BMJ Open

Development of a computable phenotype to identify a transgender sample for health research purposes: A feasibility study in a large linked provincial healthcare administrative cohort in British Columbia, Canada

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-040928
Article Type:	Original research
Date Submitted by the Author:	26-May-2020
Complete List of Authors:	Rich, Ashleigh; The University of British Columbia, School of Population & Public Health; BC Centre for Excellence in HIV/AIDS, Poteat, Tonia; University of North Carolina at Chapel Hill, Department of Social Medicine Koehoorn, Mieke; University of British Columbia, School of Population and Public Health; Li, Jenny; BC Centre for Excellence in HIV/AIDS, Ye, Monica; BC Centre for Excellence in HIV/AIDS, Sereda, Paul; British Columbia Centre for Excellence in HIV/AIDS, Salway, Travis; Simon Fraser University Hogg, R; Simon Fraser University,
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, Sexual and gender disorders < PSYCHIATRY, HIV & AIDS < INFECTIOUS DISEASES

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

TITLE PAGE

Title: Development of a computable phenotype to identify a transgender sample for health research purposes: A feasibility study in a large linked provincial healthcare administrative cohort in British Columbia, Canada

Authors: Rich AJ[1,2], Poteat T[3], Koehoorn M[1], Li J[2], Ye M[2], Sereda, P[2], Salway T[4], Hogg RS[2,4]

Affiliations:
1. School of Population and Public Health, University of British Columbia, Vancouver, Canada
2. British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada
3. Department of Social Medicine, University of North Carolina- Chapel Hill, Chapel Hill, USA
4. Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada

Corresponding author:
Ashleigh J Rich
ajrich@mail.ubc.ca

2206 East Mall
Vancouver, BC
V6T 1Z3
Canada

Word count: 3842

ABSTRACT

Objectives: Innovative methods are needed for identification of transgender people in administrative records for health research purposes. This study investigated the feasibility of using transgender-specific healthcare utilization in a Canadian population-based, health records database to develop a computable phenotype (CP) and identify the proportion of transgender people within the HIV-positive population as a public health priority.

Design: The COAST cohort comprises a data linkage between two provincial data sources: The BC Centre for Excellence in HIV/AIDS Drug Treatment Program, which coordinates HIV treatment dispensation across BC; and Population Data BC, a provincial data repository holding individual, longitudinal data for all BC residents (1996-2013).

Setting: British Columbia, Canada.

Participants: COAST participants include 13 907 BC residents living with HIV (≥19 years of age) and a 10% random sample comparison group of the HIV-negative general population (514 952 individuals).

Primary and secondary outcome measures: Healthcare records were used to identify transgender people via a CP algorithm (diagnosis codes + androgen blocker/hormone prescriptions), to examine related diagnoses and prescription concordance, and to validate the CP using an independent provider-report transgender status measure. Demographics and chronic illness burden were also characterized for the transgender sample.

Results: The best-performing CP identified 137 HIV-negative and 51 HIV-positive transgender people (total 188). In validity analyses, the best-performing CP had low sensitivity (27.5%, 95%CI:17.8-39.8), high specificity (99.8%, 95%CI:99.6-99.8), low agreement using Kappa statistics (0.3, 95%CI:0.2-0.5), and moderate positive predictive value (43.2%, 95%CI:28.7-58.9). There was high concordance between exogenous-sex hormone use and transgender-specific diagnoses.

Conclusions: The development of a validated CP opens up new opportunities for identifying transgender people for inclusion in population-based health research using administrative health data, and offers the potential for much-needed and heretofore unavailable evidence on health status, including HIV status, and the healthcare use and needs of transgender people.

KEYWORDS: Transgender Persons, Health Services, Algorithms, Canada

ARTICLE SUMMARY

Strengths and limitations of this study:

- This study demonstrates the feasibility of developing and validating a computable phenotype for identification of a transgender sample, using a population-based representative source population and healthcare records.
- A major contribution of this study is the ascertainment of the population of transgender people living with HIV in the Canadian province of British Columbia, in a universal

healthcare setting, using a computable phenotype, and capacity to estimate the prevalence of transgender status among the population living with HIV in the province.

- Development of a validated transgender computable phenotype algorithm lays the foundation for future investigation of transgender-specific research questions related to general and HIV-specific healthcare use and health outcomes for this key population.

INTRODUCTION

Limited data on transgender people

Transgender people are often overlooked within epidemiological research and population health surveillance due to small sample size, limited research designs, and other institutional and methodological erasures.[1–3] A 2017 review of Medline-indexed literature from 1950 to 2016 found 2405 published articles including transgender people, with almost half published in the last decade.[4] A 2008 United States (US)-based meta-analysis of HIV prevalence among transgender populations found 24 studies of transgender women, and five additional studies of transgender men,[5] though an updated review found 43 primary studies on transgender women and 15 on transgender men published between 2006 to 2017.[6] Despite this recent increase in transgender health research in general and for HIV specifically, much of the literature has focused on transgender-specific care, mental health and HIV/sexual health,[7,8] leaving the population understudied, in particular in the broader areas of physical health and healthcare utilization.

The erasures or exclusions of transgender persons in health studies may be explained, in part, by methodological challenges. Specific to electronic health record (EHR) data, a 2017 report identified only one transgender person among 38,5820 cancer cases in a Minnesota cancer registry,[9] clearly an undercount given that 0.4% of the US general population and 0.6% (95% credibility intervals: 0.5%-0.7%) of the Minnesota population is estimated to be transgender.[10,11] This highlights the need for improved gender ascertainment and transgender inclusion in research relying on patient records and administrative data. The establishment of best practices for measuring transgender status in survey research, such as the two-step method (measuring sex assigned at birth and current gender identity), points to a way forward for transgender-inclusive population health research.[12,13] However, innovative research methods are needed to identify transgender people in studies that rely on existing data sources (in particular EHR) and that optimize the use of transgender respondents' data in non-transgender specific research.

Methodological limitations in transgender health research

Previous research in transgender health largely comprises cross-sectional studies, case reports, and qualitative or observational research.[7] Much consists of clinic- or venue-based convenience samples or lack comparison groups.[7,8] The literature is further characterized by inconsistent transgender status measurement,[14] small sample sizes, and focus on the United States (US).[8] In response, researchers have called for advancing transgender health research methods - namely ascertainment of high-quality samples via systematic approaches - including for general population-based and health systems-based studies.[15]

Computable phenotypes for transgender health research

One opportunity for the advancement of transgender health research methods is the emerging use of computable phenotypes (CPs), also called natural language processing algorithms[16] or case algorithms, to identify transgender samples in healthcare utilization data. A computable phenotype is a clinical feature, condition, or set of characteristics that can be determined directly from EHR and other ancillary health care data systems (e.g., disease registries, insurance claims data) data.[17] CPs are developed using a combination of data elements (e.g., sociodemographic variables, clinical diagnoses) and value sets (i.e., the selection of a set of relevant values for each

data element). Development of CPs using standardized methods and definitions enables identification and inclusion of transgender persons in research, as well as replication of analyses across data sources, healthcare organizations/sites and studies. CPs have application in clinical care, surveillance, and health research.

Recently, CP and other EHR-based algorithm methods have been applied in a number of settings primarily in the US to identify transgender samples for health research.[14] Specifically, the STRONG study identified a transgender cohort (n = 6,456) using EHR data from Kaiser Permanente health plan members in California and Georgia, for investigation of general and transgender-specific health outcomes.[18] Blosnich et al identified 3,177 people with a "gender identity disorder" diagnosis among military veterans accessing care through the US Veterans Health Administration healthcare system,[19] for examination of mental health and other outcomes. Researchers with the US Centers for Medicare & Medicaid Services identified 4,098 transgender beneficiaries using national Medicare claims data,[20] and researchers at Vanderbilt University identified 234 transgender patients in their university clinic EHR data.[16] While these cohorts represent important opportunities for advancement of transgender health research, these methods have yet to be applied widely outside the US context. This is particularly important as different jurisdictions may vary in medical billing and coding practices, healthcare system patient populations, and representativeness of the general population. Specifically, in Canada, healthcare is delivered through a provincially administered universal healthcare system. As such, research using EHR provides an opportunity to develop methods for population-based, representative estimates of transgender populations within the Canadian context. Coupled with the current absence of gender ascertainment measures in population-based routinely collected data (e.g., census, national government health surveys, etc.) in Canada and many other jurisdictions, this remains an evidence need.

Summary of study rationale

This study investigated the application of emerging transgender health research methods, specifically CPs, in a Canadian context for the first time, testing the feasibility of identification of a transgender sample using EHR data from a provincial healthcare administrative data-linked cohort.

METHODS

Data Sources and Participants

The Comparative Outcomes and Service Utilization Trends Study (COAST)

COAST is a population-based cohort study focused on health services utilization research questions among all people known to be living with HIV (PLWH) in the province of British Columbia (BC) and a 10% random sample comparison group of the HIV-negative general population.[21] The COAST cohort comprises individual-level, longitudinal data from PLWH who have ever accessed HIV treatment in BC between 1996 and 2013, provided by Population Data BC (PopDataBC)[21] via data linkage between two provincial data sources, by personal health number: the Drug Treatment Program (DTP) [22] and the Ministry of Health. PopDataBC provides infrastructure for access to, and linkage of, longitudinal and individual-level administrative health data for all BC residents.[23]. The HIV-negative general population cohort was drawn randomly from the Ministry of Health registry data by PopDataBC. The COAST study has received approval from the University of British Columbia/Providence Health Care

Research Ethics Board (#H09-02905) and Simon Fraser University Office of Research Ethics (#2013 s0566). The study complies with the BC Freedom of Information and Protection of Privacy Act (FIPPA) and did not require informed consent as it is conducted using retrospective administrative and anonymized data for research and statistical purposes only.

