P, 1+ H, 2+ Rx, 2 years	0.99 (0.97 - 1.00)	0.45 (0.39 - 0.50)	0.99 (0.98 - 1.00)
15. 3+ P, 1+ H, 2+ Rx, 2 years	0.99 (0.97 - 1.00)	0.46 (0.41 - 0.52)	0.99 (0.98 - 1.00)
16. 1+ P, 3 years	1.00 (1.00 - 1.00)	0.66 (0.60 - 0.73)	1.00 (1.00 - 1.00)
17. 2+ P, 3 years	0.99 (0.97 - 1.00)	0.77 (0.71 - 0.83)	0.99 (0.98 - 1.00)
18. 3+ P, 3 years	0.99 (0.97 - 1.00)	0.84 (0.79 - 0.89)	0.99 (0.98 - 1.00)
19. 4+ P, 3 years	0.97 (0.94 - 1.00)	0.84 (0.79 - 0.89)	0.98 (0.97 - 1.00)
20. 5+ P, 3 years	0.93 (0.88 - 0.98)	0.85 (0.79 - 0.90)	0.96 (0.94 - 0.99)
21. 6+ P, 3 years	0.91 (0.85 - 0.97)	0.83 (0.78 - 0.89)	0.95 (0.93 - 0.98)
22. 1+ P, 1+ H, 3 years	1.00 (1.00 - 1.00)	0.65 (0.58 - 0.71)	1.00 (1.00 - 1.00)
23. 2+ P, 1+ H, 3 years	0.99 (0.97 - 1.00)	0.74 (0.69 - 0.80)	0.99 (0.98 - 1.00)
24. 3+ P, 1+ H, 3 years	0.99 (0.97 - 1.00)	0.80 (0.75 - 0.86)	0.99 (0.98 - 1.00)
25. 1+ P, 2+ Rx, 3 years	1.00 (1.00 - 1.00)	0.42 (0.37 - 0.48)	1.00 (1.00 - 1.00)
26. 2+ P, 2+ Rx, 3 years	0.99 (0.97 - 1.00)	0.45 (0.39 - 0.51)	0.99 (0.98 - 1.00)
27. 3+ P, 2+ Rx, 3 years	0.99 (0.97 - 1.00)	0.47 (0.41 - 0.53)	0.99 (0.98 - 1.00)
28. 1+ P, 1+ H, 2+ Rx, 3 years	1.00 (1.00 - 1.00)	0.41 (0.36 - 0.47)	1.00 (1.00 - 1.00)
29. 2+ P, 1+ H, 2+ Rx, 3 years	0.99 (0.97 - 1.00)	0.44 (0.39 - 0.50)	0.99 (0.98 - 1.00)
30. 3+ P, 1+ H, 2+ Rx, 3 years	0.99 (0.97 - 1.00)	0.46 (0.40 - 0.52)	0.99 (0.98 - 1.00)

P = physician visit; H = hospitalization; Rx = prescription; AUC = area under the receiver operating characteristic curve.

Table 4.19. Crude prevalence of HIV per 100,000 Manitoba females who received a prenatal HIV screening test at age 15+ years in the study observation period and had continuous health insurance coverage between January 1, 2016 and December 31, 2018, by algorithm, Sensitivity Analysis 4

Algorithm	Number of HIV Cases	Prevalence (95% CI)
1. 1+ P, 2 years	186	252.9 (219.1 - 292)
2. 2+ P, 2 years	145	197.2 (167.6 - 232)
3. 3+ P, 2 years	124	168.6 (141.4 - 201.1)
4. 4+ P, 2 years	115	156.4 (130.3 - 187.7)
5. 5+ P, 2 years	108	146.9 (121.6 - 177.3)
6. 6+ P, 2 years	107	145.5 (120.4 - 175.9)
7. 1+ P, 1+ H, 2 years	194	263.8 (229.2 - 303.7)
8. 2+ P, 1+ H, 2 years	154	209.4 (178.8 - 245.2)
9. 3+ P, 1+ H, 2 years	134	182.2 (153.8 - 215.8)
10. 1+ P, 2+ Rx, 2 years	343	466.4 (419.6 - 518.5)
11. 2+ P, 2+ Rx, 2 years	309	420.2 (375.8 - 469.7)
12. 3+ P, 2+ Rx, 2 years	291	395.7 (352.8 - 443.9)
13. 1+ P, 1+ H, 2+ Rx, 2 years	351	477.3 (429.9 - 529.9)
14. 2+ P, 1+ H, 2+ Rx, 2 years	318	432.4 (387.4 - 482.7)
15. 3+ P, 1+ H, 2+ Rx, 2 years	300	407.9 (364.3 - 456.8)
16. 1+ P, 3 years	186	252.9 (219.1 - 292)
17. 2+ P, 3 years	145	197.2 (167.6 - 232)
18. 3+ P, 3 years	124	168.6 (141.4 - 201.1)
19. 4+ P, 3 years	118	160.5 (134 - 192.2)
20. 5+ P, 3 years	109	148.2 (122.8 - 178.8)
21. 6+ P, 3 years	108	146.9 (121.6 - 177.3)
22. 1+ P, 1+ H, 3 years	194	263.8 (229.2 - 303.7)
23. 2+ P, 1+ H, 3 years	154	209.4 (178.8 - 245.2)
24. 3+ P, 1+ H, 3 years	134	182.2 (153.8 - 215.8)
25. 1+ P, 2+ Rx, 3 years	346	470.5 (423.4 - 522.8)
26. 2+ P, 2+ Rx, 3 years	312	424.3 (379.7 - 474)
27. 3+ P, 2+ Rx, 3 years	294	399.8 (356.6 - 448.2)
28. 1+ P, 1+ H, 2+ Rx, 3 years	354	481.4 (433.7 - 534.2)
29. 2+ P, 1+ H, 2+ Rx, 3 years	321	436.5 (391.3 - 487)
30. 3+ P, 1+ H, 2+ Rx, 3 years	303	412 (368.1 - 461.1)

P = physician visit; H = hospitalization; Rx = prescription.

Note: The population denominator for all algorithms consisted of the 73,540 Manitoban females who received a prenatal HIV screening test at age 15+ years in the study observation period and had continuous health insurance coverage between January 1, 2016 and December 31, 2018.

Chapter 5 - Discussion

5.1 Overview

This chapter summarizes the results and implications of the study beginning with a discussion of the key findings in the context of existing research (Section 5.2) and a comparison of this study to the other two Canadian HIV algorithm studies (Section 5.3). The strengths and limitations of the validation study are then outlined (Section 5.4) and the implications of the research, including possible future research are discussed (Section 5.5).

5.2 Discussion of Results

This study demonstrates that administrative data in Manitoba can be used to build a validation cohort that is reflective of HIV cases and HIV non-cases in the Manitoba population. In addition, algorithms validated using this data can be used to accurately identify Manitobans living with HIV who interact with the healthcare system, with different algorithms performing best under different scenarios.

The number of people living with HIV (n = 1,589, 0.1%) in the validation cohort aligns with the 0.1% (i.e., 92.6 per 100,000 population) prevalence of HIV in the Manitoba population, calculated as of December 31, 2017. (30) The age, sex, and geographic region of residence of HIV cases in the validation cohort aligns with the demographic characteristics of HIV cases in Manitoba produced using the same (30), and different (107), data sources. The distribution of HIV cases by income quintile aligns with previously published results using the CPL LIMS data. (30) The identification of HIV cases increases for a few years following the start of the observation period as prevalent or existing cases from prior to 2007 were identified for the first time, then declined and remained fairly steady, following the trend of HIV incidence in Manitoba. (84)

The distribution of sex and geographic region of residence of HIV non-cases in the validation cohort closely aligned with the demographics of Manitoba's overall population in 2018. (108) The average age of HIV non-cases in the validation cohort was 6.5 years lower than the average age of 39.2, identified by the 2016 Statistics Canada Census as the average age of Manitobans. (109)

The ICD-9-CM code 042, which was the only ICD-9-CM code that contributed to algorithm development when searching the Hospital Abstracts database, is simply a diagnosis of

HIV. The ICD-9-CM codes that did not contribute to algorithm development when searching the Hospital Abstracts database were codes for identifying secondary conditions resulting from HIV infection (043, 044, 795.71), asymptomatic HIV infection (V08), and HIV, Type 2 (079.53). ICD-10-CA codes B24, O98.7, R75, and Z21 were the ICD-10-CA codes that contributed to algorithm development when searching the Hospital Abstracts database. These codes diagnose unspecified HIV infection; HIV disease complicating pregnancy, childbirth, and the puerperium; laboratory evidence of HIV; and asymptomatic HIV infection, respectively. The ICD-10-CA codes that did not contribute to algorithm development when searching the Hospital Abstracts database (B20, B21, B22, B23, F02.4) were for secondary conditions resulting from HIV infection.

Of the 49 DINs that did not contribute to algorithm development when searching the DPIN database, 23 were not found in the DPIN database because the medications were either cancelled prior to the start of the study observation period or introduced after the end of the study observation period. An additional two DINs could not be found in Health Canada's Drug Product Database, (110) suggesting the DINs were not valid. It is not clear why the remaining 24 DINs did not contribute to algorithm development when searching the DPIN database. All DINs used in this study were selected because they were used in prior research. (9, 97)

The 12 algorithms which required two or more HIV antiretroviral prescriptions (Algorithms 10-15 and 25-30) produced the greatest number of false positives of all the algorithms that were investigated, ranging from 46.8% to 55.9% of the positive HIV cases. This is likely because the antiretroviral prescription list was not specific enough and included medications also used for pre-exposure prophylaxis, post exposure prophylaxis, or the treatment of other viral infections such as Hepatitis B. The accuracy of the algorithms that required HIV antiretroviral prescriptions could potentially be improved by limiting the DIN list to those medications used exclusively to treat HIV infection. However, given the results of this study, future validation studies should not be concerned if they do not have prescription data available for HIV algorithm validation.

To select the best algorithm for identifying HIV cases in a particular research study, one should consider the number of HIV cases identified by the algorithm, the seven measures of accuracy, and the prevalence of HIV calculated using the algorithm. There is no single algorithm

that can be identified as the best overall; the algorithm selected should depend on the context of the study. Table 5.1 summarizes the best algorithm to select based on the goal of the study.

Regarding the number of HIV cases identified by the algorithm: Algorithm 18, which required three or more physician visits in three years, was closest to identifying the same number of HIV cases as the reference standard, with only 13 fewer cases identified. Only 9% of the HIV cases identified by Algorithm 18 were false positives, thus, Algorithm 18 produced a sample that was representative of the reference standard, even if they were not the entirety of HIV cases from the reference standard.

Regarding the seven measures of accuracy: for a condition like HIV, where the prevalence of infection within the target population is very low, specificity and NPV are not useful in differentiating the accuracy of the algorithms because the numbers of true negatives (which, for the purposes of this study were those individuals in the validation cohort who did not have a confirmatory HIV laboratory test within the study observation period, rather than people who have tested negative for HIV) is very large relative to the number of false positives (for specificity) or false negatives (for NPV) so the values of specificity and NPV are very close to 100% for all algorithms. By extension, Youden's J Statistic (J = sensitivity + specificity – 1) becomes a reflection of the sensitivity value and does not add to the evaluation of the algorithm. (99) The AUC (which plots sensitivity against 1-specificity and takes the area under the curve) is also only influenced by the changing sensitivity value. (99) Thus, in this context, sensitivity, PPV, and kappa are the best measures of accuracy to consider when choosing an algorithm.

The algorithm with the highest sensitivity will identify as many true positive HIV cases as possible. To maximize sensitivity, the number of true positive HIV cases identified by the algorithm must be maximized, and in turn, the number of false negatives must be minimized. An unfortunate side effect of maximizing the number of true positives is that the number of false positives also tends to increase. For example, Algorithms 13 and 28 both require one or more physician visits, one or more hospitalizations, or two or more antiretroviral prescriptions in two- or three-years, respectively, and both have the highest sensitivity at 96.5% (95% CI 95.5% - 97.4%). These algorithms identify the largest number of true positive HIV cases (n = 1,534, respectively) with minimal false negatives (n = 55, respectively), but they also identify many false positive HIV cases, accounting for 55.8% and 55.9% of the total HIV cases identified (n = 1,943

and n = 1943, respectively). Due to the large number of false positives, the average age of HIV cases identified by Algorithm 13 as of the index date was the same as HIV non-cases in the validation cohort (32.7 years), rather than the 36.5 years for HIV cases in the validation cohort. The percentage of HIV cases who were male was also lower than the percentage of males in the reference standard with 55.6% males compared to 66.3% males. The majority of HIV cases identified by the algorithm resided in WRHA (76.1%) and were within the lowest income quintile (43.7%), similar to HIV cases in the validation cohort. Since Algorithm 13 requires less data than Algorithm 28 but produces the same sensitivity, Algorithm 13 would be the best choice when the goal is to identify all possible HIV cases, without concern for the number of false positives captured. An example of a study with this goal would be the development of a clinical cohort of HIV patients that requires the review of patient charts. Using administrative data to identify potential individuals for inclusion in the study will reduce the total number of client charts that need to be reviewed.

While PPV is also affected by the prevalence of the infection, with PPV decreasing as prevalence decreases, the number of false positives relative to the number of true positives is not as large, resulting in a more moderate effect on PPV. The algorithm with highest PPV has the highest proportion of true positive HIV cases out of all HIV cases identified. To maximize PPV, the number of true positive HIV cases identified by the algorithm must be maximized, and in turn, the number of false positives must be minimized. Algorithm 6 which requires six or more physician visits in two years had the highest PPV at 96.0% (95% CI 94.8% - 97.0%). This algorithm identified 1,289 true positive HIV cases and 54 false positive HIV cases. Increased PPV results in an increased number of true HIV cases falsely classified as HIV non-cases (300 false negatives for Algorithm 6). But, given the low number of false positives, Algorithm 6 produced a sample that was representative of the reference standard. The average age as of the index date was 36.9 years for HIV cases identified by the algorithm compared to 36.5 years for HIV cases identified by the reference standard. 65.7% of HIV cases identified by the algorithm were males and 81.2% were from Winnipeg, compared to 66.3% and 81.6% of HIV cases identified by the reference standard, respectively. Approximately one-half (52.0%) of HIV cases identified by the algorithm were from Q1, similar to the 51.6% of HIV cases identified by the reference standard. Algorithm 6 would be the best choice if the goal was to identify as many true positive HIV cases as possible with minimal

HIV non-cases falsely classified as HIV cases. An example of a study with this goal would be a descriptive study of the characteristics of people living with HIV with minimal bias from people without HIV being captured as false positives.

Kappa does not rely on sensitivity, specificity or prevalence and takes into consideration the agreement between the two sources that may be generated by random chance. The algorithms with highest kappa have the best balance between the number of HIV non-cases misclassified as HIV cases and the number of HIV cases misclassified as HIV non-cases. Algorithms 4, 18, and 19 all have kappa of 0.91 (95% CI 0.90 - 0.92). Algorithm 18 requires three or more physician visits in three years and Algorithms 4 and 19 require four or more physician visits in two- or three-years, respectively. Algorithm 3, which requires three or more physician visits in two years has kappa of 0.90 (95% CI 0.89 - 0.92). Algorithm 3 identified 1,427 true positives, 139 false positives, and 162 false negatives. Since false positives accounted for a small percentage (8.9%) of total HIV cases identified, Algorithm 3 also produced a sample that was representative of the reference standard. The average age as of the index date was 35.8 years for HIV cases identified by Algorithm 3, 64.7% of HIV cases identified by the algorithm were males and 81.5% were from Winnipeg. Approximately one-half (51.3%) of HIV cases identified by the algorithm were from the lowest income quintile. Algorithms 3, 4, 18, or 19 would be good choices if the goal was to compare HIV cases to HIV non-cases and therefore need to accurately distinguish the two groups. Algorithm 3, while it has slightly lower kappa than Algorithms 4, 18, and 19, would be the best choice as it requires the least amount of data over the smallest time frame. An example of a study with the goal of comparing HIV cases to HIV non-cases would be a retrospective study to determine whether HIV cases differ from HIV non-cases with respect to a health outcome, such as, cancer survival.