Drug Treatment Program

In BC, antiretroviral therapy (ART) is provided to PLWH at no cost to the patient, and distributed through the DTP.[22] The DTP contributed a provider-reported measure of transgender status for COAST.

Ministry of Health

Ministry of Health data available via COAST included insured medical service billing records for outpatient visits,[24,25] hospital (in-patient) visits,[26] prescription medications,[27,28] and vital statistics.[29]

Measures & Analyses

Transgender computable phenotypes

Identification of transgender cases was tested in COAST using International Classification of Disease (ICD) codes (9th and 10th editions) and exogenous sex hormone prescription use. Transgender-specific diagnoses in medical and hospital billing records included the ICD-9 codes 302.5 Trans-sexualism with unspecified history, 302.51 Trans-sexualism with asexual history, 302.52 Trans-sexualism with homosexual history, 302.53 Trans-sexualism with heterosexual history, 302.6 Gender Identity Disorder in children, 302.85 Gender Identity Disorder in adolescents or adults; and ICD-10 codes F64.0 Gender Identity Disorder of childhood, F64.2 Gender Identity Disorder of childhood, F64.8 Other Gender Identity Disorder, and F64.9 Gender Identity Disorder unspecified. The full list of androgen blockers and exogenous sex hormone prescriptions included in analyses is available in the supplementary material.

Concordance

To assess face validity and utility of diagnosis and prescription data over time in CP development, concordance analyses evaluated the presence of at least one included diagnosis and prescription during the COAST study follow-up period with the presence of at least one included diagnosis and prescription in the last study year. Concordance was assessed between transgender-specific diagnoses, exogenous sex hormone and androgen blocker prescriptions, and non-transgender specific diagnoses (ICD-9 259.9 Unspecified Endocrine Disorder and ICD-10 E34.9 Endocrine Disorder, Unspecified (see supplementary material)). Endocrine disorder diagnosis codes are sometimes preferred by medical providers treating transgender people in response to historic exclusions of transgender-specific care from insurance coverage and to combat the stigma of transgender-specific diagnosis codes that have historically been classified as psychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM).[30] Exogenous sex hormone use, while common in transgender populations,[5,31] is not transgender-specific. Cisgender populations also use androgen blocker and sex hormone prescriptions (e.g. estrogen to treat menopausal symptoms in cisgender women, spironolactone is used for hypertension), thus exogenous sex hormone and androgen blocker prescription use

cannot independently identify transgender people. At the same time, not all transgender people use hormones and some access via non-medical sources.[32,33]

Validation

In British Columbia, transgender status is collected in the DTP via a provider-reported sex variable ("Male", "Male to Female", "Female to Male", or "Female"). Patients reported as either "Male to Female" or "Female to Male" were classified as transgender. The provider-reported transgender measure, available for the HIV positive cohort only, was used as a 'gold standard' for CP validation. Sensitivity, specificity, positive predictive value and kappa statistics with corresponding 95% confidence intervals (CI) were calculated for identifying transgender people via the CPs.

Demographics and chronic conditions

To further assess face validity of the transgender CP for future health research, descriptive statistics were calculated for the transgender sample using the COAST study key sociodemographic and health variables, specifically laboratory confirmed HIV serostatus (HIV-positive/HIV-negative), baseline age, patient's Health Authority (five provincial regions for the administration of health services that include large urban centres, suburban regions, and rural/remote areas), and chronic illness burden based on standardized case definitions from the BC Ministry of Health [34] and the BC Cancer Agency.[35]

RESULTS

The total COAST cohort included 528 859 people, of which 514 952 were HIV-negative (10% general population random sample) and 13 907 were PLWH (Figure 1).

[Figure 1 here]

Concordance

Of the 237 people who had ever had a transgender-specific diagnosis during the study period, 19.4% also had a recent diagnosis in the last follow-up year (Table 2). None had an unspecified endocrine disorder diagnosis at any time, thus this diagnosis was excluded from all CPs. Of the 237, 79.3% had an exogenous sex hormone or androgen blocker prescription at least once during the study period and 46.4% had one in the last year.

Table 1. Concordance analyses for diagnoses and hormone measures

	N	%
≥ Transgender ICD- ever	237	100
≥ Transgender ICD- recent	46	19.4
Unspecified Endocrine Disorder use- ever	0	0.0
Unspecified Endocrine Disorder use- recent	0	0.0
≥ Hormone/blocker use- ever	188	79.3
≥ Hormone/blocker use-recent	110	46.4

Validation

While no one CP consistently performed well across all validation metrics, the CP with the best overall performance across test statistics was based on having received at least one transgender-

specific diagnosis and at least one androgen blocker/exogenous sex hormone prescription over the study follow-up period (Table 1). This CP had high specificity (99.8%, 95% CI: 99.6-99.8), low sensitivity (27.5%, 95% CI: 17.8-39.8), low to moderate Kappa coefficients (0.3, 95% CI: 0.2-0.5) and moderate positive predictive values (43.2%, 95% CI: 28.7-58.9).

Table 2. Validation measures of transgender computable phenotype (CP) with provider-report transgender status measures, in COAST HIV-positive cohort

CP	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	Kappa (95% CI)
≥ 1 transgender ICD- ever	27.5 (17.8, 39.8)	99.7 (99.6, 99.8)	40.4 (26.7, 55.7)	0.3 (0.2, 0.4)
≥ 1 transgender ICD- recent	8.7 (3.6, 18.6)	100.0 (99.9, 100.0)	85.7 (42.0, 99.4)	0.2 (0.1, 0.3)
≥ 1 transgender ICD AND ≥ 1 hormone/blocker use- ever	27.5 (17.8, 39.8)	99.8 (99.6, 99.8)	43.2 (28.7, 58.9)	0.3 (0.2, 0.5)
≥ 1 transgender ICD AND ≥ 1 hormone/blocker use- recent	7.3 (2.7, 16.8)	100.0 (99.9, 100.0)	83.3 (36.5, 99.1)	0.1 (0.0, 0.2)

Transgender phenotype

Applying the best-performing CP, 137 HIV-negative people and 51 HIV-positive people (188 total) were identified as transgender in the respective COAST cohorts (Figure 1).

Demographics and chronic conditions

Demographic characteristics and chronic conditions for the 188 transgender people identified via the best-performing CP are presented in Figures 2 to 4. Transgender people were geographical located throughout BC health regions. The Vancouver Coastal Health Authority region, which includes the largest municipal area in BC, had the highest concentration of transgender people (44.2%) while the Northern Health Authority region - a predominantly rural and remote area of the province - had the lowest (1.6%).[36] The HIV-positive group had a higher median age than the HIV-negative group (35 [Q1, Q3: 30,42] and 30 [Q1, Q3: 19,42], respectively). For the HIV-negative sample, the largest proportion of transgender people were aged 19 to 29 years (44.5%) and the smallest proportion aged 55 years and older (<3%). For the HIV-positive sample, the largest proportion were aged 30 to 34 years (25.5%) and the smallest proportion aged 55 years and older (<2%).

[Figures 2 and 3 here]

Overall, HIV-positive transgender people had a higher prevalence of at least one chronic condition (other than HIV) compared to HIV-negative transgender people (88.2% versus 85.4%, respectively), and of two or more chronic conditions (76.5% versus 52.6%, respectively). Specific chronic disease differences between transgender people living with and without HIV were most notable for a higher prevalence among the HIV-positive cohort of cardiovascular

disease, chronic kidney disease, osteoarthritis, schizophrenia and personality disorders, and chronic liver disease, but a lower prevalence for hypertension.

[Figure 4 here]

DISCUSSION

This study demonstrates the feasibility of identification of a sample of transgender people in a large linked provincial healthcare administrative database, using a CP based on prescriptions and diagnoses. Among a growing number of studies using EHR and CP methods to identify transgender samples for health research purposes, this is the first to do so in Canada., to independently validate the CP using a 'gold standard' of provider-reported transgender status, and the only to use population-based data.

Concordance

There was high concordance between transgender-specific diagnoses and exogenous sex hormone or androgen blocker prescription use in this study. That nearly half of those with at least one transgender-specific diagnosis had been dispensed hormones or blockers in the past year is consistent with findings from US and Canadian studies (48.9% and 43.0%, respectively)[20,32,33] - suggesting face validity for the current CP.

CP development and validation

The best-performing CP overall successfully identified cisgender people who were truly cisgender (specificity) and correctly identified transgender people who were truly transgender (0.2% false positive rate, results not shown). However, the selected CP had relatively low sensitivity, missing approximately 72.5% of 'true' transgender people in COAST. Though a relatively small proportion of the 'true' transgender sample was identified in this study, the impact on future analyses comparing health outcomes for transgender and cisgender groups is likely negligible, as even the large proportion of 'true' transgender people misclassified as cisgender (approximate n=496) is a very small proportion of the total COAST sample. At worst, this misclassification would bias results related to disparities between transgender and cisgender health toward the null, producing a conservative attenuated effect. Further, as discussed below, gender identity classification will likely greatly improve as transgender care shifts further into the fee-for-service system in BC. As in other Canadian administrative data studies, low sensitivity may be explained in part by provider and system billing preferences using 3-digit ICD diagnosis coding instead of the more specific 4-digit coding, and inconsistencies in the BC billing management system.[37]

The limited agreement between the CP and provider-report transgender status may be due to the widely varying transgender status prevalence depending on study design and ascertainment measures used.[14] In the BC context, the CP and the DTP measures are assessing transgender status in different ways and for different purposes. In the DTP, transgender status is ascertained in the context of HIV diagnosis and ART prescribing, during which demographics and HIV transmission risk factors are recorded. This differs from recording diagnoses in EHR for those accessing transgender-specific care as utilized in the CP. Ultimately, a single CP may not be sufficient for all intended purposes and the best applicable CP (using different types of

diagnoses, prescriptions or procedures) may differ depending on the intended healthcare, health research, or health policy application.[17]

There is limited literature on EHR-based studies with the ability to validate an administrative transgender measure using a 'gold standard' comparison measure.[16] The two previous studies that have developed and validated algorithms to identify transgender individuals have both been conducted in non-representative samples in the US, one using Medicare data[38] and one in a university medical center.[16] Similar to the current study, the Medicare study found high specificity when comparing an EHR-based and a two-step survey-based transgender measure. However, the Medicare study found that the EHR measure performed consistently well with high sensitivity and a high Kappa statistic, unlike in the current study. Using chart review as the 'gold standard' for comparison of transgender status, Ehrenfeld et al. found a low false positive rate for their best-performing algorithm (3%), though not as low as the false positive rate in the current study. The overall high levels of agreement for transgender measures in the two previous studies is likely a function of the lack of independence between the 'gold standard' and the CP or algorithm measures. Specifically, only those classified as transgender in the Medicare EHR data were offered survey participation to complete the two-step 'gold standard' survey measure, and only those cases identified as transgender in the university clinic EHR were included in chart review. Thus, previous studies could assess agreement between the two measures, but not robustly validate either. In the current study, the DTP provider-based transgender status measure is independent and thus could be used for robust CP validation.

Transgender status prevalence & ascertainment

Based on a recent meta-analysis of transgender status prevalence in population-based probability samples,[10] it was expected that an effective CP would identify 0.4% of the general population as transgender, or approximately 54 of the HIV-positive COAST cohort (n=13 907) and 3,098 of the HIV-negative cohort (n=516 340). Consistent with expectations, the best-performing CP identified 51 PLWH as transgender, equivalent to a transgender status prevalence of 0.4% among PLWH. Contrary to expectations, the best-performing CP identified less than 5% of the number of transgender persons expected in the HIV-negative cohort. This is likely a result of a number of factors including the limitation of CPs to the subset of a population accessing care as noted, and the result of most transgender people in BC receiving care currently outside the main fee-for-service healthcare delivery system. However, it is also consistent with the undercount of transgender populations using diagnostic criteria compared to other methods of ascertainment demonstrated in other studies.[14]

Using the broadest CP algorithm (any transgender-specific diagnosis ever, n=56) and those identified by provider-report together (total n=106), the total transgender PLWH sample would represent as high as 0.88% (range: 0.73-1.1%) of the prevalent HIV infections in BC in 2014.[39] This overrepresentation of transgender people among PLWH is consistent with evidence of a disproportionate HIV burden for transgender populations globally,[5,40,41] as well as in line with the only other available data on the proportion of PLWH who are transgender, from US national surveillance data (2012 data: 1.1%, 95%CI: 0.8-1.4).[42]

Demographics and chronic conditions

Despite moderate to low performance by some validation metrics, particularly low sensitivity, the CP was able to detect meaningful results in the characterization of demographics and chronic condition burden for the transgender sample - supporting CP face validity. The population density and age distribution by HIV-status of transgender people in this study is largely consistent with general population patterns, as well as the larger COAST cohort.[21,36] The overall higher burden of chronic illness for transgender people living with HIV versus without HIV in this study is consistent with elevated chronic illness risk and morbidity among non-transgender PLWH.[43] This higher chronic disease burden is linked to HIV disease processes and related inflammatory immune response.[44] While a small but growing number of studies have begun to investigate the chronic illness burden for transgender populations in other industrialized settings,[16,19,45–47] including using EHR data, findings vary widely due to differences in sampling, study design, setting and measurement.

Limitations

Findings from this study should be interpreted in the context of a few key limitations. CPs are by design only applicable to people accessing healthcare services, often motivated by illness and aided by the ability to access care. As such, this study is limited to those transgender people accessing medical transition care in BC and may only represent 24% to 47% of the total transgender population.[33] This study was also limited by the inability to validate the transgender CP among the HIV-negative COAST cohort, as a 'gold standard' provider-based transgender measure was only available for the HIV-positive cohort. It is possible that the transgender CPs would perform differently in populations living without HIV, particularly as healthcare contact is higher among populations living with HIV. Additionally, this study should be considered in light of the context in which it was conducted, an environment in which transgender healthcare delivery in BC is currently shifting from specialized care settings to the main primary care fee-for-service settings. Given that COAST only includes fee-for-service data, this study was limited by the inability to capture transgender people who access transgender care outside the fee-for-service system. However, fortunately, as the shift to the fee-for-service system occurs, transgender ascertainment via CPs in BC will likely improve. The administrative data used in this study may also be susceptible to coding error (and coding biases/practices) across conditions and settings,[48] potentially introducing misclassification bias in terms of transgender ascertainment. Finally, chronic condition prevalence data reported in this study should be interpreted with caution, given potential selection bias by serostatus in the COAST cohort; though any such bias likely resulted in conservative estimates of difference by serostatus in this analysis.

CONCLUSION

This study makes a number of important contributions to the literature on innovative methods in transgender health. Major contributions include development and validation of a transgender CP, using a population-based representative source population, in the Canadian context. Another strength is the approximately complete ascertainment of the population of transgender PLWH in BC, and capacity to estimate transgender status prevalence among PLWH. In a current funding environment of limited support for longitudinal transgender health studies in the US and none to date in Canada, this study and the methods employed offer an efficient, replicable and cost-effective way forward in creating electronic cohorts for advancing transgender health research.[15] Moreover, the recent rollback of sexual orientation and gender identity data

collection and legal changes in insurance coverage of transgender healthcare in the US potentiate decline in accurate claims coding for gender-affirming care.[30] This highlights the utility of work in this area from other jurisdictions, particularly those with transgender-inclusive universal healthcare systems such as Canada.

Future research should build upon the methods developed in this study and explore complimentary approaches for gender identity ascertainment in administrative and EHR data, such as machine learning approaches, as have been used to develop algorithms based on healthcare utilization data in other research areas. Finally, the current study lays the foundation for future work with the ability to study transgender health and healthcare use patterns over time, with linkage to laboratory data, as well as inclusion of appropriate comparison groups.[15,49]

ACKNOWLEDGEMENTS & DISCLAIMER
The authors thank the COAST study participants, the BC Centre for Excellence in HIV/AIDS, the BC Ministry of Health, BC Vital Statistics Agency, PharmaNet and the institutional data stewards for granting access to the data, and Population Data BC for facilitating the data linkage process. In addition, we would like to thank the COAST core team members and other support staff at these institutions for their administrative assistance with the data access and preparation. All inferences, opinions, and conclusions drawn in this paper are those of the authors, and do not reflect the opinions or policies of the Data Steward(s). There are no conflicts of interest to declare.

COMPETING INTERESTS
None declared.

FUNDING
This work was supported by the Canadian Institutes of Health Research, through an Operation Grant [grant number 130419], a Foundation Award to RSH [grant number 143342], and a Doctoral Research Award to AJR [grant number 152382] and support from the British Columbia Centre for Excellence in HIV/AIDS. The DTP receives funding from the provincial government of British Columbia (PharmaCare). The funders had no role in the study design, analysis, interpretation of the data, drafting of the manuscript or in the decision to submit for publication.

CONTRIBUTORS
AJR and RSH conceived of and designed the study. RSH acquired study data and funding. JL, MY and PS conducted data analysis, in consultation with AJR. AJR drafted the manuscript, with commentary on drafts by TP, MK, and TS, and approval of final version by all co-authors.

PATIENT CONSENT FOR PUBLICATION
Not required.

DATA SHARING STATEMENT

The data used for this study are held by the BC Centre for Excellence in HIV/AIDS under the authority of the BC Ministry of Health; as they contain confidential patient health records including HIV serostatus, data are cannot be made available to other parties.

FIGURES LEGENDS

Figure 1. Total transgender sample identified using a computable phenotype with electronic health records

Figure 2. Geographic distribution of transgender people across province, by health authority*
**% of transgender individuals with known health authority (n=182)*

Figure 3. Age distribution of transgender sample, by HIV serostatus

Figure 4. Co-morbidities among transgender sample, by HIV serostatus

REFERENCES

1. Bauer GR. Making sure everyone counts: Considerations for inclusion, identification, and analysis of transgender and transsexual participants in health surveys. What A Difference Sex And Gender Make, Gender Sex Health Research Casebook. Vancouver, Canada: Canadian Institutes of Health Research (CIHR), Institute for Gender and Health (IGH); 2012. p. 59–67.