Regarding the prevalence of HIV calculated using the algorithm: Algorithm 6, which required six or more physician visits in three years, was closest to estimating the true prevalence of HIV in the Manitoba population with an estimated prevalence of 92.9 cases per 100,000 population (95% CI 87.6 – 98.4). Algorithm 6 also had the lowest percent of false positive cases with only 4% of the 1,343 HIV case identified by the algorithm classified as false positives so the sample of HIV cases identified was representative of the reference standard with an average age as of the index date of 36.9 years, 65.7% of cases were male, 81.2% resided in WRHA, and 52.0% were captured within the lowest income quintile.

Table 5.1. Best algorithms for different study scenarios

Goal	Action	Algorithm Selection
Produce a sample of HIV cases that has the same number of cases as the reference standard (or as close as possible).	Chose the algorithm with smallest difference in number of HIV cases.	Algorithm 18: three or more physician visits in three years
Produce a sample of all possible HIV cases, without concern for the number of false positives captured.	Chose the algorithm with highest sensitivity.	Algorithm 13: one or more physician visits or one or more hospitalizations or two or more ART prescriptions in two years
Produce a sample of HIV cases with as many true positives as possible with minimal false positives.	Chose the algorithm with highest PPV.	Algorithm 6: six or more physician visits in two years
Produce a sample of HIV cases that is most representative of the reference standard.	Choose the algorithm with fewest false positives.	Algorithm 6: six or more physician visits in two years
Produce a sample of HIV cases and HIV non-cases that minimizes the misclassification of both HIV cases and HIV non-cases and allows for most accurate comparison between the two groups.	Choose the algorithm with highest kappa.	Algorithm 3: three or more physician visits in two years
Produce a sample of HIV cases and HIV non-cases that has the same prevalence of HIV as the Manitoba population (or as close as possible).	Choose the algorithm with prevalence closest to that of the Manitoba population.	Algorithm 6: six or more physician visits in two years
Overall best algorithm	Highest possible PPV (≥90%) while maximizing sensitivity (≥90%) over the shortest duration examined (in years).	Algorithm 3: three or more physician visits in two years

When the algorithm selected as best under a specific study scenario in Table 5.1 was chosen due to the value of a performance estimate, it is important to consider the precision of the estimate.

When the 95% confidence interval for the estimate of the selected algorithm overlaps with the 95% confidence interval of another algorithm, there is little difference in the validity of the algorithms, meaning the other algorithm would perform similarly to achieve the stated goal.

The 95% confidence interval for the sensitivity estimate of Algorithm 13 overlapped with the 95% confidence intervals for all algorithms that required two or more ART prescriptions and all algorithms that required one or more physician visits or one or more hospitalization. Any of these algorithms could be used to reliably generate a sample of all possible HIV cases without regard for false positives.

The 95% confidence interval for the PPV estimate of Algorithm 6 overlapped with the 95% confidence interval for the algorithm that required six or more physician visits in three years and all algorithms that required four or more physician visits or five or more physician visits. Any of these algorithms could be used to reliably generate a sample of HIV cases with as many true positives as possible with minimal false positives.

5.2.1 Sensitivity Analyses

The first sensitivity analysis applied the 30 algorithms using only ICD-9-CM codes 042-044 and ICD-10-CA codes B20-B24 to produce algorithm estimates that were comparable to the results produced by Antoniou et al. (9) in Ontario, Canada. The addition of ICD-9-CM codes V08, 795.51 and ICD-10-CA codes R75 and Z71 codes in the main analysis slightly increased sensitivity and slightly decreased PPV by allowing for the identification of slightly more HIV cases per algorithm but also a slightly higher proportion of false positives. The difference between the results in the main analysis and the first sensitivity analysis were so minimal that the interpretation of the results for the sensitivity analysis does not differ from the main analysis. This indicates the addition of ICD-9-CM codes V08, 795.51 and ICD-10-CA codes R75 and Z71 codes in the main analysis was not necessary.

The second sensitivity analysis applied the 30 algorithms using only the ICD-9-CM codes 042-044, V08, 795.71 and ICD-10-CA codes B20-B24, R75, Z21 to produce algorithm estimates that were comparable to the results produced by Nosyk et al. (10) in B.C., Canada. There was no difference between the results in the main analysis and the results in the second sensitivity analysis, so the addition of ICD-9-CM codes 079.53, F02.4 and O98.7 was not necessary.

The third sensitivity analysis applied the 30 algorithms to a reference standard consisting of adults only. All performance estimates for the third sensitivity analysis had the same, or higher, values when compared with the main analysis, except for specificity. Specificity was marginally lower than the main analysis, ranging from 99.6% (95% CI 99.6% - 99.7%) to 100.0% (95% CI 99.9% - 100.0%), due to the lower number of individuals in the validation cohort (966,507 people for the third sensitivity analysis compared to 1,454,010 people in the main analysis). The results of the third sensitivity analysis suggest that the algorithms performed better when applied to adults only; very few cases of HIV are diagnosed in people under 18 years of age, so by removing their records from the analysis, we reduced misclassification.

The fourth sensitivity analysis applied the 30 algorithms to a reference standard of females who received a prenatal HIV screening test in the study observation period, and who were at least 15 years old at the time of the test to validate the algorithms in a cohort that contained only true HIV positive and true HIV negative individuals. Sensitivity in the fourth sensitivity analysis was higher than in the main analysis while PPV was lower. Multiple algorithms had sensitivity of 100% (Algorithms 1, 7, 10, 13, 16, 22, 25, 28) and the algorithm with the highest PPV was Algorithm 20 (five or more physician visits in three years) compared to Algorithm 6 (six or more physician visits in two years) in the main analysis. The fourth sensitivity analysis may be used to exclude 16 algorithms from consideration as the best algorithm due to having PPV $\leq$ 49.8%; this includes all algorithms that required antiretroviral prescriptions and those that required only one physician visit (Algorithms 1, 7, 10-16, 22, 25-30).

Overall, the sensitivity analyses were helpful to show that in future applications of the algorithms, it is not necessary to include ICD-9-CM codes V08, 795.51, 079.53, F02.4 and O98.7 and ICD-10-CA codes R75 and Z71 when searching the hospitalization and physician visit data. They sensitivity analyses also showed that while the algorithms can be applied to data for children, they perform better when applied to data for adults only. Lastly, the results of the fourth sensitivity analysis, while not generalizable to the population of Manitoba because it is limited pregnant women only, helps to reinforce that the algorithms that require antiretroviral prescriptions or only one physician visit are not reliable for identifying HIV cases because they capture too many false positives. The algorithms selected as best under the main analysis in Table 5.1 are supported by the findings of the sensitivity analysis; the only difference would be the overall best algorithm

selected from the cohort of true positive and true negative cases, in the fourth sensitivity analysis because no algorithms had PPV greater than 90%. For the fourth sensitivity analysis, the algorithm with the highest possible PPV while maximizing sensitivity over the shortest duration examined (in years) would be Algorithm 5, requiring five or more HIV-related physician visits in two years.

5.3 Comparison to Literature

These findings extend those of previous Canadian HIV algorithm validation studies performed by Antoniou et al. (9) and Nosyk et al. (10) Antoniou et al. identified that the optimal algorithm to identify HIV cases in the ICES administrative databases was at least three HIV-related physician visits in three years with specificity of 99.6% (95% CI: 99.1%-99.8%), sensitivity of 96.2% (95% CI: 94.0%-97.7%), kappa statistic of 0.97 (95% CI: 0.95-0.98), and AUC of 0.979. (9) Nosyk et al. identified that the optimal algorithm to identify HIV cases using data held by the BCMoH, the BCCDC, and the BCCfE was at least three HIV-related physician visits or at least one HIV-related hospital admission within 15 years with a sensitivity of 88%.

This study had a larger validation cohort than both the Antoniou et al. and Nosyk et al. studies, a lower prevalence of HIV in the validation cohort, and a different method for ascertaining the validation cohort. Antoniou et al. used a validation cohort of chart abstractions from two primary care clinics in downtown Toronto where approximately 20% of patients were known to have HIV. (9) Nosyk et al. used a validation cohort of confirmed HIV cases (i.e. an HIV-positive test in the BCCDC surveillance database or a viral load test with detectable viral load, a CD4 test, or an HIV-related antiretroviral dispensation in the BCCfE database) who had linkable records in each of the BCCfE, BCCDC, and BCMoH databases. (10) This study differs from the Antoniou et al. and Nosyk et al. studies most notably in that, rather than identifying one “best” algorithm, this study details several algorithm options to use in different scenarios.

The recommendation of overall best algorithm when using Manitoba’s administrative data (Table 5.1), defined as having the highest possible PPV (≥90%) while maximizing sensitivity (≥90%) over the shortest duration examined (in years), was similar to that of Antoniou et al., (9) with the only difference being the number of years required. In the first sensitivity analysis in this study, which utilized the same ICD-9-CM codes and ICD-10-CA codes as Antoniou et al., (9), the algorithm requiring three physician claims in three years had specificity of 100.0% (95% CI

100.0% - 100.0%), sensitivity of 89.7% (95% CI 88.1% - 91.1%), and kappa of 0.91 (95% CI 0.89 - 0.92).

Conversely, none of the algorithms recommended by this study were similar to the one chosen by Nosyk et al. (10) The only algorithm recommended in this study that required hospitalization data was Algorithm 13 (one or more physician visits, 1 or more hospitalizations, or two or more antiretroviral prescriptions in two years), which was recommended for use when wanting to identify every possible HIV case, regardless of the number of false positives produced. The performance of algorithms in this study may differ from the performance of algorithms in the Nosyk et al. (10) study because the Nosyk et al. (10) did not include HIV non-cases in their validation cohort. That said, the algorithm chosen by Nosyk et al. (10) still performed well in this study. The results of the second sensitivity analysis, which utilized the same ICD-9-CM codes and ICD-10-CA codes as Nosyk et al. (10) had sensitivity of 92.8% (95% CI 91.4% - 94.1%) for the algorithm requiring at least three physician visits or at least one hospital admission in two years (Algorithm 9) and 93.2% (95% CI 91.9% - 94.4%) for the algorithm requiring at least three physician visits or at least one hospital admission in three years (Algorithm 24).

5.4 Study Strengths and Limitations

The use of administrative data for research comes with inherent strengths and limitations. The biggest strength of administrative data lies in its population-based nature which allowed for the algorithms to be validated against a reference standard that reflected the low prevalence of HIV in the Manitoba population. All HIV testing in Manitoba during the study observation period was processed by CPL, making the reference standard very complete for those who sought out and received laboratory testing. (16)

The diagnostic HIV tests performed by CPL are the current gold standard with sensitivity and specificity >99% three months after exposure to the virus. (14, 15) Other confirmatory HIV tests, such as viral load tests, are only performed on diagnosed HIV cases. The presence of these types of tests in the CPL data indicate the person was previously diagnosed with HIV. (Personal communication, Cadham Provincial Laboratory, 2018) The only HIV cases that should be missing from the CPL data are those who have not been tested. In 2018, it was estimated that 13% of HIV cases in Canada were unaware of their HIV status. (26) Assuming Manitoba HIV cases were

similar, approximately 240 people living with HIV who never received HIV testing would be misclassified as HIV non-cases in the reference standard.

Administrative data only capture those people who sought and received care. The algorithms are only able to identify people who have regular contacts with the healthcare system, so misclassification of cases does occur. People who are undiagnosed and unaware of their status and/or those who were diagnosed pre-2007 and have been lost to follow up will be missed. This has implications for HIV research in populations that are underserved by current healthcare systems

The biggest threat to the value of this work is that the algorithms were validated against a reference standard that did not contain true negatives, making it difficult to ascertain if a false positive case was truly an HIV negative individual, or whether they were a person living with HIV who simply did not have a linkable, positive result, on a confirmatory laboratory test since 2007, but were being correctly identified by the algorithm. This was evaluated in the fourth sensitivity analysis by validating the algorithms against a subset of the population (pregnant women) for whom positive and negative HIV tests were available.

Data from the reference standard were not available prior to 2007. It is uncertain whether the algorithms will perform with similar accuracy in time periods before and after the study observation period due to changes in the ICD-9-CM and ICD-10-CA codes used over time and potential revisions to ICD codes in the future, (95, 96) healthcare practices, availability of antiretroviral therapy, and even prevalence. To validate the algorithms prior to 2007, a different reference standard would need to be used.

The generalizability of the algorithms for application in other provinces, territories, and countries outside of Canada depends on both whether the demographics of the populations align and the availability and structure of similar datasets and similar provision of healthcare. It cannot be assumed that the algorithms validated in this study will perform the same in all administrative health data sources.

This study was based on two previous Canadian validation studies that validated the same HIV case definitions using administrative data in Ontario and BC. (9, 10) The methods for the Antoniou et al. (9) study were well documented and provided guidance for my research, including which algorithms to validate and which DINs to include. These two previous Canadian HIV

algorithm validation studies also allowed for the results of the performance estimates to be compared to the results of the same algorithms applied in two other Canadian populations. (9, 10)

5.5 Future Research

Having validated algorithms will allow for more longitudinal HIV research using administrative data in Manitoba. All HIV tests performed prior to January 1, 2007 do not have a valid identifier so any HIV cases diagnosed prior to 2007, who did not have confirmatory testing after 2007, cannot be linked to the other administrative datasets (12) While the percentage of HIV cases with valid identifiers has increased since January 1, 2007, it has not reached 100%. (12) In addition, the CPL data only capture those people who have had confirmatory testing. It does not contain information on any HIV cases who were previously diagnosed outside the province, moved to Manitoba, and continued their regimen of care without being re-tested. Therefore, a validated algorithm that uses administrative data sources such as physician billings, hospitalizations, and prescriptions can fill the gaps in the CPL testing data.

Although the algorithms were not validated outside of the study observation period, the algorithms could be applied with caution to time periods prior to when the reference standard is available. Selecting an algorithm which has similar data available before the study observation period as during the study observation would allow for more confidence that the algorithm will perform similarly outside the study observation period. For example, algorithms that rely on prescription data may perform differently in different time periods due to when ART was introduced, changes in what medications were available, public policy around drug coverage, and prescribing practices whereas algorithms which rely on hospitalizations and physician visits are likely to produce more stable results over time.

The HIV algorithm validated for use in Manitoba may be considered alongside the other validated Canadian HIV algorithms to identify HIV cases for inclusion in pan-Canadian research studies through the Health Data Research Network or PHAC's Canadian Chronic Disease Surveillance System (CCDSS). (111, 112) It is important to note that the ICD-9-CM and ICD-10-CA codes that did not contribute to algorithm development when searching the Hospital Abstracts database were valid ICD codes for identifying HIV infection, but were not used by coders in Manitoba. These ICD-9-CM and ICD-10-CM codes may be used in other jurisdictions and may therefore be useful for constructing algorithms with pan-Canadian applicability.