2. Bauer GR, Hammond R, Travers R, Kaay M, Hohenadel KM, Boyce M. "I don't think this is theoretical; this is our lives": How erasure impacts health care for transgender people. JANAC. 2009;20:348–361.

3. Poteat T, German D, Flynn C. The conflation of gender and sex: Gaps and opportunities in HIV data among transgender women and MSM. Glob Public Health. 2016;11:835–848.

4. Wanta JW, Unger CA. Review of the transgender literature: Where do we go from here? Transgender Health. 2017;2:119–128.

5. Herbst JH, Jacobs ED, Finlayson TJ, McKleroy VS, Neumann MS, Crepaz N, et al. Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a systematic review. AIDS Behav. 2008;12:1–17.

6. Becasen JS, Denard CL, Mullins MM, Higa DH, Sipe TA. Estimating the prevalence of HIV and sexual behaviors among the US transgender population: A systematic review and meta-analysis, 2006–2017. Am J Public Health. 2019;109:e1–8.

7. Marshall Z, Welch V, Thomas J, Brunger F, Swab M, Shemilt I, et al. Documenting research with transgender and gender diverse people: protocol for an evidence map and thematic analysis. Syst Rev. 2017;6:35.

8. Reisner SL, Poteat T, Keatley J, Cabral M, Mothopeng T, Dunham E, et al. Global health burden and needs of transgender populations: a review. The Lancet. 2016;388:412–436.

9. Cathcart-Rake EJ, Lightner DJ, Quevedo FJ, Ertz DC, Jatoi A. Cancer in transgender patients: One case in 385,820 is indicative of a paucity of data. J Oncol Pract. 2017;JOP2017027714.

10. Meerwijk EL, Sevelius JM. Transgender population size in the United States: A meta-regression of population-based probability samples. Am J Public Health. 2017;107:e1–e8.

11. Flores AR, Herman JL, Gates GJ, Brown TNT. How Many Adults Identify as Transgender in the United States? [Internet]. Los Angeles, CA: Williams Institute; 2016 Jun p. 13. Available from: https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf

12. The GenIUSS Group. Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys. Los Angeles, CA: The Williams Institute; 2014.

13. Bauer GR, Braimoh J, Scheim AI, Dharma C. Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. PLOS ONE. 2017;12:e0178043.

14. Collin L, Reisner SL, Tangpricha V, Goodman M. Prevalence of transgender depends on the "case" definition: A systematic review. J Sex Med. 2016;13:613–626.

15. Reisner SL, Deutsch MB, Bhasin S, Bockting W, Brown GR, Feldman J, et al. Advancing methods for US transgender health research. Curr Opin Endocrinol Diabetes Obes. 2016;23:198–207.

16. Ehrenfeld JM, Gottlieb KG, Beach LB, Monahan SE, Fabbri D. Development of a Natural Language Processing Algorithm to Identify and Evaluate Transgender Patients in Electronic Health Record System. Ethn Dis. 2019;29:441–50.

17. Richesson R, Smerek M, Rusincovitch S, Zozus MN, Chaudhuri PS, Hammond WE, et al. Electronic Health Records-Based Phenotyping. In: Uhlenbrauck G, editor. NIH Collab Living Textb Pragmatic Clin Trials. Bethesda, Maryland; 2014.

18. Quinn VP, Nash R, Hunkeler E, Contreras R, Cromwell L, Becerra-Culqui TA, et al. Cohort profile: Study of Transition, Outcomes and Gender (STRONG) to assess health status of transgender people. BMJ Open. 2017;7:e018121.

19. Blosnich JR, Brown GR, Shipherd Phd JC, Kauth M, Piegari RI, Bossarte RM. Prevalence of gender identity disorder and suicide risk among transgender veterans utilizing veterans health administration care. Am J Public Health. 2013;103:e27–32.

20. Proctor K, Haffer SC, Ewald E, Hodge C, James CV. Identifying the transgender population in the Medicare program. Transgender Health. 2016;1:250–265.

21. Eyawo O, Hull MW, Salters K, Samji H, Cescon A, Sereda P, et al. Cohort profile: the Comparative Outcomes And Service Utilization Trends (COAST) Study among people living with and without HIV in British Columbia, Canada. BMJ Open. 2018;8:e019115.

22. Drug Treatment Program. BC Centre for Excellence. HIVAIDS. 2010 [cited 2019 Sep 27]. Available from: http://www.cfenet.ubc.ca/drug-treatment-program

23. Population Data BC. About PopData [Internet]. Population Data BC. 2017 [cited 2019 Sep 27]. Available from: https://www.popdata.bc.ca/aboutus

24. British Columbia Ministry of Health (2014). Medical Services Plan (MSP) Payment Information File. Population Data BC. Data Extract. MOH (2014). http://www.popdata.bc.ca/data

25. British Columbia Ministry of Health (2014). Consolidation File (MSP Registration & Premium Billing). Population Data BC. Data Extract. MOH (2014). http://www.popdata.bc.ca/data

26. Canadian Institute for Health Information (2014). Discharge Abstract Database (Hospital Separations). Population Data BC. Data Extract. MOH (2014). http://www.popdata.bc.ca/data

27. British Columbia Ministry of Health (2014). PharmaCare. Population Data BC. Data Extract. MOH (2014). http://www.popdata.bc.ca/data

28. British Columbia Ministry of Health (2014). PharmaNet. Population Data BC. Data Extract. Data Stewardship Committee (2014). http://www.popdata.bc.ca/data

29. British Columbia Vital Statistics Agency (2014). Vital Statistics Deaths. Population Data BC. Data Extract. Vital Statistics Agency (2014). http://www.popdata.bc.ca

30. Arnold JD, Nelson AE, Loubier EM. Trends in Insurance Coverage for Gender-Affirming Surgeries. JAMA Surg. 2018;153:972.

31. Radix A, Sevelius J, Deutsch MB. Transgender women, hormonal therapy and HIV treatment: a comprehensive review of the literature and recommendations for best practices. J Int AIDS Soc. 2016;19:20810.

32. Rotondi NK, Bauer GR, Scanlon K, Kaay M, Travers R, Travers A. Nonprescribed hormone use and self-performed surgeries: "do-it-yourself" transitions in transgender communities in Ontario, Canada. Am J Public Health. 2013;103:1830–1836.

33. Scheim AI, Bauer GR. Sex and gender diversity among transgender persons in Ontario, Canada: results from a respondent-driven sampling survey. J Sex Res. 2015;52:1–14.

34. Chronic Disease Information Working Group. BC Chronic Disease and Selected Procedure Case Definitions [Internet]. British Columbia, Canada: BC Ministry of Health; 2019 Apr p. 1–132. Available from: https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf

35. British Columbia Cancer Agency (2014). BC Cancer Registry Data. Population Data BC. BC Cancer (2014). http://www.popdata.bc.ca/data

36. BC Stats. Sub-provincial Population Estimates [Internet]. 2018 [cited 2019 Feb 1]. Available from: file://localhost/Users/AJRich/Zotero/storage/5VM2LBUN/PopulationEstimates.html

37. Hu W. Diagnostic Codes in MSP Claim Data. Program Monitoring and Information Management Branch, Resource Management Division, Medical Services Plan; 1996 p. 39.

38. Gerth J, Becerra-Culqui T, Bradlyn A, Getahun D, Hunkeler EM, Lash TL, et al. Agreement between medical records and self-reports: Implications for transgender health research. Rev Endocr Metab Disord. 2018;19:263–269.

39. BC Centre for Disease Control. HIV in British Columbia: Annual Surveillance Report 2015. Vancouver, Canada: BC Centre for Disease Control; 2017.

40. Poteat T, Scheim A, Xavier J, Reisner S, Baral S. Global epidemiology of HIV infection and related syndemics affecting transgender people. J Acquir Immune Defic Syndr. 2016;72 Suppl 3:S210–9.

41. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV in transgender women: a systematic review and meta-analysis. Lancet Infect Dis. 2013;13:214–222.

42. Centers for Disease Control and Prevention. Behavioral and Clinical Characteristics of Persons Receiving Medical Care for HIV Infection—Medical Monitoring Project, United States, 2012 [Internet]. Washington, D.C.: Centers for Disease Control; 2015 Aug p. 42. Report No.: 12. Available from: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf

43. Grabovac I, Veronese N, Stefanac S, Haider S, Jackson SE, Koyanagi A, et al. Human Immunodeficiency Virus Infection and Diverse Physical Health Outcomes: An Umbrella Review of Meta-analyses of Observational Studies. Clin Infect Dis. 2019;ciz539.

44. Nasi M, De Biasi S, Gibellini L, Bianchini E, Pecorini S, Bacca V, et al. Ageing and inflammation in patients with HIV infection: Ageing, inflammation and HIV infection. Clin Exp Immunol. 2017;187:44–52.

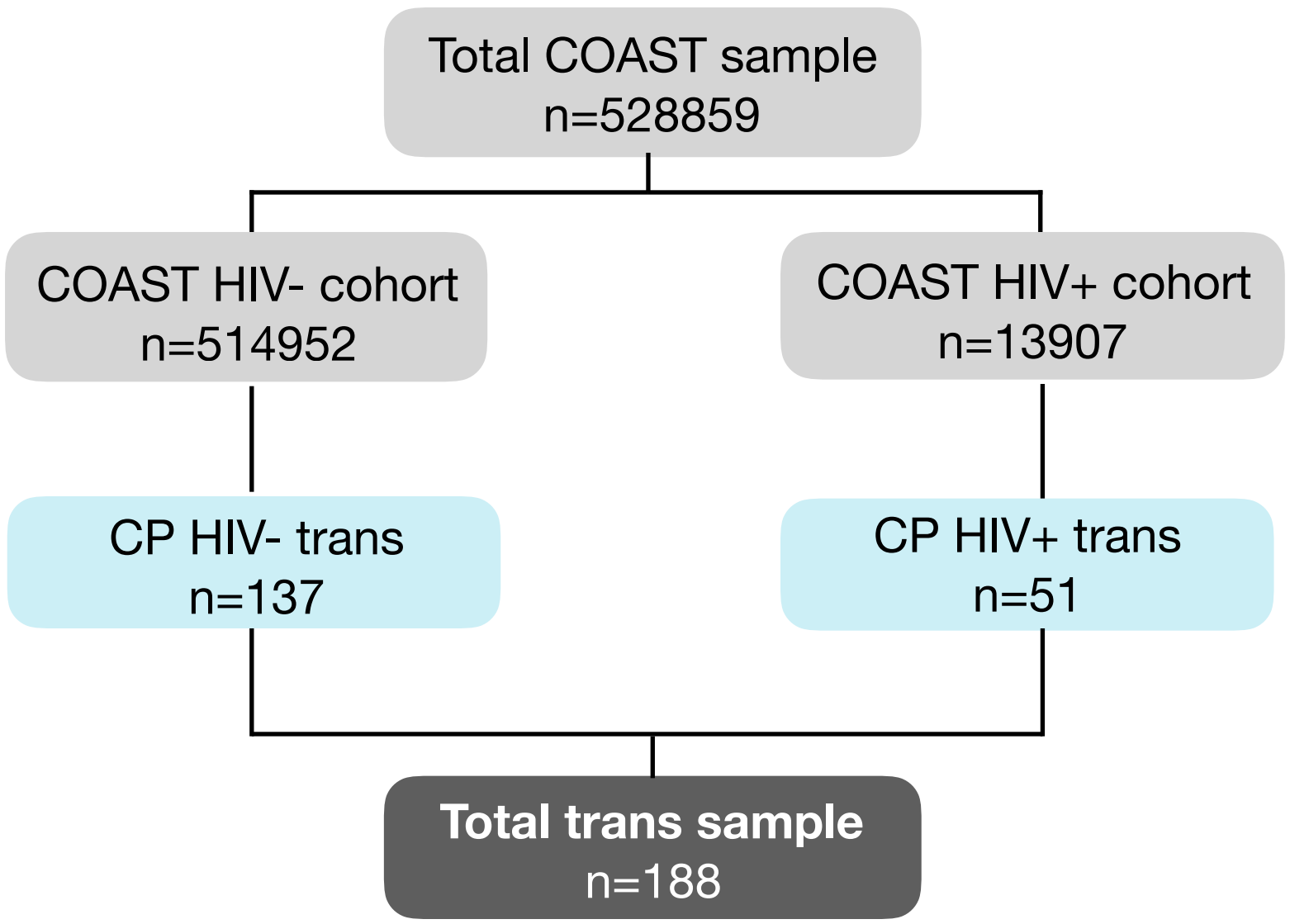
45. Judge C., O'Donovan C., Callaghan G., Gaoatswe G., O'Shea D. Gender dysphoria - prevalence and co-morbidities in an Irish adult population. Front Endocrinol. 2014;5:87.

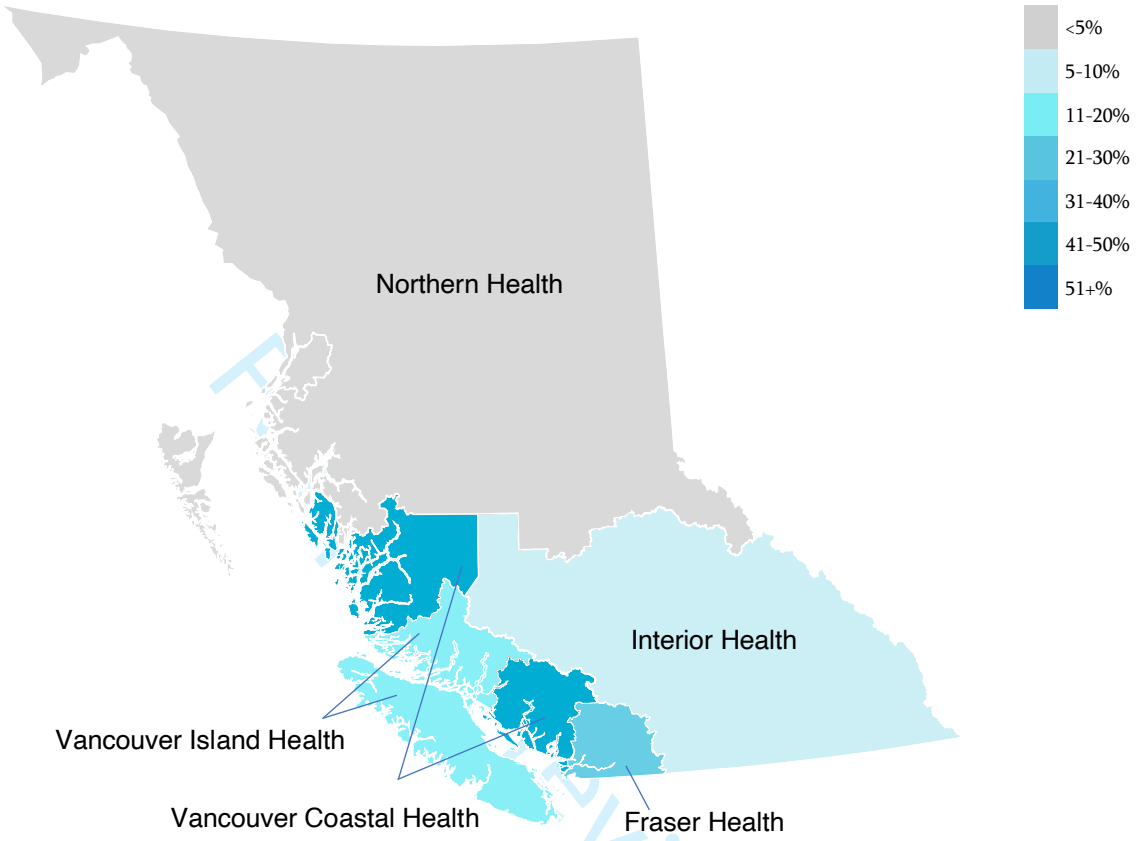
46. Cheung AS, Ooi O, Leemaqz S, Cundill P, Silberstein N, Bretherton I, et al. Sociodemographic and clinical characteristics of transgender adults in Australia. Transgender Health. 2018;3:229–38.

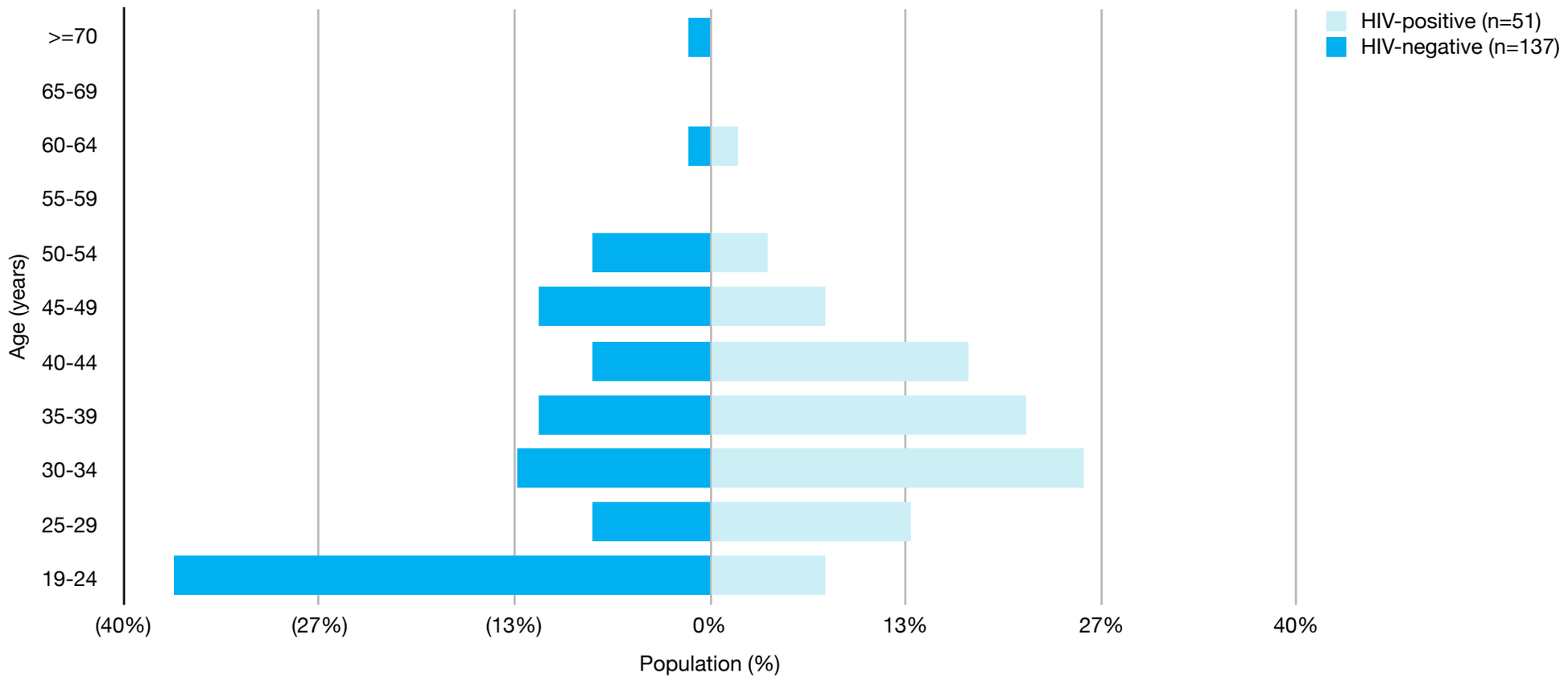
47. Swartz JA, Ducheny K, Holloway T, Stokes L, Willis S, Kuhns LM. A Latent Class Analysis of Chronic Health Conditions Among HIV-Positive Transgender Women of Color. AIDS Behav [Internet]. 2019 [cited 2019 Oct 2]; Available from: http://link.springer.com/10.1007/s10461-019-02543-3