Future research may develop Manitoba-specific algorithms to validate. An additional source of administrative data which could be considered are electronic medical records. Electronic medical records were used in several HIV algorithm validation studies in the USA. (74-77) A model-based approach to algorithm development, rather than a deterministic approach to algorithm development could also be considered. A model-based approach to algorithm development would use statistical or machine learning methods to identify other factors that may contribute to HIV case identification, such as time between physician visits, temporal relationship between physician visits and prescriptions, or age. (113-115)

5.6 Conclusion

This study demonstrates that administrative data in Manitoba can be used to accurately identify people living with HIV who interact with the healthcare system. Different algorithms perform best under different scenarios. For example, the algorithm that required six or more physician visits in two years (Algorithm 6) will produce a sample of HIV cases that has characteristics mostly closely resembling the reference standard. The algorithm that required one or more physician visits or one or more hospitalizations or two or more ART prescriptions in two years (Algorithm 13) will produce a sample of all possible HIV cases without concern for the number of false positives captured. The overall best algorithm was three physician visits in two years, chosen for having high PPV and high sensitivity over the shortest time frame examined. I recommend this algorithm for use in HIV research to allow for ease of comparison of HIV cases and results across future studies of people living with HIV in Manitoba.

The findings of this study will inform future population-based research and will serve to improve HIV research across Canada and abroad. The validated algorithms may be used to study progression of disease, comorbid illnesses, long-term outcomes, access to HIV care and treatment, or any other HIV-related research with the aim of improving the wellbeing of people living with HIV and contributing to the UNAIDS' and the United Nations General Assembly's goal of ending the AIDS epidemic by 2030. (6-8)

References

1. Government of Canada. Risks of HIV and AIDS 2019 [updated 2015-11-302019-08-16]. Available from: https://www.canada.ca/en/public-health/services/diseases/hiv-aids/risks-hiv-aids.html.
2. Government of Canada. Causes of HIV and AIDS 2019 [updated 2015-11-302019-08-16]. Available from: https://www.canada.ca/en/public-health/services/diseases/hiv-aids/causes-hiv-aids.html.
3. Government of Manitoba. Communicable Disease Management Protocol: Hepatitis C. Winnipeg, Manitoba; 2014.
4. Marcus LJ, Chao RC, Leyden AW, Xu PL, Quesenberry BC, Klein JD, et al. Narrowing the Gap in Life Expectancy Between HIV-Infected and HIV-Uninfected Individuals With Access to Care. JAIDS Journal of Acquired Immune Deficiency Syndromes. 2016;73(1):39-46.
5. Joint United Nations Programme on HIV/AIDS. Fact Sheet – World AIDS Day 2021. 2021.
6. United Nations. Transforming our World: The 2030 Agenda for Sustainable Development. 2015.
7. Joint United Nations Programme on HIV/AIDS. Global AIDS Strategy 2021–2026: End Inequalities. End AIDS. Geneva; 2022.
8. Joint United Nations Programme on HIV/AIDS. Ending Inequalities and Getting on Track to End AIDS by 2030. Geneva; 2022.
9. Antoniou T, Zagorski B, Loutfy MR, Strike C, Glazier RH. Validation of Case-Finding Algorithms Derived from Administrative Data for Identifying Adults Living with Human Immunodeficiency Virus Infection. PLoS ONE. 2011;6(6).
10. Nosyk B, Colley G, Yip B, Chan K, Heath K, Lima VD, et al. Application and Validation of Case-Finding Algorithms for Identifying Individuals with Human Immunodeficiency Virus from Administrative Data in British Columbia, Canada. PLoS ONE. 2013;8(1).
11. McClarty LM, Cheuk E, Ireland L, Kendall C, Bibeau C, Loeppky C, et al. Cohort profile: The LHIV-Manitoba clinical cohort of people living with HIV in Manitoba, Canada. BMJ Open. 2020;10(5).
12. Lix L, Smith M, Azimaee M, Dahl M, Nicol P, Burchill C, et al. A Systematic Investigation of Manitoba's Provincial Laboratory Data. Winnipeg, MB; 2012.
13. Hardy D. Fundamentals of HIV medicine 2017. Hardy WD, editor. New York: Oxford University Press; 2017.
14. Public Health Agency of Canada. Human Immunodeficiency Virus - HIV Screening and Testing Guide Ottawa, ON2019 [updated 2014-08-05. Available from: https://www.canada.ca/en/public-health/services/hiv-aids/hiv-screening-testing-guide.html#e5.
15. Bullard J. HIV Point of Care Testing in Manitoba Report: 2016 to 2017. Winnipeg, MB: Cadham Provincial Laboratory; 2018.
16. Government of Manitoba. Communicable Disease Management Protocol: HIV/AIDS. Winnipeg, Manitoba; 2010.
17. Altfeld M, Barmeyer A, Behrens G, Berzow D, Boesecke C, Braun P, et al. HIV 2015/16. 23 ed. Hoffmann C, Rockstroh J, editors. Hamburg: Medizin Fokus Verlag; 2015.
18. Hogg RS, Heath KV, Yip B, Craib KJP, O'Shaughnessy MV, Schechter MT, et al. Improved survival among HIV-infected individuals following initiation of antiretroviral therapy. Jama-Journal of the American Medical Association. 1998;279(6):450-4.
19. Public Health Agency of Canada. HIV Factsheet: U=U for Health Professionals. 2019.
20. World Health Organization. Global Health Sector Strategy on HIV 2016–2021 Towards Ending AIDS. 2016.
21. Joint United Nations Programme on HIV/AIDS. On the Fast-Track to End AIDS: 2016-2021 Strategy. 2016.

22. Joint United Nations Programme on HIV/AIDS. Unequal, Unprepared, Under Threat: Why bold action against inequalities is needed to end AIDS, stop COVID-19 and prepare for future pandemics. Geneva; 2021.
23. Joint United Nations Programme on HIV/AIDS. Confronting Inequalities: Lessons for pandemic responses from 40 years of AIDS. Geneva; 2021.
24. Joint United Nations Programme on HIV/AIDS. AIDSinfo: Global data on HIV epidemiology and response 2020 [Available from: https://aidsinfo.unaids.org.
25. World Health Organization. The top 10 causes of death Geneva2019 [updated December 9, 2020. Available from: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.
26. Public Health Agency of Canada. Estimates of HIV Incidence, Prevalence and Canada's Progress On Meeting The 90-90-90 HIV Targets, 2018. Ottawa, Ontario; 2020.
27. Public Health Agency of Canada. HIV in Canada: 2020 Surveillance highlights. Ottawa, Ontario: Government of Canada,; 2021 December 1, 2021.
28. Haddad N, Weeks A, Robert A, Totten S. HIV in Canada-surveillance report, 2019. Canada communicable disease report. 2021;47(1):77-86.
29. Statistics Canada. Table 13-10-0394-01 Leading causes of death, total population, by age group [updated 2022-01-24. Available from: https://www150.statcan.gc.ca/t1/tbl1/en/cv!recreate.action?pid=1310039401&selectedNodeIds=4D13&checkedLevels=0D1,1D1,2D1,4D1&refPeriods=20000101,20200101&dimensionLayouts=layout2,layout2,layout2,layout2,layout2,layout3&vectorDisplay=false.
30. Fransoo R, Mahar A, The Need To Know Team, Anderson A, Prior H, Koseva I, et al. The 2019 RHA Indicators Atlas. Winnipeg, MB: Manitoba Centre for Health Policy; 2019.
31. Manitoba HIV Program. 2018 Manitoba HIV Program Update. Winnipeg, Manitoba; 2019.
32. Government of Manitoba MHaSC, Epidemiology and Surveillance,. 2020 Annual Surveillance Update: HIV in Manitoba. 2021.
33. Roos LL, Nicol JP, Cageorge SM. Using administrative data for longitudinal research: Comparisons with primary data collection. Journal of Chronic Diseases. 1987;40(1):41-9.
34. Elias P. Administrative Data. In: Dusa A, Nelle D, Stock G, Wagner G, editors. Facing the Future: European Research Infrastructures for the Humanities
and Social Sciences. 1 ed. Berlin: SCIVERO; 2014.
35. De Coster C, Quan HD, Finlayson A, Gao M, Halfon P, Humphries K, et al. Identifying priorities in methodological research using ICD-9-ICM and ICD-10 administrative data: report from an international consortium. Bmc Health Services Research. 2006;6(1).
36. Roos LL, Menec V, Currie RJ. Policy analysis in an information-rich environment. Social Science & Medicine. 2004;58(11):2231-41.
37. Katz A, Enns J, Smith M, Burchill C, Turner K, Towns D. Population Data Centre Profile: The Manitoba Centre for Health Policy. International journal of population data science. 2020;4(2).
38. Andrews RM, Keyes MA, Pine PL. Acquired Immunodeficiency Syndrome in California's Medicaid Program, 1981-84. Health Care Financ R. 1988;10(1):95-103.
39. Keyes M, Andrews R, Mason ML. A Methodology for Building an AIDS Research File Using Medicaid Claims and Administrative Data Bases. J Acq Immun Def Synd. 1991;4(10):1015-24.
40. Fanning TR, Turner BJ, Cosler LE, OetjenGerdes L, Markson LE, McKee L, et al. The Quality of Medicaid Data for HIV/AIDS Research: Examination of a Statewide Data Base. Aids Public Policy J. 1995;10(1):39-47.
41. Thornton C, Fasciano N, Turner BJ, Cherlow A, Bencio DS. Methods for Identifying AIDS Cases in Medicare and Medicaid Claims Data. Princeton, NJ: Mathematica Policy Research: Health Care Financing Administration; 1997.
42. Fasciano NJ, Cherlow AL, Turner BJ, Thornton CV. Profile of Medicare Beneficiaries with AIDS: Application of an AIDS Casefinding Algorithm. Health Care Financ R. 1998;19(3):19-38.

43. Mrus JM, Shireman TI, Moomaw CJ, Tsevat J. Development of an HIV Research Database Using Medicaid Claims Data. Aids Public Policy J. 2001;16(3-4):48-54.
44. Chesnut TJ, Laufer FN, Carrascal AF, Feldman IS. An Expenditure Analysis of High-Cost Medicaid Recipients with HIV Disease in New York State. Journal of Health Care for the Poor and Underserved. 2011;22(1):330-45.
45. Leibowitz AA, Desmond K. Identifying a Sample of HIV-Positive Beneficiaries from Medicaid Claims Data and Estimating their Treatment Costs. 2015;105(3):567.
46. Walkup JT, Wei WH, Sambamoorthi U, Crystal S. Sensitivity of an AIDS Case-Finding Algorithm: Who Are We Missing? Medical Care. 2004;42(8):756-63.
47. Manitoba Centre for Health Policy. Manitoba Population Research Data Repository Data Descriptions: Drug Program Information Network Winnipeg, MB: University of Manitoba; 2020 [Available from: http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/resources/repository/descriptions.html?ds=DPIN.
48. BC Centre for Excellence in HIV/AIDS. How to Obtain HIV Medication through the Drug Treatment Program. 2020.
49. Manitoba Centre for Health Policy. Manitoba Population Research Data Repository Data Descriptions: CPL - Laboratory Information Management System (LIMS) Winnipeg, MB: University of Manitoba; 2020 [Available from: http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/resources/repository/descriptions.html?ds=CPL_LIMS.
50. Institute for Clinical Evaluative Sciences. The value of Ontario's electronic health data infrastructure: A brief report from the perspective of the Institute for Clinical Evaluative Sciences. 2016.
51. Ontario HIV Epidemiology and Surveillance Initiative. HIV Datamart 2020 [Available from: http://www.ohesi.ca/technical-notes-2/.
52. Antoniou T, Zagorski B, Loutfy MR, Strike C, Glazier RH. Socio-economic and sex-related disparities in rates of hospital admission among patients with HIV infection in Ontario: A population-based study. Open Medicine. 2012;6(4):146-54.
53. Antoniou T, Zagorski B, Bayoumi AM, Loutfy MR, Strike C, Raboud J, et al. Trends in HIV prevalence, new HIV diagnoses, and mortality among adults with HIV who entered care in Ontario, 1996/1997 to 2009/2010: A population-based study. Open Medicine. 2013;7(4):98-106.
54. Kendall CE, Wong J, Taljaard M, Glazier RH, Hogg W, Younger J, et al. A cross-sectional, population-based study measuring comorbidity among people living with HIV in Ontario. BMC Public Health. 2014;14(1).
55. Antoniou T, Zagorski B, Macdonald EM, Bayoumi AM, Raboud J, Brophy J, et al. Trends in live birth rates and adverse neonatal outcomes among HIV-positive women in Ontario, Canada, 2002–2009: a descriptive population-based study. International Journal of STD and AIDS. 2014;25(13):960-6.
56. Ng R, Macdonald EM, Loutfy MR, Yudin MH, Raboud J, Masinde KI, et al. Adequacy of prenatal care among women living with human immunodeficiency virus: A population-based study. BMC Public Health. 2015;15(1).
57. Macdonald EM, Ng R, Yudin MH, Bayoumi AM, Loutfy M, Raboud J, et al. Postpartum Maternal and Neonatal Hospitalizations among Women with HIV: A Population-Based Study. AIDS Research and Human Retroviruses. 2015;31(10):967-72.
58. Macdonald EM, Ng R, Bayoumi AM, Raboud J, Brophy J, Masinde KI, et al. Adverse Neonatal Outcomes Among Women Living With HIV: A Population-Based Study. Journal of Obstetrics and Gynaecology Canada. 2015;37(4):302-9.
59. Kendall CE, Younger J, Manuel DG, Hogg W, Glazier RH, Taljaard M. Health administrative data can be used to define a shared care typology for people with HIV. Journal of Clinical Epidemiology. 2015;68(11):1301-11.

60. Kendall CE, Wong J, Taljaard M, Glazier RH, Hogg W, Younger J, et al. A cross-sectional, population-based study of HIV physicians and outpatient health care use by people with HIV in Ontario. BMC Health Services Research. 2015;15(1).
61. Kendall CE, Walmsley S, Lau C, Jembere N, Burchell AN, Loutfy M, et al. A cross-sectional population-based study of breast cancer screening among women with HIV in Ontario, Canada. CMAJ open. 2017;5(3):E673-E81.
62. Durbin A, Brown HK, Bansal S, Antoniou T, Jung JKH, Lunsky Y. How HIV affects health and service use for adults with intellectual and developmental disabilities. Journal of Intellectual Disability Research. 2017;61(7):682-96.
63. Burchell AN, Kendall CE, Cheng SY, Lofters A, Cotterchio M, Bayoumi AM, et al. Cervical cancer screening uptake among HIV-positive women in Ontario, Canada: A population-based retrospective cohort study. Preventive Medicine. 2018;107:14-20.
64. Wilton J, Liu J, Sullivan A, Rachlis B, Marchand-Austin A, Giles M, et al. Trends in HIV care cascade engagement among diagnosed people living with HIV in Ontario, Canada: A retrospective, population-based cohort study. PLoS ONE. 2019;14(1).
65. Van Meer R, Antoniou T, McCormack D, Khanna S, Kendall C, Kiefer L, et al. A comparison of the observed and expected prevalence of HIV in persons released from Ontario provincial prisons in 2010. Canadian Journal of Public Health. 2019;110(5):675-82.
66. Sparrow-Downes VM, Loutfy M, Antoniou T, Vigod SN. Postpartum mental health service utilization in women with Human Immunodeficiency Virus (HIV): a population-based study. AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV. 2019;31(11):1332-9.
67. Khanna S, Leah J, Fung K, Antoniou T, Kouyoumdjian F. Health care utilization by people with HIV on release from provincial prison in Ontario, Canada in 2010: a retrospective cohort study. AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV. 2019;31(7):785-92.
68. Kouyoumdjian FG, Lamarche L, McCormack D, Rowe J, Kiefer L, Kroch A, et al. 90-90-90 for everyone?: Access to HIV care and treatment for people with HIV who experience imprisonment in Ontario, Canada. AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV. 2020;32(9):1168-76.
69. Stewart A, Antoniou T, Graves E, Plumptre L, Carusone SC. Health care utilization in medically complex people living with HIV before and after admission to an HIV-specific community facility: a pre-post comparison study. CMAJ open. 2021;9(2):E460-E5.
70. Shoemaker ES, Saiyin T, Smith S, Loutfy M, Darling L, Walker M, et al. Patterns of cesarean birth among women living with HIV in Ontario: A cross-sectional, population-level study. Birth. 2021;48(3):357-65.
71. Mbuagbaw L, Logie CH, Thabane L, Smaill F, Smieja M, Burchell AN, et al. The impact of routine HIV drug resistance testing in Ontario: A controlled interrupted time series study. PLoS ONE. 2021;16(4 April).
72. Keyes MA, Andrews, R., Pine, P., et al. A Comparison of Two Methods of Identifying AIDS Cases in Medical Claims Analyses. Paper presented at American Public Health Association Annual Meeting; New Orleans. 1987.
73. Turner BJ, Mckee L, Markson L, Fanning T. Benefits of Shared Care for Advanced HIV-Infection. Clin Res. 1993;41(2):A552-A.
74. Fultz SL, Skanderson M, Mole LA, Gandhi N, Bryant K, Crystal S, et al. Development and Verification of a “Virtual" Cohort Using the National VA Health Information System. Medical Care. 2006;44(8):S25-S30.
75. Felsen UR, Bellin EY, Cunningham CO, Zingman BS. Development of an Electronic Medical Record-Based Algorithm to Identify Patients with Unknown HIV Status. Aids Care-Psychological and Socio-Medical Aspects of Aids/Hiv. 2014;26(10):1318-25.

76. Goetz MB, Hoang T, Kan VL, Rimland D, Rodriguez-Barradas M. Development and Validation of an Algorithm to Identify Patients Newly Diagnosed with HIV Infection from Electronic Health Records. AIDS Research and Human Retroviruses. 2014;30(7):626-33.
77. Paul DW, Neely NB, Clement M, Riley I, Al-Hegela M, Phelan M, et al. Development and Validation of an Electronic Medical Record (EMR)-Based Computed Phenotype of HIV-1 Infection. Journal of the American Medical Informatics Association. 2018;25(2):150-7.
78. Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttmann A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. Journal of Clinical Epidemiology. 2011;64(8):821-9.
79. Roos LL, Brownell M, Lix L, Roos NP, Walld R, MacWilliam L. From health research to social research: Privacy, methods, approaches. Social science & medicine (1982). 2008;66(1):117-29.
80. Manitoba Centre for Health Policy. Concept: Manitoba Population Research Data Repository (Repository) Winnipeg, MB: University of Manitoba; 2019 [Available from: http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1419.
81. Roos LL, Nicol JP. A Research Registry: Uses, Development, and Accuracy. Journal of Clinical Epidemiology. 1999;52(1):39-47.
82. Lix L, Smith M, Pitz M, Ahmed R, Quon H, Griffith J, et al. Cancer Data Linkage in Manitoba: Expanding the Infrastructure for Research. Winnipeg, MB; 2016.
83. Manitoba Centre for Health Policy. Manitoba Population Research Data Repository Data Descriptions: CPL - Cadham Provincial Laboratory Winnipeg, MB: University of Manitoba; 2020 [Available from: http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/resources/repository/descriptions.html?ds=Cadham.
84. Government of Manitoba MH, Seniors and Active Living, Epidemiology and Surveillance,. 2018 Annual Statistical Update: HIV in Manitoba. 2019.
85. Manitoba Centre for Health Policy. Administrative Health Data: Drug Data. Winnipeg, MB: University of Manitoba; 2009.
86. Manitoba Centre for Health Policy. Manitoba Population Research Data Repository Data Descriptions: Hospital Abstracts Winnipeg, MB: University of Manitoba; 2020 [Available from: http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/resources/repository/descriptions.html?ds=Hospital.
87. Manitoba Centre for Health Policy. Concept: International Classification of Disease (ICD) Coding in MCHP Data Winnipeg, MB: University of Manitoba; 2019 [updated 2019-01-17. Available from: http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?conceptID=1028#DxClaims.
88. Manitoba Centre for Health Policy. Manitoba Population Research Data Repository Data Descriptions: Medical Claims / Medical Services Winnipeg, MB: University of Manitoba; 2020 [Available from: http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/resources/repository/descriptions.html?ds=MedicalClaims.
89. Manitoba Centre for Health Policy. Manitoba Population Research Data Repository Data Descriptions: Manitoba Health Insurance Registry Winnipeg, MB: University of Manitoba; 2020 [Available from: http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/resources/repository/descriptions.html?ds=Insurance.
90. Manitoba Centre for Health Policy. Manitoba Population Research Data Repository Data Descriptions: Canada Census Winnipeg, MB: University of Manitoba; 2020 [Available from: http://umanitoba.ca/faculties/health_sciences/medicine/units/chs/departmental_units/mchp/resources/repository/descriptions.html?ds=Census.

91. Lix L, Yogendran M, Burchill C, Metge C, McKeen N, Moore D, et al. Defining and Validating Chronic Diseases: An Administrative Data Approach. Winnipeg, Manitoba; 2006.
92. National Center for Health Statistics. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM). Centers for Disease Control and Prevention,; 2011.
93. Canadian Institute for Health Information. International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Canada (ICD-10-CA) - Tabular List2015.
94. Canadian Institute for Health Information. Canadian Coding Standards for Version 2018 ICD-10-CA and CCI. Ottawa, Ontario: CIHI; 2018.
95. Lurie P. Official Authorized Addenda: Human Immunodeficiency Virus Infection Codes and Official Guidelines for Coding and Reporting ICD-9-CM. Centers for Disease Control and Prevention; 1994.
96. National Center for Health Statistics, Centers for Medicare & Medicaid Services. Conversion Table of New ICD-9-CM Codes, October 2013. Centers for Disease Control and Prevetion; 2013.
97. McClarty L. A critical exploration into Manitoba's HIV care cascade. 2020.
98. Canadian Institute for Health Information. In Pursuit of Health Equity: Defining Stratifiers for Measuring Health Inequality — A Focus on Age, Sex, Gender, Income, Education and Geographic Location. Ottawa, ON:: CIHI; 2018.
99. Kuhn M, Johnson K. Applied Predictive Modelling. USA: Springer; 2016.
100. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall; 1991.
101. Hosmer D, Lemeshow S, Sturdivant R. Applied Logistic Regression. 3 ed: John Wiley & Sons, Inc.; 2013.
102. Leeflang MMG, Bossuyt PMM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. Journal of clinical epidemiology. 2009;62(1):5-12.
103. Leeflang MMG, Rutjes AWS, Reitsma JB, Hooft L, Bossuyt PMM. Variation of a test's sensitivity and specificity with disease prevalence. Canadian Medical Association journal (CMAJ). 2013;185(11):E537-E44.
104. Brenner H, Gefeller O. Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. Statistics in medicine. 1997;16(9):981.
105. Legislative Assembly of Manitoba. The Personal Health Information Act C.C.S.M. c. P33.5 [Internet] Canada [Available from: http://web2.gov.mb.ca/laws/statutes/ccsm/p033-5e.php.
106. Legislative Assembly of Manitoba. The Freedom of Information and Protection of Privacy Act C.C.S.M. c. F175 [Internet] Canada [Available from: https://web2.gov.mb.ca/laws/statutes/ccsm/f175e.php.
107. Government of Manitoba MH, Seniors and Active Living, Epidemiology and Surveillance,. 2019 Annual Statistical Update: HIV in Manitoba. 2020.
108. Government of Manitoba. Population Report June 1, 2018. Winnipeg, MB: Manitoba Health, Seniors and Active Living; 2018.
109. Statistics Canada. Manitoba [Province] and Canada [Country] (table). Census Profile. Ottawa. 2017. [updated November 29, 2017]. Available from: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E
110. Drug Product Database [Internet]. Health Canada. 2021 [cited January 12, 2022]. Available from: https://health-products.canada.ca/dpd-bdpp/index-eng.jsp.
111. Health Data Research Network Canada. About 2020 [Available from: https://www.hdrn.ca/en/about.
112. Government of Canada. The Canadian Chronic Disease Surveillance System – An Overview Ottawa, Ontario: Public Health Agency of Canada,; 2021 [updated May 6, 2021. Available from: https://www.canada.ca/en/public-health/services/publications/canadian-chronic-disease-surveillance-system-factsheet.html.

113. Lix LM, Yogendran MS, Leslie WD, Shaw SY, Baumgartner R, Bowman C, et al. Using multiple data features improved the validity of osteoporosis case ascertainment from administrative databases. Journal of Clinical Epidemiology. 2008;61(12):1250-60.
114. Kroeker K, Widdifield J, Muthukumarana S, Jiang D, Lix LM. Model-based methods for case definitions from administrative health data: application to rheumatoid arthritis. BMJ open. 2017;7(6):e016173-e.
115. van Walraven C, Austin PC, Manuel D, Knoll G, Jennings A, Forster AJ. The usefulness of administrative databases for identifying disease cohorts is increased with a multivariate model. Journal of clinical epidemiology. 2010;63(12):1332-41.

Appendix A - Tables of descriptive characteristics for the 30 HIV administrative data algorithms