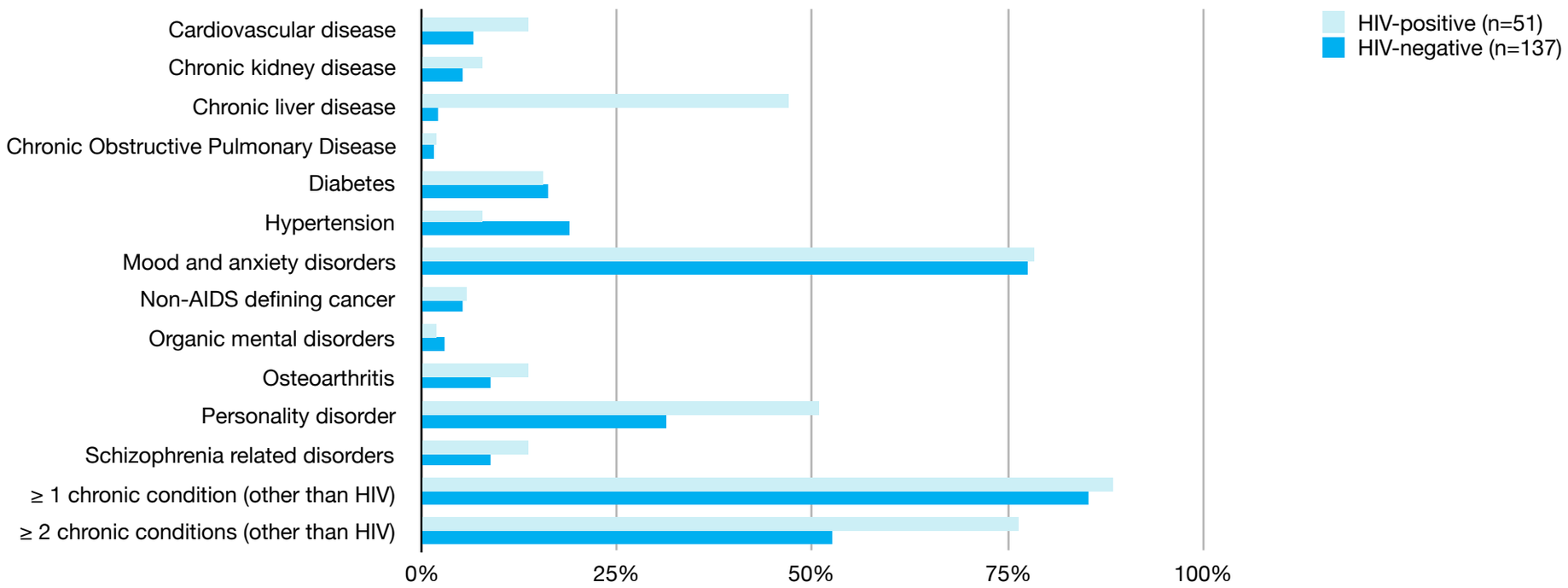
48. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 2011;

49. Reisner SL, Radix A, Deutsch MB. Integrated and Gender-Affirming Transgender Clinical Care and Research. J Acquir Immune Defic Syndr 1999. 2016;72 Suppl 3:S235–42.








Cardiovascular disease
Chronic kidney disease
Chronic liver disease
Chronic Obstructive Pulmonary Disease
Diabetes
Hypertension
Mood and anxiety disorders
Non-AIDS defining cancer
Organic mental disorders
Osteoarthritis
Personality disorder
Schizophrenia related disorders
≥ 1 chronic condition (other than HIV)
≥ 2 chronic conditions (other than HIV)
0%
25%
50%
75%
100%
HIV-positive (n=51)
HIV-negative (n=137)

Supplementary material

Data sources and description of data elements

Data source	Data steward	Description	Data elements provided for this study
Medical Services Plan	British Columbia Ministry of Health	For individuals covered by the provincial universal health insurance plan- Medically necessary services provided by fee-for-service physicians and other healthcare providers, laboratory services, diagnostic procedures, dental/oral surgery	ICD-9 and ICD-10 codes
Consolidation file	British Columbia Ministry of Health	Demographic data on individuals receiving or registered to receive care in BC, pooled from multiple PopData sources	Sociodemographics
Discharge Abstract Database	British Columbia Ministry of Health	Demographic, administrative and clinical data for inpatient hospital discharges and day surgeries	ICD-9 and ICD-10 codes
Vital statistics deaths	British Columbia Vital Statistics Agency	Records of all registered deaths in BC	Death data
Pharmacare	British Columbia Ministry of Health	Data related to prescription drugs dispensed under the BC public drug insurance program.	Prescription drug name/identifier
Pharmanet	British Columbia Ministry of Health Data Stewardship Committee	Data related to prescription drugs dispensed by community and outpatient pharmacies	Prescription drug name/identifier

Supplementary material

Drug Treatment Program and laboratory	British Columbia Centre for Excellence in HIV/AIDS	Antiretroviral therapy use history, laboratory testing, immunological and virologic testing, and demographic data on PLWH who have accessed antiretrovirals in BC	Providers-reported transgender status, laboratory confirmed HIV serostatus

Prescription drugs with drug identification numbers (DIN)s

	Generic Name	DIN
Transfeminine		
Androgen Blockers		
Spironolactone		
	SPIRONOLACTONE	28606
	SPIRONOLACTONE	613215
	SPIRONOLACTONE	285455
	SPIRONOLACTONE	613223
	SPIRONOLACT/HYDROCHLOROTHIAZID	180408
	SPIRONOLACT/HYDROCHLOROTHIAZID	613231
	SPIRONOLACT/HYDROCHLOROTHIAZID	594377
	SPIRONOLACT/HYDROCHLOROTHIAZID	657182
Cyproterone		
	ETHINYL ESTRADIOL/CYPROTERONE	2233542
	NO GENERIC FORMULARY	634514
	CYPROTERONE ACETATE	704431
	CYPROTERONE ACETATE	2229449
	CYPROTERONE ACETATE	2229723
	CYPROTERONE ACETATE	2232872
	CYPROTERONE ACETATE	2245898
	CYPROTERONE ACETATE	704423
Finasteride		
	FINASTERIDE	2010909
	FINASTERIDE	2238213
Dutasteride		
	DUTASTERIDE	2247813
Estrogens		
Estrogen		
	ESTROGENS,CONJUGATED	830240
	ESTROGENS,CONJUGATED	831395

Supplementary material

ETHYNODIOL D-ETHINYL ESTRADIOL	28630
ETHYNODIOL D-ETHINYL ESTRADIOL	469327
NORETHINDRONE-MESTRANOL	22608
NORETHINDRONE-MESTRANOL	22659
NORETHINDRONE A-E ESTRADIOL	297143
NORETHINDRONE A-E ESTRADIOL	315966
NORETHINDRONE-ETHINYL ESTRAD	317047
NORETHINDRONE-ETHINYL ESTRAD	372846
NORETHINDRONE-ETHINYL ESTRAD	373265
NORETHINDRONE-ETHINYL ESTRAD	531006
NORETHINDRONE-ETHINYL ESTRAD	538590
NORETHINDRONE-ETHINYL ESTRAD	602957
NORETHINDRONE-ETHINYL ESTRAD	620947
NORETHINDRONE-ETHINYL ESTRAD	2187086
NORETHINDRONE-ETHINYL ESTRAD	2187108
NORETHINDRONE-ETHINYL ESTRAD	2189054
NORGESTREL-ETHINYL ESTRADIOL	34207
NORGESTREL-ETHINYL ESTRADIOL	300640
LEVONORGESTREL-ETH ESTRA	579386
LEVONORGESTREL-ETH ESTRA	707600
LEVONORGESTREL-ETH ESTRA	782416
LEVONORGESTREL-ETH ESTRA	782432
LEVONORGESTREL-ETH ESTRA	2042320
NORGESTREL-ETHINYL ESTRADIOL	2043033
LEVONORGESTREL-ETH ESTRA	2043726
NORGESTIMATE-ETHINYL ESTRADIOL	2258560
NORETHINDRONE-MESTRANOL	30333
NORETHINDRONE-MESTRANOL	30341
LEVONORGESTREL-ETH ESTRA	2236974
ETHYNODIOL D-ETHINYL ESTRADIOL	471526
NORETHINDRONE-ETHINYL ESTRAD	340731
NORETHINDRONE-MESTRANOL	340758
NORETHINDRONE A-E ESTRADIOL	343838
NORETHINDRONE A-E ESTRADIOL	353027
NORETHINDRONE-ETHINYL ESTRAD	372838
NORETHINDRONE-ETHINYL ESTRAD	373273
NORETHINDRONE-ETHINYL ESTRAD	531014
NORETHINDRONE-ETHINYL ESTRAD	602965
NORETHINDRONE-ETHINYL ESTRAD	695734
NORETHINDRONE-ETHINYL ESTRAD	2187094