Table A.1. Description of Algorithm 1 (1+ physician claims in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,518)	**False Positives (n = 1,416)**	**Overall (N = 2,934)**	**True Negatives (n = 1,451,005)**	**False Negatives (n = 71)**	**Overall (N = 1,451,076)**
Age at index date, years	37 ± 12.4	29.3 ± 17.9	33.3 ± 15.8	32.7 ± 23.3	25.4 ± 12.7	32.7 ± 23.3
Sex						
Male	1,004 (66.1)	661 (46.7)	1,665 (56.7)	722,296 (49.8)	49 (69)	722,345 (49.8)
Female	514 (33.9)	755 (53.3)	1,269 (43.3)	728,709 (50.2)	22 (31)	728,731 (50.2)
Area of residence (RHA)						
IERHA	74 (4.9)	79 (5.6)	153 (5.2)	131,401 (9.1)	7 (9.9)	131,408 (9.1)
NHR	64 (4.2)	99 (7)	163 (5.6)	s	s	88,224 (6.1)
SH-SS	63 (4.2)	82 (5.8)	145 (4.9)	s	s	202,349 (13.9)
PMH	62 (4.1)	105 (7.4)	167 (5.7)	s	s	189,401 (13.1)
WRHA	1,244 (81.9)	1,027 (72.5)	2,271 (77.4)	834,691 (57.5)	52 (73.2)	834,743 (57.5)
Public Trustee/Not Found	11 (0.7)	24 (1.7)	35 (1.2)	s	s	4,951 (0.3)
Income quintile						
Q1-Lowest/Not Found	788 (51.9)	533 (37.6)	1,321 (45)	316,354 (21.8)	32 (45.1)	316,386 (21.8)
Q2	284 (18.7)	262 (18.5)	546 (18.6)	291,869 (20.1)	21 (29.6)	291,890 (20.1)
Q3	204 (13.4)	198 (14)	402 (13.7)	283,196 (19.5)	11 (15.5)	283,207 (19.5)
Q4	134 (8.8)	228 (16.1)	362 (12.3)	s	s	283,579 (19.5)
Q5-Highest	108 (7.1)	195 (13.8)	303 (10.3)	s	s	276,014 (19)
Health insurance coverage, years	10.4 ± 2.6	10.5 ± 2.6	10.4 ± 2.6	10.3 ± 2.8	9.6 ± 2.9	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.2. Description of Algorithm 2 (2+ physician claims in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,462)	**False Positives (n = 268)**	**Overall (N = 1,730)**	**True Negatives (n = 1,452,153)**	**False Negatives (n = 127)**	**Overall (N = 1,452,280)**
Age at index date, years	36.9 ± 12.4	26.8 ± 20.7	35.3 ± 14.4	32.7 ± 23.3	32.2 ± 15	32.7 ± 23.3
Sex						
Male	965 (66)	131 (48.9)	1,096 (63.4)	722,826 (49.8)	88 (69.3)	722,914 (49.8)
Female	497 (34)	137 (51.1)	634 (36.6)	729,327 (50.2)	39 (30.7)	729,366 (50.2)
Area of residence (RHA)						
IERHA	74 (5.1)	19 (7.1)	93 (5.4)	131,461 (9.1)	7 (5.5)	131,468 (9.1)
NHR	64 (4.4)	17 (6.3)	81 (4.7)	s	s	88,306 (6.1)
SH-SS	63 (4.3)	17 (6.3)	80 (4.6)	s	s	202,414 (13.9)
PMH	s	s	67 (3.9)	189,490 (13)	11 (8.7)	189,501 (13)
WRHA	1,193 (81.6)	200 (74.6)	1,393 (80.5)	835,518 (57.5)	103 (81.1)	835,621 (57.5)
Public Trustee/Not Found	s	s	16 (0.9)	4,970 (0.3)	0 (0)	4,970 (0.3)
Income quintile						
Q1-Lowest/Not Found	753 (51.5)	122 (45.5)	875 (50.6)	316,765 (21.8)	67 (52.8)	316,832 (21.8)
Q2	279 (19.1)	42 (15.7)	321 (18.6)	292,089 (20.1)	26 (20.5)	292,115 (20.1)
Q3	196 (13.4)	36 (13.4)	232 (13.4)	283,358 (19.5)	19 (15)	283,377 (19.5)
Q4	129 (8.8)	34 (12.7)	163 (9.4)	283,772 (19.5)	6 (4.7)	283,778 (19.5)
Q5-Highest	105 (7.2)	34 (12.7)	139 (8)	276,169 (19)	9 (7.1)	276,178 (19)
Health insurance coverage, years	10.5 ± 2.6	9.7 ± 3.3	10.3 ± 2.7	10.3 ± 2.8	8.6 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.3. Description of Algorithm 3 (3+ physician claims in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,427)	False Positives (n = 139)	Overall (N = 1,566)	True Negatives (n = 1,452,282)	False Negatives (n = 162)	Overall (N = 1,452,444)
Age at index date, years	37 ± 12.3	23.6 ± 22.3	35.8 ± 14	32.7 ± 23.3	32.1 ± 14.6	32.7 ± 23.3
Sex						
Male	939 (65.8)	74 (53.2)	1,013 (64.7)	722,883 (49.8)	114 (70.4)	722,997 (49.8)
Female	488 (34.2)	65 (46.8)	553 (35.3)	729,399 (50.2)	48 (29.6)	729,447 (50.2)
Area of residence (RHA)						
IERHA	71 (5)	8 (5.8)	79 (5)	131,472 (9.1)	10 (6.2)	131,482 (9.1)
NHR	63 (4.4)	8 (5.8)	71 (4.5)	s	s	88,316 (6.1)
SH-SS	60 (4.2)	6 (4.3)	66 (4.2)	202,421 (13.9)	7 (4.3)	202,428 (13.9)
PMH	s	s	60 (3.8)	189,497 (13)	11 (6.8)	189,508 (13)
WRHA	1,166 (81.7)	110 (79.1)	1,276 (81.5)	835,608 (57.5)	130 (80.2)	835,738 (57.5)
Public Trustee/Not Found	s	s	14 (0.9)	s	s	4,972 (0.3)
Income quintile						
Q1-Lowest/Not Found	733 (51.4)	70 (50.4)	803 (51.3)	316,817 (21.8)	87 (53.7)	316,904 (21.8)
Q2	274 (19.2)	23 (16.5)	297 (19)	292,108 (20.1)	31 (19.1)	292,139 (20.1)
Q3	192 (13.5)	13 (9.4)	205 (13.1)	283,381 (19.5)	23 (14.2)	283,404 (19.5)
Q4	127 (8.9)	21 (15.1)	148 (9.5)	283,785 (19.5)	8 (4.9)	283,793 (19.5)
Q5-Highest	101 (7.1)	12 (8.6)	113 (7.2)	276,191 (19)	13 (8)	276,204 (19)
Health insurance coverage, years	10.5 ± 2.5	8.8 ± 3.7	10.4 ± 2.7	10.3 ± 2.8	8.7 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.4. Description of Algorithm 4 (4+ physician claims in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,388)	False Positives (n = 87)	HIV Case (N = 1,475)	True Negatives (n = 1,452,334)	False Negatives (n = 201)	HIV Non-Case (N = 1,452,535)
Age at index date, years	37.1 ± 12.3	21.5 ± 22.1	36.2 ± 13.6	32.7 ± 23.3	32.0 ± 14.1	32.7 ± 23.3
Sex						
Male	913 (65.8)	48 (55.2)	961 (65.2)	722,909 (49.8)	140 (69.7)	723,049 (49.8)
Female	475 (34.2)	39 (44.8)	514 (34.8)	729,425 (50.2)	61 (30.3)	729,486 (50.2)
Area of residence (RHA)						
IERHA	70 (5)	7 (8)	77 (5.2)	131,473 (9.1)	11 (5.5)	131,484 (9.1)
NHR	s	s	62 (4.2)	s	s	88,325 (6.1)
SH-SS	s	s	62 (4.2)	202,425 (13.9)	7 (3.5)	202,432 (13.9)
PMH	s	s	59 (4)	189,497 (13)	12 (6)	189,509 (13)
WRHA	1,133 (81.6)	70 (80.5)	1,203 (81.6)	835,648 (57.5)	163 (81.1)	835,811 (57.5)
Public Trustee/Not Found	s	s	12 (0.8)	s	s	4,974 (0.3)
Income quintile						
Q1-Lowest/Not Found	715 (51.5)	47 (54)	762 (51.7)	316,840 (21.8)	105 (52.2)	316,945 (21.8)
Q2	262 (18.9)	13 (14.9)	275 (18.6)	292,118 (20.1)	43 (21.4)	292,161 (20.1)
Q3	186 (13.4)	8 (9.2)	194 (13.2)	283,386 (19.5)	29 (14.4)	283,415 (19.5)
Q4	125 (9)	12 (13.8)	137 (9.3)	283,794 (19.5)	10 (5)	283,804 (19.5)
Q5-Highest	100 (7.2)	7 (8)	107 (7.3)	276,196 (19)	14 (7)	276,210 (19)
Health insurance coverage, years	10.5 ± 2.5	8.3 ± 3.9	10.4 ± 2.7	10.3 ± 2.8	8.8 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.5. Description of Algorithm 5 (5+ physician claims in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,342)	False Positives (n = 66)	Overall (N = 1,408)	True Negatives (n = 1,452,355)	False Negatives (n = 247)	Overall (N = 1,452,602)
Age at index date, years	37.3 ± 12.3	22.9 ± 23	36.7 ± 13.3	32.7 ± 23.3	31.9 ± 13.7	32.7 ± 23.3
Sex						
Male	884 (65.9)	39 (59.1)	923 (65.6)	722,918 (49.8)	169 (68.4)	723,087 (49.8)
Female	458 (34.1)	27 (40.9)	485 (34.4)	729,437 (50.2)	78 (31.6)	729,515 (50.2)
Area of residence (RHA)						
IERHA	s	s	70 (5)	131,477 (9.1)	14 (5.7)	131,491 (9.1)
NHR	s	s	60 (4.3)	s	s	88,327 (6.1)
SH-SS	s	s	60 (4.3)	202,426 (13.9)	8 (3.2)	202,434 (13.9)
PMH	s	s	58 (4.1)	189,497 (13)	13 (5.3)	189,510 (13)
WRHA	1,093 (81.4)	56 (84.8)	1,149 (81.6)	835,662 (57.5)	203 (82.2)	835,865 (57.5)
Public Trustee/Not Found	s	s	11 (0.8)	s	s	4,975 (0.3)
Income quintile						
Q1-Lowest/Not Found	689 (51.3)	36 (54.5)	725 (51.5)	316,851 (21.8)	131 (53)	316,982 (21.8)
Q2	253 (18.9)	11 (16.7)	264 (18.8)	292,120 (20.1)	52 (21.1)	292,172 (20.1)
Q3	s	s	186 (13.2)	283,388 (19.5)	35 (14.2)	283,423 (19.5)
Q4	121 (9)	9 (13.6)	130 (9.2)	283,797 (19.5)	14 (5.7)	283,811 (19.5)
Q5-Highest	s	s	103 (7.3)	276,199 (19)	15 (6.1)	276,214 (19)
Health insurance coverage, years	10.6 ± 2.5	8.2 ± 3.9	10.4 ± 2.6	10.3 ± 2.8	9 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.6. Description of Algorithm 6 (6+ physician claims in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,289)	False Positives (n = 54)	Overall (N = 1,343)	True Negatives (n = 1,452,367)	False Negatives (n = 300)	Overall (N = 1,452,667)
Age at index date, years	37.5 ± 12.3	21.9 ± 23.4	36.9 ± 13.3	32.7 ± 23.3	32 ± 13.2	32.7 ± 23.3
Sex						
Male	852 (66.1)	31 (57.4)	883 (65.7)	722,926 (49.8)	201 (67.0)	723,127 (49.8)
Female	437 (33.9)	23 (42.6)	460 (34.3)	729,441 (50.2)	99 (33.0)	729,540 (50.2)
Area of residence (RHA)						
IERHA	s	s	68 (5.1)	131,477 (9.1)	16 (5.3)	131,493 (9.1)
NHR	s	s	59 (4.4)	s	s	88,328 (6.1)
SH-SS	s	s	58 (4.3)	202,426 (13.9)	10 (3.3)	202,436 (13.9)
PMH	s	s	57 (4.2)	189,497 (13.0)	14 (4.7)	189,511 (13.0)
WRHA	1,047 (81.2)	44 (81.5)	1,091 (81.2)	835,674 (57.5)	249 (83.0)	835,923 (57.5)
Public Trustee/Not Found	s	s	10 (0.7)	s	s	4,976 (0.3)
Income quintile						
Q1-Lowest/Not Found	667 (51.7)	31 (57.4)	698 (52.0)	316,856 (21.8)	153 (51.0)	317,009 (21.8)
Q2	247 (19.2)	9 (16.7)	256 (19.1)	292,122 (20.1)	58 (19.3)	292,180 (20.1)
Q3	s	s	173 (12.9)	283,390 (19.5)	46 (15.3)	283,436 (19.5)
Q4	115 (8.9)	6 (11.1)	121 (9.0)	283,800 (19.5)	20 (6.7)	283,820 (19.5)
Q5-Highest	s	s	95 (7.1)	276,199 (19.0)	23 (7.7)	276,222 (19.0)
Health insurance coverage, years	10.6 ± 2.5	7.8 ± 3.9	10.4 ± 2.6	10.3 ± 2.8	9.3 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.7. Description of Algorithm 7 (1+ physician claims in 2 years or 1+ hospital discharges in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,514)	False Positives (n = 794)	Overall (N = 2,308)	True Negatives (n = 1,451,627)	False Negatives (n = 75)	Overall (N = 1,451,702)
Age at index date, years	36.9 ± 12.5	29.3 ± 20.1	34.3 ± 15.9	32.7 ± 23.3	27.5 ± 13.1	32.7 ± 23.3
Sex						
Male	998 (65.9)	401 (50.5)	1,399 (60.6)	722,556 (49.8)	55 (73.3)	722,611 (49.8)
Female	516 (34.1)	393 (49.5)	909 (39.4)	729,071 (50.2)	20 (26.7)	729,091 (50.2)
Area of residence (RHA)						
IERHA	76 (5)	54 (6.8)	130 (5.6)	s	s	131,431 (9.1)
NHR	64 (4.2)	46 (5.8)	110 (4.8)	s	s	88,277 (6.1)
SH-SS	64 (4.2)	76 (9.6)	140 (6.1)	s	s	202,354 (13.9)
PMH	57 (3.8)	42 (5.3)	99 (4.3)	189,459 (13.1)	10 (13.3)	189,469 (13.1)
WRHA	1,241 (82)	568 (71.5)	1,809 (78.4)	835,150 (57.5)	55 (73.3)	835,205 (57.5)
Public Trustee/Not Found	12 (0.8)	8 (1)	20 (0.9)	4,966 (0.3)	0 (0)	4,966 (0.3)
Income quintile						
Q1-Lowest/Not Found	788 (52)	319 (40.2)	1,107 (48)	316,568 (21.8)	32 (42.7)	316,600 (21.8)
Q2	285 (18.8)	143 (18)	428 (18.5)	291,988 (20.1)	20 (26.7)	292,008 (20.1)
Q3	201 (13.3)	131 (16.5)	332 (14.4)	283,263 (19.5)	14 (18.7)	283,277 (19.5)
Q4	133 (8.8)	109 (13.7)	242 (10.5)	s	s	283,699 (19.5)
Q5-Highest	107 (7.1)	92 (11.6)	199 (8.6)	s	s	276,118 (19)
Health insurance coverage, years	10.4 ± 2.6	10.3 ± 2.7	10.4 ± 2.7	10.3 ± 2.8	9.1 ± 3.1	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.8. Description of Algorithm 8 (2+ physician claims in 2 years or 1+ hospital discharges in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,499)	False Positives (n = 357)	Overall (N = 1,856)	True Negatives (n = 1,452,064)	False Negatives (n = 90)	Overall (N = 1,452,154)
Age at index date, years	37.1 ± 12.4	29.7 ± 20	35.6 ± 14.5	32.7 ± 23.3	27.2 ± 12.7	32.7 ± 23.3
Sex						
Male	990 (66)	184 (51.5)	1,174 (63.3)	722,773 (49.8)	63 (70)	722,836 (49.8)
Female	509 (34)	173 (48.5)	682 (36.7)	729,291 (50.2)	27 (30)	729,318 (50.2)
Area of residence (RHA)						
IERHA	75 (5)	23 (6.4)	98 (5.3)	131,457 (9.1)	6 (6.7)	131,463 (9.1)
NHR	64 (4.3)	24 (6.7)	88 (4.7)	s	s	88,299 (6.1)
SH-SS	64 (4.3)	21 (5.9)	85 (4.6)	s	s	202,409 (13.9)
PMH	56 (3.7)	21 (5.9)	77 (4.1)	189,480 (13)	11 (12.2)	189,491 (13)
WRHA	1,228 (81.9)	262 (73.4)	1,490 (80.3)	835,456 (57.5)	68 (75.6)	835,524 (57.5)
Public Trustee/Not Found	12 (0.8)	6 (1.7)	18 (1)	4,968 (0.3)	0 (0)	4,968 (0.3)
Income quintile						
Q1-Lowest/Not Found	780 (52)	168 (47.1)	948 (51.1)	316,719 (21.8)	40 (44.4)	316,759 (21.8)
Q2	282 (18.8)	56 (15.7)	338 (18.2)	292,075 (20.1)	23 (25.6)	292,098 (20.1)
Q3	200 (13.3)	51 (14.3)	251 (13.5)	283,343 (19.5)	15 (16.7)	283,358 (19.5)
Q4	130 (8.7)	42 (11.8)	172 (9.3)	s	s	283,769 (19.5)
Q5-Highest	107 (7.1)	40 (11.2)	147 (7.9)	s	s	276,170 (19)
Health insurance coverage, years	10.4 ± 2.6	9.9 ± 3.1	10.3 ± 2.7	10.3 ± 2.8	9.2 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.9. Description of Algorithm 9 (3+ physician claims in 2 years or 1+ hospital discharges in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,475)	False Positives (n = 228)	Overall (N = 1,703)	True Negatives (n = 1,452,193)	False Negatives (n = 114)	Overall (N = 1,452,307)
Age at index date, years	37.2 ± 12.4	29.3 ± 20.9	36.1 ± 14.1	32.7 ± 23.3	27.3 ± 12.8	32.7 ± 23.3
Sex						
Male	973 (66)	127 (55.7)	1,100 (64.6)	722,830 (49.8)	80 (70.2)	722,910 (49.8)
Female	502 (34)	101 (44.3)	603 (35.4)	729,363 (50.2)	34 (29.8)	729,397 (50.2)
Area of residence (RHA)						
IERHA	73 (4.9)	12 (5.3)	85 (5)	131,468 (9.1)	8 (7)	131,476 (9.1)
NHR	63 (4.3)	15 (6.6)	78 (4.6)	s	s	88,309 (6.1)
SH-SS	s	s	72 (4.2)	s	s	202,422 (13.9)
PMH	56 (3.8)	14 (6.1)	70 (4.1)	189,487 (13)	11 (9.6)	189,498 (13)
WRHA	1,210 (82)	172 (75.4)	1,382 (81.2)	835,546 (57.5)	86 (75.4)	835,632 (57.5)
Public Trustee/Not Found	s	s	16 (0.9)	s	s	4,970 (0.3)
Income quintile						
Q1-Lowest/Not Found	767 (52)	116 (50.9)	883 (51.8)	316,771 (21.8)	53 (46.5)	316,824 (21.8)
Q2	280 (19)	37 (16.2)	317 (18.6)	292,094 (20.1)	25 (21.9)	292,119 (20.1)
Q3	196 (13.3)	28 (12.3)	224 (13.2)	283,366 (19.5)	19 (16.7)	283,385 (19.5)
Q4	128 (8.7)	29 (12.7)	157 (9.2)	283,777 (19.5)	7 (6.1)	283,784 (19.5)
Q5-Highest	104 (7.1)	18 (7.9)	122 (7.2)	276,185 (19)	10 (8.8)	276,195 (19)
Health insurance coverage, years	10.4 ± 2.6	9.4 ± 3.3	10.3 ± 2.7	10.3 ± 2.8	9.3 ± 3.1	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.10. Description of Algorithm 10 (1+ physician claims in 2 years or 2+ antiretroviral prescriptions in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n =1,528)	False Positives (n = 1,858)	Overall (N = 3,386)	True Negatives (n = 1,450,563)	False Negatives (n = 61)	Overall (N = 1,450,624)
Age at index date, years	36.8 ± 12.4	29 ± 18.6	32.5 ± 16.6	32.7 ± 23.3	28 ± 15.2	32.7 ± 23.3
Sex						
Male	1,007 (65.9)	872 (46.9)	1,879 (55.5)	722,085 (49.8)	46 (75.4)	722,131 (49.8)
Female	521 (34.1)	986 (53.1)	1,507 (44.5)	728,478 (50.2)	15 (24.6)	728,493 (50.2)
Area of residence (RHA)						
IERHA	76 (5)	117 (6.3)	193 (5.7)	s	s	131,368 (9.1)
NHR	64 (4.2)	122 (6.6)	186 (5.5)	s	s	88,201 (6.1)
SH-SS	64 (4.2)	142 (7.6)	206 (6.1)	s	s	202,288 (13.9)
PMH	62 (4.1)	123 (6.6)	185 (5.5)	s	s	189,383 (13.1)
WRHA	1,250 (81.8)	1,330 (71.6)	2,580 (76.2)	834,388 (57.5)	46 (75.4)	834,434 (57.5)
Public Trustee/Not Found	12 (0.8)	24 (1.3)	36 (1.1)	4,950 (0.3)	0 (0)	4,950 (0.3)
Income quintile						
Q1-Lowest/Not Found	792 (51.8)	682 (36.7)	1,474 (43.5)	316,205 (21.8)	28 (45.9)	316,233 (21.8)
Q2	287 (18.8)	342 (18.4)	629 (18.6)	291,789 (20.1)	18 (29.5)	291,807 (20.1)
Q3	207 (13.5)	281 (15.1)	488 (14.4)	283,113 (19.5)	8 (13.1)	283,121 (19.5)
Q4	133 (8.7)	298 (16)	431 (12.7)	s	s	283,510 (19.5)
Q5-Highest	109 (7.1)	255 (13.7)	364 (10.8)	s	s	275,953 (19)
Health insurance coverage, years	10.4 ± 2.6	10.6 ± 2.5	10.5 ± 2.6	10.3 ± 2.8	9.2 ± 3.1	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.11. Description of Algorithm 11 (2+ physician claims in 2 years or 2+ antiretroviral prescriptions in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,518)	False Positives (n = 1,438)	Overall (N = 2,956)	True Negatives (n = 1,450,983)	False Negatives (n = 71)	Overall (N = 1,451,054)
Age at index date, years	36.9 ± 12.4	28.9 ± 18	33 ± 15.9	32.7 ± 23.3	28.5 ± 14.6	32.7 ± 23.3
Sex						
Male	1,002 (66)	659 (45.8)	1,661 (56.2)	722,298 (49.8)	51 (71.8)	722,349 (49.8)
Female	516 (34)	779 (54.2)	1,295 (43.8)	728,685 (50.2)	20 (28.2)	728,705 (50.2)
Area of residence (RHA)						
IERHA	75 (4.9)	86 (6)	161 (5.4)	131,394 (9.1)	6 (8.5)	131,400 (9.1)
NHR	64 (4.2)	100 (7)	164 (5.5)	s	s	88,223 (6.1)
SH-SS	63 (4.2)	88 (6.1)	151 (5.1)	s	s	202,343 (13.9)
PMH	62 (4.1)	101 (7)	163 (5.5)	s	s	189,405 (13.1)
WRHA	1,242 (81.8)	1,041 (72.4)	2,283 (77.2)	834,677 (57.5)	54 (76.1)	834,731 (57.5)
Public Trustee/Not Found	12 (0.8)	22 (1.5)	34 (1.2)	4,952 (0.3)	0 (0)	4,952 (0.3)
Income quintile						
Q1-Lowest/Not Found	785 (51.7)	535 (37.2)	1,320 (44.7)	316,352 (21.8)	35 (49.3)	316,387 (21.8)
Q2	285 (18.8)	260 (18.1)	545 (18.4)	291,871 (20.1)	20 (28.2)	291,891 (20.1)
Q3	206 (13.6)	202 (14)	408 (13.8)	283,192 (19.5)	9 (12.7)	283,201 (19.5)
Q4	133 (8.8)	235 (16.3)	368 (12.4)	s	s	283,573 (19.5)
Q5-Highest	109 (7.2)	206 (14.3)	315 (10.7)	s	s	276,002 (19)
Health insurance coverage, years	10.4 ± 2.6	10.5 ± 2.6	10.4 ± 2.6	10.3 ± 2.8	9.2 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.12. Description of Algorithm 12 (3+ physician claims in 2 years or 2+ antiretroviral prescriptions in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,503)	False Positives (n = 1,323)	Overall (N = 2,826)	True Negatives (n = 1,451,098)	False Negatives (n = 86)	Overall (N = 1,451,184)
Age at index date, years	37 ± 12.4	28.8 ± 17.9	33.1 ± 15.7	32.7 ± 23.3	28.2 ± 14.1	32.7 ± 23.3
Sex						
Male	992 (66)	609 (46)	1,601 (56.7)	722,348 (49.8)	61 (70.9)	722,409 (49.8)
Female	511 (34)	714 (54)	1,225 (43.3)	728,750 (50.2)	25 (29.1)	728,775 (50.2)
Area of residence (RHA)						
IERHA	74 (4.9)	75 (5.7)	149 (5.3)	131,405 (9.1)	7 (8.1)	131,412 (9.1)
NHR	64 (4.3)	92 (7)	156 (5.5)	s	s	88,231 (6.1)
SH-SS	61 (4.1)	77 (5.8)	138 (4.9)	202,350 (13.9)	6 (7)	202,356 (13.9)
PMH	62 (4.1)	95 (7.2)	157 (5.6)	s	s	189,411 (13.1)
WRHA	1,231 (81.9)	963 (72.8)	2,194 (77.6)	834,755 (57.5)	65 (75.6)	834,820 (57.5)
Public Trustee/Not Found	11 (0.7)	21 (1.6)	32 (1.1)	s	s	4,954 (0.3)
Income quintile						
Q1-Lowest/Not Found	778 (51.8)	487 (36.8)	1,265 (44.8)	316,400 (21.8)	42 (48.8)	316,442 (21.8)
Q2	283 (18.8)	244 (18.4)	527 (18.6)	291,887 (20.1)	22 (25.6)	291,909 (20.1)
Q3	203 (13.5)	182 (13.8)	385 (13.6)	283,212 (19.5)	12 (14)	283,224 (19.5)
Q4	132 (8.8)	222 (16.8)	354 (12.5)	s	s	283,587 (19.5)
Q5-Highest	107 (7.1)	188 (14.2)	295 (10.4)	s	s	276,022 (19)
Health insurance coverage, years	10.4 ± 2.6	10.5 ± 2.6	10.4 ± 2.6	10.3 ± 2.8	9.2 ± 3.1	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.13. Description of Algorithm 13 (1+ physician claims in 2 years or 1+ hospital discharges in 2 years or 2+ antiretroviral prescriptions in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,534)	False Positives (n = 1,934)	Overall (N = 3,468)	True Negatives (n = 1,450,487)	False Negatives (n = 55)	Overall (N = 1,450,542)
Age at index date, years	36.9 ± 12.5	29.4 ± 18.5	32.7 ± 16.6	32.7 ± 23.3	25.5 ± 13.1	32.7 ± 23.3
Sex						
Male	1,012 (66)	917 (47.4)	1,929 (55.6)	722,040 (49.8)	41 (74.5)	722,081 (49.8)
Female	522 (34)	1,017 (52.6)	1,539 (44.4)	728,447 (50.2)	14 (25.5)	728,461 (50.2)
Area of residence (RHA)						
IERHA	76 (5)	120 (6.2)	196 (5.7)	s	s	131,365 (9.1)
NHR	64 (4.2)	129 (6.7)	193 (5.6)	s	s	88,194 (6.1)
SH-SS	64 (4.2)	145 (7.5)	209 (6)	s	s	202,285 (13.9)
PMH	62 (4)	132 (6.8)	194 (5.6)	s	s	189,374 (13.1)
WRHA	1,256 (81.9)	1,383 (71.5)	2,639 (76.1)	834,335 (57.5)	40 (72.7)	834,375 (57.5)
Public Trustee/Not Found	12 (0.8)	25 (1.3)	37 (1.1)	4,949 (0.3)	0 (0)	4,949 (0.3)
Income quintile						
Q1-Lowest/Not Found	797 (52)	720 (37.2)	1,517 (43.7)	316,167 (21.8)	23 (41.8)	316,190 (21.8)
Q2	287 (18.7)	355 (18.4)	642 (18.5)	291,776 (20.1)	18 (32.7)	291,794 (20.1)
Q3	207 (13.5)	295 (15.3)	502 (14.5)	283,099 (19.5)	8 (14.5)	283,107 (19.5)
Q4	134 (8.7)	303 (15.7)	437 (12.6)	s	s	283,504 (19.5)
Q5-Highest	109 (7.1)	261 (13.5)	370 (10.7)	s	s	275,947 (19)
Health insurance coverage, years	10.3 ± 2.6	10.6 ± 2.5	10.5 ± 2.6	10.3 ± 2.8	9.5 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.14. Description of Algorithm 14 (2+ physician claims in 2 years or 1+ hospital discharges in 2 years or 2+ antiretroviral prescriptions in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,526)	False Positives (n = 1,519)	Overall (N = 3,045)	True Negatives (n = 1,450,902)	False Negatives (n = 63)	Overall (N = 1,450,965)
Age at index date, years	36.9 ± 12.5	29.4 ± 18	33.2 ± 15.9	32.7 ± 23.3	25.9 ± 12.6	32.7 ± 23.3
Sex						
Male	1,009 (66.1)	708 (46.6)	1,717 (56.4)	722,249 (49.8)	44 (69.8)	722,293 (49.8)
Female	517 (33.9)	811 (53.4)	1,328 (43.6)	728,653 (50.2)	19 (30.2)	728,672 (50.2)
Area of residence (RHA)						
IERHA	75 (4.9)	89 (5.9)	164 (5.4)	131,391 (9.1)	6 (9.5)	131,397 (9.1)
NHR	64 (4.2)	107 (7)	171 (5.6)	s	s	88,216 (6.1)
SH-SS	64 (4.2)	91 (6)	155 (5.1)	s	s	202,339 (13.9)
PMH	62 (4.1)	111 (7.3)	173 (5.7)	s	s	189,395 (13.1)
WRHA	1,249 (81.8)	1,098 (72.3)	2,347 (77.1)	834,620 (57.5)	47 (74.6)	834,667 (57.5)
Public Trustee/Not Found	12 (0.8)	23 (1.5)	35 (1.1)	4,951 (0.3)	0 (0)	4,951 (0.3)
Income quintile						
Q1-Lowest/Not Found	792 (51.9)	576 (37.9)	1,368 (44.9)	316,311 (21.8)	28 (44.4)	316,339 (21.8)
Q2	285 (18.7)	274 (18)	559 (18.4)	291,857 (20.1)	20 (31.7)	291,877 (20.1)
Q3	206 (13.5)	216 (14.2)	422 (13.9)	283,178 (19.5)	9 (14.3)	283,187 (19.5)
Q4	134 (8.8)	241 (15.9)	375 (12.3)	s	s	283,566 (19.5)
Q5-Highest	109 (7.1)	212 (14)	321 (10.5)	s	s	275,996 (19)
Health insurance coverage, years	10.3 ± 2.6	10.5 ± 2.5	10.4 ± 2.6	10.3 ± 2.8	9.6 ± 2.9	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.15. Description of Algorithm 15 (3+ physician claims in 2 years or 1+ hospital discharges in 2 years or 2+ antiretroviral prescriptions in 2 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,514)	False Positives (n = 1,404)	Overall (N = 2,918)	True Negatives (n = 1,451,017)	False Negatives (n = 75)	Overall (N = 1,451,092)
Age at index date, years	37 ± 12.4	29.4 ± 17.9	33.3 ± 15.8	32.7 ± 23.3	25.9 ± 12.6	32.7 ± 23.3
Sex						
Male	1,002 (66.2)	658 (46.9)	1,660 (56.9)	722,299 (49.8)	51 (68)	722,350 (49.8)
Female	512 (33.8)	746 (53.1)	1,258 (43.1)	728,718 (50.2)	24 (32)	728,742 (50.2)
Area of residence (RHA)						
IERHA	74 (4.9)	78 (5.6)	152 (5.2)	131,402 (9.1)	7 (9.3)	131,409 (9.1)
NHR	64 (4.2)	99 (7.1)	163 (5.6)	s	s	88,224 (6.1)
SH-SS	63 (4.2)	80 (5.7)	143 (4.9)	s	s	202,351 (13.9)
PMH	62 (4.1)	105 (7.5)	167 (5.7)	s	s	189,401 (13.1)
WRHA	1,240 (81.9)	1,020 (72.6)	2,260 (77.5)	834,698 (57.5)	56 (74.7)	834,754 (57.5)
Public Trustee/Not Found	11 (0.7)	22 (1.6)	33 (1.1)	s	s	4,953 (0.3)
Income quintile						
Q1-Lowest/Not Found	787 (52)	528 (37.6)	1,315 (45.1)	316,359 (21.8)	33 (44)	316,392 (21.8)
Q2	283 (18.7)	258 (18.4)	541 (18.5)	291,873 (20.1)	22 (29.3)	291,895 (20.1)
Q3	203 (13.4)	196 (14)	399 (13.7)	283,198 (19.5)	12 (16)	283,210 (19.5)
Q4	133 (8.8)	228 (16.2)	361 (12.4)	s	s	283,580 (19.5)
Q5-Highest	108 (7.1)	194 (13.8)	302 (10.3)	s	s	276,015 (19)
Health insurance coverage, years	10.4 ± 2.6	10.5 ± 2.6	10.4 ± 2.6	10.3 ± 2.8	9.6 ± 2.9	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.16. Description of Algorithm 16 (1+ physician claims in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,484)	False Positives (n = 710)	Overall (N = 2,194)	True Negatives (n = 1,451,711)	False Negatives (n = 105)	Overall (N = 1,451,816)
Age at index date, years	36.8 ± 12.4	28.2 ± 20.4	34 ± 15.9	32.7 ± 23.3	32.4 ± 15.2	32.7 ± 23.3
Sex						
Male	977 (65.8)	352 (49.6)	1,329 (60.6)	722,605 (49.8)	76 (72.4)	722,681 (49.8)
Female	507 (34.2)	358 (50.4)	865 (39.4)	729,106 (50.2)	29 (27.6)	729,135 (50.2)
Area of residence (RHA)						
IERHA	75 (5.1)	50 (7)	125 (5.7)	131,430 (9.1)	6 (5.7)	131,436 (9.1)
NHR	64 (4.3)	39 (5.5)	103 (4.7)	s	s	88,284 (6.1)
SH-SS	64 (4.3)	72 (10.1)	136 (6.2)	s	s	202,358 (13.9)
PMH	57 (3.8)	33 (4.6)	90 (4.1)	189,468 (13.1)	10 (9.5)	189,478 (13.1)
WRHA	1,212 (81.7)	510 (71.8)	1,722 (78.5)	835,208 (57.5)	84 (80)	835,292 (57.5)
Public Trustee/Not Found	12 (0.8)	6 (0.8)	18 (0.8)	4,968 (0.3)	0 (0)	4,968 (0.3)
Income quintile						
Q1-Lowest/Not Found	767 (51.7)	276 (38.9)	1,043 (47.5)	316,611 (21.8)	53 (50.5)	316,664 (21.8)
Q2	282 (19)	130 (18.3)	412 (18.8)	292,001 (20.1)	23 (21.9)	292,024 (20.1)
Q3	198 (13.3)	116 (16.3)	314 (14.3)	283,278 (19.5)	17 (16.2)	283,295 (19.5)
Q4	132 (8.9)	102 (14.4)	234 (10.7)	s	s	283,707 (19.5)
Q5-Highest	105 (7.1)	86 (12.1)	191 (8.7)	s	s	276,126 (19)
Health insurance coverage, years	10.4 ± 2.6	10.3 ± 2.8	10.4 ± 2.6	10.3 ± 2.8	8.4 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.17. Description of Algorithm 17 (2+ physician claims in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,465)	False Positives (n = 270)	Overall (N = 1,735)	True Negatives (n = 1,452,151)	False Negatives (n = 124)	Overall (N = 1,452,275)
Age at index date, years	36.9 ± 12.3	26.8 ± 20.7	35.3 ± 14.4	32.7 ± 23.3	32.2 ± 15.2	32.7 ± 23.3
Sex						
Male	968 (66.1)	132 (48.9)	1,100 (63.4)	722,825 (49.8)	85 (68.5)	722,910 (49.8)
Female	497 (33.9)	138 (51.1)	635 (36.6)	729,326 (50.2)	39 (31.5)	729,365 (50.2)
Area of residence (RHA)						
IERHA	74 (5.1)	19 (7)	93 (5.4)	131,461 (9.1)	7 (5.6)	131,468 (9.1)
NHR	64 (4.4)	17 (6.3)	81 (4.7)	s	s	88,306 (6.1)
SH-SS	63 (4.3)	17 (6.3)	80 (4.6)	s	s	202,414 (13.9)
PMH	s	s	67 (3.9)	189,490 (13)	11 (8.9)	189,501 (13)
WRHA	1,196 (81.6)	202 (74.8)	1,398 (80.6)	835,516 (57.5)	100 (80.6)	835,616 (57.5)
Public Trustee/Not Found	s	s	16 (0.9)	4,970 (0.3)	0 (0)	4,970 (0.3)
Income quintile						
Q1-Lowest/Not Found	753 (51.4)	123 (45.6)	876 (50.5)	316,764 (21.8)	67 (54)	316,831 (21.8)
Q2	279 (19)	42 (15.6)	321 (18.5)	292,089 (20.1)	26 (21)	292,115 (20.1)
Q3	196 (13.4)	37 (13.7)	233 (13.4)	283,357 (19.5)	19 (15.3)	283,376 (19.5)
Q4	132 (9)	34 (12.6)	166 (9.6)	s	s	283,775 (19.5)
Q5-Highest	105 (7.2)	34 (12.6)	139 (8)	s	s	276,178 (19)
Health insurance coverage, years	10.5 ± 2.6	9.7 ± 3.3	10.3 ± 2.7	10.3 ± 2.8	8.6 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.18. Description of Algorithm 18 (3+ physician claims in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,434)	False Positives (n = 142)	Overall (N = 1,576)	True Negatives (n = 1,452,279)	False Negatives (n = 155)	Overall (N = 1,452,434)
Age at index date, years	37 ± 12.3	23.5 ± 22.3	35.8 ± 14	32.7 ± 23.3	31.8 ± 14.7	32.7 ± 23.3
Sex						
Male	944 (65.8)	76 (53.5)	1,020 (64.7)	722,881 (49.8)	109 (70.3)	722,990 (49.8)
Female	490 (34.2)	66 (46.5)	556 (35.3)	729,398 (50.2)	46 (29.7)	729,444 (50.2)
Area of residence (RHA)						
IERHA	71 (5)	8 (5.6)	79 (5)	131,472 (9.1)	10 (6.5)	131,482 (9.1)
NHR	63 (4.4)	8 (5.6)	71 (4.5)	s	s	88,316 (6.1)
SH-SS	60 (4.2)	7 (4.9)	67 (4.3)	202,420 (13.9)	7 (4.5)	202,427 (13.9)
PMH	s	s	60 (3.8)	189,497 (13)	11 (7.1)	189,508 (13)
WRHA	1,173 (81.8)	112 (78.9)	1,285 (81.5)	835,606 (57.5)	123 (79.4)	835,729 (57.5)
Public Trustee/Not Found	s	s	14 (0.9)	s	s	4,972 (0.3)
Income quintile						
Q1-Lowest/Not Found	737 (51.4)	72 (50.7)	809 (51.3)	316,815 (21.8)	83 (53.5)	316,898 (21.8)
Q2	275 (19.2)	23 (16.2)	298 (18.9)	292,108 (20.1)	30 (19.4)	292,138 (20.1)
Q3	193 (13.5)	14 (9.9)	207 (13.1)	283,380 (19.5)	22 (14.2)	283,402 (19.5)
Q4	128 (8.9)	21 (14.8)	149 (9.5)	283,785 (19.5)	7 (4.5)	283,792 (19.5)
Q5-Highest	101 (7)	12 (8.5)	113 (7.2)	276,191 (19)	13 (8.4)	276,204 (19)
Health insurance coverage, years	10.5 ± 2.5	8.8 ± 3.7	10.3 ± 2.7	10.3 ± 2.8	8.6 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.19. Description of Algorithm 19 (4+ physician claims in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,402)	False Positives (n = 90)	Overall (N = 1,492)	True Negatives (n = 1,452,331)	False Negatives (n = 187)	Overall (N = 1,452,518)
Age at index date, years	37.1 ± 12.3	21.1 ± 21.9	36.1 ± 13.6	32.7 ± 23.3	32.3 ± 14.3	32.7 ± 23.3
Sex						
Male	922 (65.8)	50 (55.6)	972 (65.1)	722,907 (49.8)	131 (70.1)	723,038 (49.8)
Female	480 (34.2)	40 (44.4)	520 (34.9)	729,424 (50.2)	56 (29.9)	729,480 (50.2)
Area of residence (RHA)						
IERHA	70 (5)	7 (7.8)	77 (5.2)	131,473 (9.1)	11 (5.9)	131,484 (9.1)
NHR	s	s	62 (4.2)	s	s	88,325 (6.1)
SH-SS	s	s	62 (4.2)	202,425 (13.9)	7 (3.7)	202,432 (13.9)
PMH	s	s	59 (4)	189,497 (13)	12 (6.4)	189,509 (13)
WRHA	1,147 (81.8)	73 (81.1)	1,220 (81.8)	835,645 (57.5)	149 (79.7)	835,794 (57.5)
Public Trustee/Not Found	s	s	12 (0.8)	s	s	4,974 (0.3)
Income quintile						
Q1-Lowest/Not Found	720 (51.4)	48 (53.3)	768 (51.5)	316,839 (21.8)	100 (53.5)	316,939 (21.8)
Q2	266 (19)	14 (15.6)	280 (18.8)	292,117 (20.1)	39 (20.9)	292,156 (20.1)
Q3	189 (13.5)	8 (8.9)	197 (13.2)	283,386 (19.5)	26 (13.9)	283,412 (19.5)
Q4	127 (9.1)	13 (14.4)	140 (9.4)	283,793 (19.5)	8 (4.3)	283,801 (19.5)
Q5-Highest	100 (7.1)	7 (7.8)	107 (7.2)	276,196 (19)	14 (7.5)	276,210 (19)
Health insurance coverage, years	10.5 ± 2.5	8.3 ± 3.9	10.4 ± 2.7	10.3 ± 2.8	8.7 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.20. Description of Algorithm 20 (5+ physician claims in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,363)	False Positives (n = 68)	Overall (N = 1,431)	True Negatives (n = 1,452,353)	False Negatives (n = 226)	Overall (N = 1,452,579)
Age at index date, years	37.3 ± 12.2	23.1 ± 23.2	36.6 ± 13.3	32.7 ± 23.3	31.9 ± 14	32.7 ± 23.3
Sex						
Male	899 (66)	40 (58.8)	939 (65.6)	722,917 (49.8)	154 (68.1)	723,071 (49.8)
Female	464 (34)	28 (41.2)	492 (34.4)	729,436 (50.2)	72 (31.9)	729,508 (50.2)
Area of residence (RHA)						