Supplementary material

NORETHINDRONE-ETHINYL ESTRAD	2187116
NORETHINDRONE-ETHINYL ESTRAD	2189062
ETHINYL ESTRADIOL/NORETH AC	2242531
NORGESTREL-ETHINYL ESTRADIOL	340766
NORGESTREL-ETHINYL ESTRADIOL	342815
LEVONORGESTREL-ETH ESTRA	586609
LEVONORGESTREL-ETH ESTRA	707503
LEVONORGESTREL-ETH ESTRA	782424
LEVONORGESTREL-ETH ESTRA	782440
LEVONORGESTREL-ETH ESTRA	2042339
NORGESTREL-ETHINYL ESTRADIOL	2043041
LEVONORGESTREL-ETH ESTRA	2043734
NORGESTIMATE-ETHINYL ESTRADIOL	2258587
LEVONORGESTREL-ETH ESTRA	2236975
NORGESTIMATE-ETHINYL ESTRADIOL	1968440
NORGESTIMATE-ETHINYL ESTRADIOL	2028700
NORGESTIMATE-ETHINYL ESTRADIOL	1992872
NORGESTIMATE-ETHINYL ESTRADIOL	2029421
DESOGESTREL-ETHINYL ESTRADIOL	2042487
DESOGESTREL-ETHINYL ESTRADIOL	2042541
DESOGESTREL-ETHINYL ESTRADIOL	2042479
DESOGESTREL-ETHINYL ESTRADIOL	2042533
ESTRADIOL/NORETH AC	2241835
ESTRADIOL/NORETH AC	2241837
LEVONORGESTREL	2241674
ESTROGEN,CON/M-PROGEST ACET	2242878
ESTROGEN,CON/M-PROGEST ACET	2242879
ESTRADIOL/NORETH AC	2243529
ESTRADIOL/NORETH AC	2243530
ETHINYL ESTRADIOL/DROSPIRENONE	2261723
ETHINYL ESTRADIOL/DROSPIRENONE	2261731
ETONOGESTREL/ETHINYL ESTRADIOL	2253186
ETHINYL ESTRADIOL/NORELGEST	2248297
DIENESTROL	441295
DIETHYLSTILBESTROL	3360
DIETHYLSTILBESTROL	2091461
DIETHYLSTILBESTROL	2091488
ESTRADIOL	464791
ESTRADIOL	2148587
ESTRADIOL	464805

Supplementary material

ESTRADIOL	2148595
ESTRADIOL VALERATE	29238
ESTRADIOL	756849
ESTRADIOL	2237807
ESTRADIOL	2243722
ESTRADIOL	2245676
ESTRADIOL	756857
ESTRADIOL	2204428
ESTRADIOL	2231509
ESTRADIOL	2237808
ESTRADIOL	2243724
ESTRADIOL	2244000
ESTRADIOL	2246967
ESTRADIOL	756792
ESTRADIOL	2204444
ESTRADIOL	2231510
ESTRADIOL	2244002
ESTRADIOL	2246969
ESTRADIOL	2168898
ESTRADIOL	2204436
ESTRADIOL	2244001
ESTRADIOL	2246968
ESTRADIOL	2225190
ESTRADIOL	2204401
ESTRADIOL	2238704
ESTRADIOL	2243999
ESTRADIOL	2241332
ESTRADIOL	2247499
ESTRADIOL	2247500
ESTROGENS,CONJUGATED	2569
ESTROGENS,CONJUGATED	2043394
ESTROGENS,CONJUGATED	2230891
ESTROGENS,CONJUGATED	2239654
ESTROGENS,CONJUGATED	2577
ESTROGENS,CONJUGATED	265470
ESTROGENS,CONJUGATED	587281
ESTROGENS,CONJUGATED	2043408
ESTROGENS,CONJUGATED	2089
ESTROGENS,CONJUGATED	2043440
ESTROGENS,CONJUGATED	403466

Supplementary material

	ESTROGENS,CONJUGATED	2043416
	ESTROGENS,CONJUGATED	2230892
	ESTROGENS,CONJUGATED	2239655
	ESTROGENS,CONJUGATED	2585
	ESTROGENS,CONJUGATED	265489
	ESTROGENS,CONJUGATED	587303
	ESTROGENS,CONJUGATED	2043424
	ESTROGENS,CONJUGATED	2043432
	ESTROGENS,CONJUGATED	2043386
	ME-TESTOSTERONE/ESTROGEN,CON	53538
	ESTROPIPATE	282685
	ESTROPIPATE	2089769
	ESTROPIPATE	282677
	ESTROPIPATE	2089777
	ESTROPIPATE	2089793
	ESTRADIOL/NORETH AC	2108186
	NORGESTIMATE-ETHINYL ESTRADIOL	2229064
	NORGESTIMATE-ETHINYL ESTRADIOL	2229218
	NORGESTIMATE-ETHINYL ESTRADIOL	2229226
	ETHINYL ESTRADIOL/CYPROTERONE	2233542
	ETHINYL ESTRADIOL/NORELGEST	2246340
	NO GENERIC FORMULARY	66124057
	NO GENERIC FORMULARY	66124058
	NO GENERIC FORMULARY	66124060
	NO GENERIC FORMULARY	66124061
	NO GENERIC FORMULARY	66124062
	NO GENERIC FORMULARY	66124063
	NO GENERIC FORMULARY	66124064
Progestogens		
Progesterone		
	PROGESTERONE,MICRONIZED	2241013
	MEDROXYPROGESTERONE ACET	30848
	MEDROXYPROGESTERONE ACET	30856
	MEDROXYPROGESTERONE ACET	585092
	NO GENERIC FORMULARY	66123240
	MEDROXYPROGESTERONE ACET	708917
	MEDROXYPROGESTERONE ACET	2148552
	MEDROXYPROGESTERONE ACET	2221284
	MEDROXYPROGESTERONE ACET	2229838
	MEDROXYPROGESTERONE ACET	2244726

Supplementary material

	MEDROXYPROGESTERONE ACET	2246627
	MEDROXYPROGESTERONE ACET	30937
	MEDROXYPROGESTERONE ACET	2010739
	MEDROXYPROGESTERONE ACET	2148560
	MEDROXYPROGESTERONE ACET	2221292
	MEDROXYPROGESTERONE ACET	2229839
	MEDROXYPROGESTERONE ACET	2244727
	MEDROXYPROGESTERONE ACET	2246628
	MEDROXYPROGESTERONE ACET	729973
	MEDROXYPROGESTERONE ACET	2010933
	MEDROXYPROGESTERONE ACET	2148579
	MEDROXYPROGESTERONE ACET	2221306
	MEDROXYPROGESTERONE ACET	2229840
	MEDROXYPROGESTERONE ACET	2246629
	MEDROXYPROGESTERONE ACET	30945
	MEDROXYPROGESTERONE ACET	2267640
	NORETHINDRONE	37605
	PROGESTERONE,MICRONIZED	2166704
	PROGESTERONE	739952
	PROGESTERONE	1977652
	PROGESTERONE	2128470
	LEVONORGESTREL	2243005
Transmasculine		
Testosterone		
	TESTOSTERONE	2249499
	TESTOSTERONE CYPIONATE	30783
	TESTOSTERONE PROPIONATE	1977571
	TESTOSTERONE CYPIONATE	1977601
	TESTOSTERONE CYPIONATE	2220318
	TESTOSTERONE CYPIONATE	2246063
	TESTOSTERONE ENANTHATE	29246
	TESTOSTERONE ENANTHATE	716936
	TESTOSTERONE ENANTHATE	739944
	TESTOSTERONE UNDECANOATE	782327
	TESTOSTERONE ENANTHATE/ESTRAD	108278
	TESTOSTERONE ENANTHATE/ESTRAD	2061031
	TESTOSTERONE	2239653
	TESTOSTERONE	2245346
	TESTOSTERONE	2245345
	TESTOSTERONE	2245972