IERHA	s	s	74 (5.2)	131,476 (9.1)	11 (4.9)	131,487 (9.1)
NHR	s	s	61 (4.3)	s	s	88,326 (6.1)
SH-SS	s	s	61 (4.3)	202,426 (13.9)	7 (3.1)	202,433 (13.9)
PMH	s	s	58 (4.1)	189,497 (13)	13 (5.8)	189,510 (13)
WRHA	1,110 (81.4)	56 (82.4)	1,166 (81.5)	835,662 (57.5)	186 (82.3)	835,848 (57.5)
Public Trustee/Not Found	s	s	11 (0.8)	s	s	4,975 (0.3)
Income quintile						
Q1-Lowest/Not Found	699 (51.3)	38 (55.9)	737 (51.5)	316,849 (21.8)	121 (53.5)	316,970 (21.8)
Q2	257 (18.9)	11 (16.2)	268 (18.7)	292,120 (20.1)	48 (21.2)	292,168 (20.1)
Q3	s	s	191 (13.3)	283,388 (19.5)	30 (13.3)	283,418 (19.5)
Q4	123 (9)	9 (13.2)	132 (9.2)	283,797 (19.5)	12 (5.3)	283,809 (19.5)
Q5-Highest	s	s	103 (7.2)	276,199 (19)	15 (6.6)	276,214 (19)
Health insurance coverage, years	10.5 ± 2.5	8.2 ± 3.9	10.4 ± 2.6	10.3 ± 2.8	8.9 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.21. Description of Algorithm 21 (6+ physician claims in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,321)	False Positives (n = 57)	Overall (N = 1,378)	True Negatives (n = 1,452,364)	False Negatives (n = 268)	Overall (N = 1,452,632)
Age at index date, years	37.4 ± 12.2	22.2 ± 23.5	36.8 ± 13.2	32.7 ± 23.3	31.8 ± 13.6	32.7 ± 23.3
Sex						
Male	870 (65.9)	34 (59.6)	904 (65.6)	722,923 (49.8)	183 (68.3)	723,106 (49.8)
Female	451 (34.1)	23 (40.4)	474 (34.4)	729,441 (50.2)	85 (31.7)	729,526 (50.2)
Area of residence (RHA)						
IERHA	s	s	69 (5)	131,476 (9.1)	16 (6)	131,492 (9.1)
NHR	s	s	59 (4.3)	s	s	88,328 (6.1)
SH-SS	s	s	60 (4.4)	202,426 (13.9)	8 (3)	202,434 (13.9)
PMH	s	s	57 (4.1)	189,497 (13)	14 (5.2)	189,511 (13)
WRHA	1,076 (81.5)	46 (80.7)	1,122 (81.4)	835,672 (57.5)	220 (82.1)	835,892 (57.5)
Public Trustee/Not Found	s	s	11 (0.8)	s	s	4,975 (0.3)
Income quintile						
Q1-Lowest/Not Found	682 (51.6)	33 (57.9)	715 (51.9)	316,854 (21.8)	138 (51.5)	316,992 (21.8)
Q2	252 (19.1)	10 (17.5)	262 (19)	292,121 (20.1)	53 (19.8)	292,174 (20.1)
Q3	s	s	181 (13.1)	283,390 (19.5)	38 (14.2)	283,428 (19.5)
Q4	116 (8.8)	6 (10.5)	122 (8.9)	283,800 (19.5)	19 (7.1)	283,819 (19.5)
Q5-Highest	s	s	98 (7.1)	276,199 (19)	20 (7.5)	276,219 (19)
Health insurance coverage, years	10.6 ± 2.5	7.8 ± 3.9	10.5 ± 2.6	10.3 ± 2.8	9 ± 3.1	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.22. Description of Algorithm 22 (1+ physician claims in 3 years or 1+ hospital discharges in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,514)	False Positives (n = 794)	Overall (N = 2,308)	True Negatives (n = 1,451,627)	False Negatives (n = 75)	Overall (N = 1,451,702)
Age at index date, years	36.9 ± 12.5	29.3 ± 20.1	34.3 ± 15.9	32.7 ± 23.3	27.5 ± 13.1	32.7 ± 23.3
Sex						
Male	998 (65.9)	401 (50.5)	1,399 (60.6)	722,556 (49.8)	55 (73.3)	722,611 (49.8)
Female	516 (34.1)	393 (49.5)	909 (39.4)	729,071 (50.2)	20 (26.7)	729,091 (50.2)
Area of residence (RHA)						
IERHA	76 (5)	54 (6.8)	130 (5.6)	s	s	131,431 (9.1)
NHR	64 (4.2)	46 (5.8)	110 (4.8)	s	s	88,277 (6.1)
SH-SS	64 (4.2)	76 (9.6)	140 (6.1)	s	s	202,354 (13.9)
PMH	57 (3.8)	42 (5.3)	99 (4.3)	189,459 (13.1)	10 (13.3)	189,469 (13.1)
WRHA	1,241 (82)	568 (71.5)	1,809 (78.4)	835,150 (57.5)	55 (73.3)	835,205 (57.5)
Public Trustee/Not Found	12 (0.8)	8 (1)	20 (0.9)	4,966 (0.3)	0 (0)	4,966 (0.3)
Income quintile						
Q1-Lowest/Not Found	788 (52)	319 (40.2)	1,107 (48)	316,568 (21.8)	32 (42.7)	316,600 (21.8)
Q2	285 (18.8)	143 (18)	428 (18.5)	291,988 (20.1)	20 (26.7)	292,008 (20.1)
Q3	201 (13.3)	131 (16.5)	332 (14.4)	283,263 (19.5)	14 (18.7)	283,277 (19.5)
Q4	133 (8.8)	109 (13.7)	242 (10.5)	s	s	283,699 (19.5)
Q5-Highest	107 (7.1)	92 (11.6)	199 (8.6)	s	s	276,118 (19)
Health insurance coverage, years	10.4 ± 2.6	10.3 ± 2.7	10.4 ± 2.7	10.3 ± 2.8	9.1 ± 3.1	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.23. Description of Algorithm 23 (2+ physician claims in 3 years or 1+ hospital discharges in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,502)	False Positives (n = 359)	Overall (N = 1,861)	True Negatives (n = 1,452,062)	False Negatives (n = 87)	Overall (N = 1,452,149)
Age at index date, years	37 ± 12.4	29.6 ± 20	35.6 ± 14.5	32.7 ± 23.3	27 ± 12.9	32.7 ± 23.3
Sex						
Male	993 (66.1)	185 (51.5)	1,178 (63.3)	722,772 (49.8)	60 (69)	722,832 (49.8)
Female	509 (33.9)	174 (48.5)	683 (36.7)	729,290 (50.2)	27 (31)	729,317 (50.2)
Area of residence (RHA)						
IERHA	75 (5)	23 (6.4)	98 (5.3)	131,457 (9.1)	6 (6.9)	131,463 (9.1)
NHR	64 (4.3)	24 (6.7)	88 (4.7)	s	s	88,299 (6.1)
SH-SS	64 (4.3)	21 (5.8)	85 (4.6)	s	s	202,409 (13.9)
PMH	56 (3.7)	21 (5.8)	77 (4.1)	189,480 (13)	11 (12.6)	189,491 (13)
WRHA	1,231 (82)	264 (73.5)	1,495 (80.3)	835,454 (57.5)	65 (74.7)	835,519 (57.5)
Public Trustee/Not Found	12 (0.8)	6 (1.7)	18 (1)	4,968 (0.3)	0 (0)	4,968 (0.3)
Income quintile						
Q1-Lowest/Not Found	780 (51.9)	169 (47.1)	949 (51)	316,718 (21.8)	40 (46)	316,758 (21.8)
Q2	282 (18.8)	56 (15.6)	338 (18.2)	292,075 (20.1)	23 (26.4)	292,098 (20.1)
Q3	200 (13.3)	52 (14.5)	252 (13.5)	283,342 (19.5)	15 (17.2)	283,357 (19.5)
Q4	133 (8.9)	42 (11.7)	175 (9.4)	s	s	283,766 (19.5)
Q5-Highest	107 (7.1)	40 (11.1)	147 (7.9)	s	s	276,170 (19)
Health insurance coverage, years	10.4 ± 2.6	9.8 ± 3.1	10.3 ± 2.7	10.3 ± 2.8	9.2 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.24. Description of Algorithm 24 (3+ physician claims in 3 years or 1+ hospital discharges in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,481)	False Positives (n = 231)	Overall (N = 1,712)	True Negatives (n = 1,452,190)	False Negatives (n = 108)	Overall (N = 1,452,298)
Age at index date, years	37.2 ± 12.3	29.1 ± 21	36.1 ± 14.1	32.7 ± 23.3	26.8 ± 12.7	32.7 ± 23.3
Sex						
Male	977 (66)	129 (55.8)	1,106 (64.6)	722,828 (49.8)	76 (70.4)	722,904 (49.8)
Female	504 (34)	102 (44.2)	606 (35.4)	729,362 (50.2)	32 (29.6)	729,394 (50.2)
Area of residence (RHA)						
IERHA	73 (4.9)	12 (5.2)	85 (5)	131,468 (9.1)	8 (7.4)	131,476 (9.1)
NHR	63 (4.3)	15 (6.5)	78 (4.6)	s	s	88,309 (6.1)
SH-SS	s	s	73 (4.3)	s	s	202,421 (13.9)
PMH	56 (3.8)	14 (6.1)	70 (4.1)	189,487 (13)	11 (10.2)	189,498 (13)
WRHA	1,216 (82.1)	174 (75.3)	1,390 (81.2)	835,544 (57.5)	80 (74.1)	835,624 (57.5)
Public Trustee/Not Found	s	s	16 (0.9)	s	s	4,970 (0.3)
Income quintile						
Q1-Lowest/Not Found	770 (52)	118 (51.1)	888 (51.9)	316,769 (21.8)	50 (46.3)	316,819 (21.8)
Q2	281 (19)	37 (16)	318 (18.6)	292,094 (20.1)	24 (22.2)	292,118 (20.1)
Q3	197 (13.3)	29 (12.6)	226 (13.2)	283,365 (19.5)	18 (16.7)	283,383 (19.5)
Q4	129 (8.7)	29 (12.6)	158 (9.2)	283,777 (19.5)	6 (5.6)	283,783 (19.5)
Q5-Highest	104 (7)	18 (7.8)	122 (7.1)	276,185 (19)	10 (9.3)	276,195 (19)
Health insurance coverage, years	10.4 ± 2.6	9.4 ± 3.3	10.3 ± 2.7	10.3 ± 2.8	9.2 ± 3.1	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.25. Description of Algorithm 25 (1+ physician claims in 3 years or 2+ antiretroviral prescriptions in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,528)	False Positives (n = 1,867)	Overall (N = 3,395)	True Negatives (n = 1,450,554)	False Negatives (n = 61)	Overall (N = 1,450,615)
Age at index date, years	36.8 ± 12.4	28.9 ± 18.5	32.5 ± 16.6	32.7 ± 23.3	28 ± 15.2	32.7 ± 23.3
Sex						
Male	1,007 (65.9)	873 (46.8)	1,880 (55.4)	722,084 (49.8)	46 (75.4)	722,130 (49.8)
Female	521 (34.1)	994 (53.2)	1,515 (44.6)	728,470 (50.2)	15 (24.6)	728,485 (50.2)
Area of residence (RHA)						
IERHA	76 (5)	118 (6.3)	194 (5.7)	s	s	131,367 (9.1)
NHR	64 (4.2)	122 (6.5)	186 (5.5)	s	s	88,201 (6.1)
SH-SS	64 (4.2)	143 (7.7)	207 (6.1)	s	s	202,287 (13.9)
PMH	62 (4.1)	123 (6.6)	185 (5.4)	s	s	189,383 (13.1)
WRHA	1,250 (81.8)	1,335 (71.5)	2,585 (76.1)	834,383 (57.5)	46 (75.4)	834,429 (57.5)
Public Trustee/Not Found	12 (0.8)	26 (1.4)	38 (1.1)	4,948 (0.3)	0 (0)	4,948 (0.3)
Income quintile						
Q1-Lowest/Not Found	792 (51.8)	685 (36.7)	1,477 (43.5)	316,202 (21.8)	28 (45.9)	316,230 (21.8)
Q2	287 (18.8)	346 (18.5)	633 (18.6)	291,785 (20.1)	18 (29.5)	291,803 (20.1)
Q3	207 (13.5)	282 (15.1)	489 (14.4)	283,112 (19.5)	8 (13.1)	283,120 (19.5)
Q4	133 (8.7)	298 (16)	431 (12.7)	s	s	283,510 (19.5)
Q5-Highest	109 (7.1)	256 (13.7)	365 (10.8)	s	s	275,952 (19)
Health insurance coverage, years	10.4 ± 2.6	10.6 ± 2.5	10.5 ± 2.6	10.3 ± 2.8	9.2 ± 3.1	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.26. Description of Algorithm 26 (2+ physician claims in 3 years or 2+ antiretroviral prescriptions in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,518)	False Positives (n = 1,449)	Overall (N = 2,967)	True Negatives (n = 1,450,972)	False Negatives (n = 71)	Overall (N = 1,451,043)
Age at index date, years	36.9 ± 12.4	28.9 ± 18	33 ± 15.9	32.7 ± 23.3	28.5 ± 14.6	32.7 ± 23.3
Sex						
Male	1,002 (66)	661 (45.6)	1,663 (56)	722,296 (49.8)	51 (71.8)	722,347 (49.8)
Female	516 (34)	788 (54.4)	1,304 (44)	728,676 (50.2)	20 (28.2)	728,696 (50.2)
Area of residence (RHA)						
IERHA	75 (4.9)	87 (6)	162 (5.5)	131,393 (9.1)	6 (8.5)	131,399 (9.1)
NHR	64 (4.2)	100 (6.9)	164 (5.5)	s	s	88,223 (6.1)
SH-SS	63 (4.2)	89 (6.1)	152 (5.1)	s	s	202,342 (13.9)
PMH	62 (4.1)	101 (7)	163 (5.5)	s	s	189,405 (13.1)
WRHA	1,242 (81.8)	1,048 (72.3)	2,290 (77.2)	834,670 (57.5)	54 (76.1)	834,724 (57.5)
Public Trustee/Not Found	12 (0.8)	24 (1.7)	36 (1.2)	4,950 (0.3)	0 (0)	4,950 (0.3)
Income quintile						
Q1-Lowest/Not Found	785 (51.7)	539 (37.2)	1,324 (44.6)	316,348 (21.8)	35 (49.3)	316,383 (21.8)
Q2	285 (18.8)	264 (18.2)	549 (18.5)	291,867 (20.1)	20 (28.2)	291,887 (20.1)
Q3	206 (13.6)	204 (14.1)	410 (13.8)	283,190 (19.5)	9 (12.7)	283,199 (19.5)
Q4	133 (8.8)	235 (16.2)	368 (12.4)	s	s	283,573 (19.5)
Q5-Highest	109 (7.2)	207 (14.3)	316 (10.7)	s	s	276,001 (19)
Health insurance coverage, years	10.4 ± 2.6	10.5 ± 2.6	10.4 ± 2.6	10.3 ± 2.8	9.2 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.27. Description of Algorithm 27 (3+ physician claims in 3 years or 2+ antiretroviral prescriptions in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,507)	False Positives (n = 1,335)	Overall (N = 2,842)	True Negatives (n = 1,451,086)	False Negatives (n = 82)	Overall (N = 1,451,168)
Age at index date, years	37 ± 12.4	28.7 ± 17.9	33.1 ± 15.8	32.7 ± 23.3	27.9 ± 14.4	32.7 ± 23.3
Sex						
Male	994 (66)	612 (45.8)	1,606 (56.5)	722,345 (49.8)	59 (72)	722,404 (49.8)
Female	513 (34)	723 (54.2)	1,236 (43.5)	728,741 (50.2)	23 (28)	728,764 (50.2)
Area of residence (RHA)						
IERHA	74 (4.9)	76 (5.7)	150 (5.3)	131,404 (9.1)	7 (8.5)	131,411 (9.1)
NHR	64 (4.2)	92 (6.9)	156 (5.5)	s	s	88,231 (6.1)
SH-SS	61 (4)	79 (5.9)	140 (4.9)	202,348 (13.9)	6 (7.3)	202,354 (13.9)
PMH	62 (4.1)	95 (7.1)	157 (5.5)	s	s	189,411 (13.1)
WRHA	1,235 (82)	970 (72.7)	2,205 (77.6)	834,748 (57.5)	61 (74.4)	834,809 (57.5)
Public Trustee/Not Found	11 (0.7)	23 (1.7)	34 (1.2)	s	s	4,952 (0.3)
Income quintile						
Q1-Lowest/Not Found	779 (51.7)	492 (36.9)	1,271 (44.7)	316,395 (21.8)	41 (50)	316,436 (21.8)
Q2	284 (18.8)	248 (18.6)	532 (18.7)	291,883 (20.1)	21 (25.6)	291,904 (20.1)
Q3	204 (13.5)	184 (13.8)	388 (13.7)	283,210 (19.5)	11 (13.4)	283,221 (19.5)
Q4	133 (8.8)	222 (16.6)	355 (12.5)	s	s	283,586 (19.5)
Q5-Highest	107 (7.1)	189 (14.2)	296 (10.4)	s	s	276,021 (19)
Health insurance coverage, years	10.4 ± 2.6	10.5 ± 2.6	10.4 ± 2.6	10.3 ± 2.8	9.2 ± 3.1	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.28. Description of Algorithm 28 (1+ physician claims in 3 years or 1+ hospital discharges in 3 years or 2+ antiretroviral prescriptions in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,534)	False Positives (n = 1,943)	Overall (N = 3,477)	True Negatives (n = 1,450,478)	False Negatives (n = 55)	Overall (N = 1,450,533)
Age at index date, years	36.9 ± 12.5	29.3 ± 18.5	32.7 ± 16.6	32.7 ± 23.3	25.5 ± 13.1	32.7 ± 23.3
Sex						
Male	1,012 (66)	918 (47.2)	1,930 (55.5)	722,039 (49.8)	41 (74.5)	722,080 (49.8)
Female	522 (34)	1,025 (52.8)	1,547 (44.5)	728,439 (50.2)	14 (25.5)	728,453 (50.2)
Area of residence (RHA)						
IERHA	76 (5)	121 (6.2)	197 (5.7)	s	s	131,364 (9.1)
NHR	64 (4.2)	129 (6.6)	193 (5.6)	s	s	88,194 (6.1)
SH-SS	64 (4.2)	146 (7.5)	210 (6)	s	s	202,284 (13.9)
PMH	62 (4)	132 (6.8)	194 (5.6)	s	s	189,374 (13.1)
WRHA	1,256 (81.9)	1,388 (71.4)	2,644 (76)	834,330 (57.5)	40 (72.7)	834,370 (57.5)
Public Trustee/Not Found	12 (0.8)	27 (1.4)	39 (1.1)	4,947 (0.3)	0 (0)	4,947 (0.3)
Income quintile						
Q1-Lowest/Not Found	797 (52)	723 (37.2)	1,520 (43.7)	316,164 (21.8)	23 (41.8)	316,187 (21.8)
Q2	287 (18.7)	359 (18.5)	646 (18.6)	291,772 (20.1)	18 (32.7)	291,790 (20.1)
Q3	207 (13.5)	296 (15.2)	503 (14.5)	283,098 (19.5)	8 (14.5)	283,106 (19.5)
Q4	134 (8.7)	303 (15.6)	437 (12.6)	s	s	283,504 (19.5)
Q5-Highest	109 (7.1)	262 (13.5)	371 (10.7)	s	s	275,946 (19)
Health insurance coverage, years	10.3 ± 2.6	10.6 ± 2.5	10.5 ± 2.6	10.3 ± 2.8	9.5 ± 3	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.29. Description of Algorithm 29 (2+ physician claims in 3 years or 1+ hospital discharges in 3 years or 2+ antiretroviral prescriptions in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,526)	False Positives (n = 1,530)	Overall (N = 3,056)	True Negatives (n = 1,450,891)	False Negatives (n = 63)	Overall (N = 1,450,954)
Age at index date, years	36.9 ± 12.5	29.4 ± 17.9	33.2 ± 15.9	32.7 ± 23.3	25.9 ± 12.6	32.7 ± 23.3
Sex						
Male	1,009 (66.1)	710 (46.4)	1,719 (56.3)	722,247 (49.8)	44 (69.8)	722,291 (49.8)
Female	517 (33.9)	820 (53.6)	1,337 (43.8)	728,644 (50.2)	19 (30.2)	728,663 (50.2)
Area of residence (RHA)						
IERHA	75 (4.9)	90 (5.9)	165 (5.4)	131,390 (9.1)	6 (9.5)	131,396 (9.1)
NHR	64 (4.2)	107 (7)	171 (5.6)	s	s	88,216 (6.1)
SH-SS	64 (4.2)	92 (6)	156 (5.1)	s	s	202,338 (13.9)
PMH	62 (4.1)	111 (7.3)	173 (5.7)	s	s	189,395 (13.1)
WRHA	1,249 (81.8)	1,105 (72.2)	2,354 (77)	834,613 (57.5)	47 (74.6)	834,660 (57.5)
Public Trustee/Not Found	12 (0.8)	25 (1.6)	37 (1.2)	4,949 (0.3)	0 (0)	4,949 (0.3)
Income quintile						
Q1-Lowest/Not Found	792 (51.9)	580 (37.9)	1,372 (44.9)	316,307 (21.8)	28 (44.4)	316,335 (21.8)
Q2	285 (18.7)	278 (18.2)	563 (18.4)	291,853 (20.1)	20 (31.7)	291,873 (20.1)
Q3	206 (13.5)	218 (14.2)	424 (13.9)	283,176 (19.5)	9 (14.3)	283,185 (19.5)
Q4	134 (8.8)	241 (15.8)	375 (12.3)	s	s	283,566 (19.5)
Q5-Highest	109 (7.1)	213 (13.9)	322 (10.5)	s	s	275,995 (19)
Health insurance coverage, years	10.3 ± 2.6	10.5 ± 2.6	10.4 ± 2.6	10.3 ± 2.8	9.6 ± 2.9	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.

Table A.30. Description of Algorithm 30 (3+ physician claims in 3 years or 1+ hospital discharges in 3 years or 2+ antiretroviral prescriptions in 3 years) application

Characteristic	HIV Case			HIV Non-Case		
	True Positives (n = 1,518)	False Positives (n = 1,416)	Overall (N = 2,934)	True Negatives (n = 1,451,005)	False Negatives (n = 71)	Overall (N = 1,451,076)
Age at index date, years	37 ± 12.4	29.3 ± 17.9	33.3 ± 15.8	32.7 ± 23.3	25.4 ± 12.7	32.7 ± 23.3
Sex						
Male	1,004 (66.1)	661 (46.7)	1,665 (56.7)	722,296 (49.8)	49 (69)	722,345 (49.8)
Female	514 (33.9)	755 (53.3)	1,269 (43.3)	728,709 (50.2)	22 (31)	728,731 (50.2)
Area of residence (RHA)						
IERHA	74 (4.9)	79 (5.6)	153 (5.2)	131,401 (9.1)	7 (9.9)	131,408 (9.1)
NHR	64 (4.2)	99 (7)	163 (5.6)	s	s	88,224 (6.1)
SH-SS	63 (4.2)	82 (5.8)	145 (4.9)	s	s	202,349 (13.9)
PMH	62 (4.1)	105 (7.4)	167 (5.7)	s	s	189,401 (13.1)
WRHA	1,244 (81.9)	1,027 (72.5)	2,271 (77.4)	834,691 (57.5)	52 (73.2)	834,743 (57.5)
Public Trustee/Not Found	11 (0.7)	24 (1.7)	35 (1.2)	s	s	4,951 (0.3)
Income quintile						
Q1-Lowest/Not Found	788 (51.9)	533 (37.6)	1,321 (45)	316,354 (21.8)	32 (45.1)	316,386 (21.8)
Q2	284 (18.7)	262 (18.5)	546 (18.6)	291,869 (20.1)	21 (29.6)	291,890 (20.1)
Q3	204 (13.4)	198 (14)	402 (13.7)	283,196 (19.5)	11 (15.5)	283,207 (19.5)
Q4	134 (8.8)	228 (16.1)	362 (12.3)	s	s	283,579 (19.5)
Q5-Highest	108 (7.1)	195 (13.8)	303 (10.3)	s	s	276,014 (19)
Health insurance coverage, years	10.4 ± 2.6	10.5 ± 2.6	10.4 ± 2.6	10.3 ± 2.8	9.6 ± 2.9	10.3 ± 2.8

Notes:
Data are mean ± standard deviation or n (%)
IERHA = Interlake Eastern Regional Health Authority; NHR = Northern Health Region; SH-SS = Southern Health – Santé Sud; PMH = Prairie Mountain Health; WRHA = Winnipeg Regional Health Authority.