Supplementary material

Chronic condition case definitions

Chronic condition		Case definition	Codes
Cardiovascular disease*	Acute myocardial infarction	1 or more hospitalizations with relevant ICD codes	ICD-10: I21 Acute myocardial infarction I22 Subsequent myocardial infarction ICD-9: 410 Acute myocardial infarction
	Ischemic heart disease	At least one of the following: 2 medical visits with Angina ICD-9 code 413 plus 1 heart disease prescription in 365 days; or 1 specialist visit with Angina ICD-9 code 413 plus 1 prescription in 365 days; or 2 medical visits with two ICD9 codes 410, 411, 412, 413, 414 in 365 days; or 1 CCI/CCP CABG,PCI/PCTA procedure code; or 1 hospitalization with relevant ICD code.	ICD-10: I20 Angina pectoris I21 Acute myocardial infarction I 22 Subsequent myocardial infarction I23 Certain current complications following acute myocardial infarction I24 Other acute ischaemic heart diseases I25 Chronic ischaemic heart disease ICD-9: 410 Acute myocardial infarction 411 Other acute and subacute forms of ischaemic heart disease 412 Old myocardial infarction 413 Angina pectoris 414 Other forms of chronic ischaemic heart disease
	Chronic heart failure	1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	ICD-10: I50 Heart failure ICD-9: 428 Heart failure
	Stroke- hospital	1 or more hospitalizations with relevant ICD codes	ICD-10: H34.1 Central retinal artery occlusion I60 Subarachnoid hemorrhage I61 Intracerebral haemorrhage I63 Cerebral infarction (exclude I63.6 Cerebral infarction due to

Supplementary material

			cerebral venous thrombosis, nonpyogenic) I64 Stroke, not specified as haemorrhage or infarction 362.3 Retinal vascular occlusion 430 Subarachnoid hemorrhage 431 Intracerebral hemorrhage 433.x1 Occlusion and stenosis of precerebral arteries 434.x Occlusion cerebral arteries 436 Acute but ill-defined cerebrovascular disease *Excludes any traumatic brain injury*
	Transient ischemic attack	1 or more hospitalizations with relevant ICD codes	ICD-10: H34.0 Transient retinal artery occlusion G45.0 Vertebro-basilar artery syndrome G45.1 Carotid artery syndrome (hemispheric) G45.2 Multiple and bilateral precerebral artery syndromes G45.3 Amaurosis fugax G45.8 Other transient cerebral ischemic attacks and related syndromes G45.9 Transient cerebral ischemic attack, unspecified ICD-9: 435 Transient cerebral ischemia *Excludes any traumatic brain injury*
Chronic kidney disease*		1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	ICD-10: N01 Rapidly progressive nephritic syndrome N03 Chronic nephritic syndrome N04 Nephrotic syndrome N05 Unspecified nephritic syndrome N06 Isolated proteinuria with specified morphological lesion N07 Hereditary nephropathy, not elsewhere classified

Supplementary material

		N18 Chronic kidney disease N19 Unspecified kidney failure N26 Unspecified contracted kidney N27 Small kidney of unknown cause ICD-9: 581 Nephrotic syndrome 582 Chronic glomerulonephritis 583 Nephritis and nephropathy, not specified as acute or chronic 585 Chronic renal failure 586 Renal failure, unspecified 587 Renal sclerosis, unspecified 589 Small kidney of unknown cause
Chronic liver disease	1 or more hospitalization or medical visit with relevant diagnosis within 365 days	ICD-9: 571.0 Alcoholic fatty liver 571.2 Alcoholic cirrhosis of liver 571.3 Alcoholic liver damage, unspecified 571.4 Chronic hepatitis 571.5 Cirrhosis of liver without mention of alcohol 571.6 Billiary cirrhosis 571.8 Other chronic nonalcoholic liver disease 571.9 Unspecified chronic liver disease without mention of alcohol 070.3 Viral hepatitis B without mention of hepatic coma 070.30 Viral hepatitis B without mention of hepatic coma, acute or unspecified, without mention of hepatitis delta 070.31 Viral hepatitis B without mention of hepatic coma, acute or unspecified, with hepatitis delta 070.32 Viral hepatitis B without mention of hepatic coma, chronic, without mention of hepatitis delta 070.33 Viral hepatitis B without mention of hepatic coma, chronic, with hepatitis delta

Supplementary material

		070.52 Hepatitis delta without mention of active Hepatitis B disease or hepatic coma V02.61 Hepatitis B carrier 070.42 Hepatitis delta without mention of active Hepatitis B disease with hepatic coma 070.54 Chronic hepatitis C without mention of hepatic coma V02.62 Hepatitis C carrier
Chronic Obstructive Pulmonary Disease*	1 or more hospitalization or 2 or more medical visits within 365 days	ICD-10: J41 Simple and mucopurulent chronic bronchitis J42 Unspecified chronic bronchitis J43 Emphysema J44 Other chronic obstructive pulmonary disease ICD-9: 491 Chronic bronchitis 492 Emphysema 496 Chronic airways obstruction, not elsewhere classified
Diabetes Mellitus*	At least 1 of the following: 1 hospitalization or 2 medical visits in 365 days with relevant ICD codes; or 2 or more insulin prescriptions in 365 days; or 2 or more oral antihyperglycemic (not including metformin) prescriptions in 365 days; or 1 insulin and 1 oral antihyperglycemic (including metformin) in 365 days; or 2 metformin prescriptions and 1 medical visit in one year with relevant ICD codes.	ICD-10: E10 Type 1 diabetes mellitus E11 Type 2 diabetes mellitus E13 Other specified diabetes mellitus E14 Unspecified diabetes mellitus ICD-9: 250 Diabetes mellitus

Supplementary material

	Excludes gestational diabetes.	
Hypertension*	1 or more hospitalizations or 2 or more medical visits within 2 years with relevant ICD codes. *Excludes gestational hypertension.*	ICD-10: I10 Essential (primary) hypertension I11 Hypertensive heart disease I12 Hypertensive renal disease I13 Hypertensive heart and renal disease I15 Secondary hypertension ICD-9: 401 Essential hypertension 402 Hypertensive heart disease 403 Hypertensive renal disease 404 Hypertensive heart and renal disease 405 Secondary hypertension
Mood and anxiety disorders*	1 or more hospitalizations with a relevant ICD code or 2 or more medical visits with a relevant code within 2 years	ICD-10: F30 Manic episode F31 Bipolar affective disorder F32 Depressive episode F33 Recurrent depressive disorder F34 Persistent mood [affective] disorders F38 Other mood [affective] disorders F39 Unspecified mood [affective] disorder F40 Phobic anxiety disorders F41 Other anxiety disorders F42 Obsessive-compulsive disorder F43 Reaction to severe stress, and adjustment disorders F44 Dissociative (conversion) disorders F45 Somatoform disorders F48 Other neurotic disorders F68 Other disorders of adult personality & behavior ICD-9:

Supplementary material

			296 Affective psychoses 300 Neurotic disorders 311 Depressive disorder, not elsewhere classified MSP DX Code: 50B Anxiety/Depression
Non-AIDS defining cancer†	All prevalent cancer cases were included, with the exception of AIDS defining malignancies (Kaposi's sarcoma, non-Hodgkin's lymphoma, invasive cervical cancer)		*Cancer case definition details available from the British Columbia Cancer Agency: http://www.bccancer.bc.ca/health-info/types-of-cancer*
Organic mental disorders		1 or more medical visits or hospitalizations with relevant diagnoses within 365 days	ICD-9: 290.x Dementias 294.x Other organic psychotic conditions 331.x Alzheimer's ICD-10: F00.x Dementia in Alzheimer's disease F01.x Vascular Dementia F02.x Dementia in other disease classified elsewhere F03.x Unspecified dementia F04 Amnestic disorder due to physiological condition F06 Other mental disorders due to known physiological condition F09 Unspecified mental disorder due to known physiological condition G30 Alzheimer's disease with early onset
Osteoarthritis*		1 or more hospitalization or 2 or more medical visits in 365 days with a relevant ICD code	ICD-10: M15 Polyarthrosis M16 Coxarthrosis [arthrosis of hip] M17 Gonarthrosis [arthrosis of knee] M18 Arthrosis of first carpometacarpal joint M19 Other arthrosis

Supplementary material

		ICD-9: 715 Osteoarthrosis and allied disorders
Personality disorder	1 or more hospitalizations or medical visits with a relevant diagnosis within 365 days	ICD-9: 301.x Personality disorders ICD-10: F60.x Specified personality disorders F62 Enduring personality changes, not attributable to brain damage and disease F68.1 Intentional production or feigning of symptoms or disabilities, either physical or psychological F68.8 Other specified disorders or adult personality and behaviour F69 Unspecified disorder or adult personality and behaviour
Schizophrenia related disorder	1 or more medical visit or hospitalizations with relevant diagnoses within 365 days	ICD-9: 295.x Schizophrenic disorders 297.0 Paranoid state, simple 297.1 Delusional disorder 297.2 Paraphrenia 297.3 Shared psychotic disorder ICD-10: F20.x Paranoid schizophrenia F21.x Schizotypal disorder F23.2 Acute schizophrenia-like psychotic disorder F25.x Schizoaffective disorders

* Case definition adapted from British Columbia Ministry of Health version 2017, April 4 2019 update

† Case-definition adapted from British Columbia Cancer Agency

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	1.1-3: Title page
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			Introduction (pp 4-5)
Objectives	3	State specific objectives, including any prespecified hypotheses			Introduction (page 5)
Methods					
Study Design	4	Present key elements of study design early in the paper			Methods (page 5); Supplementary Material
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			Methods (page 5)

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	6.1: Methods (pp 6-7); Supplementary Material 6.2: Methods (pp 6-7) 6.3: Figure 1
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	7.1: Methods (pp 6-7); Supplementary Material
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			Methods (pp 6-7)

Bias	9	Describe any efforts to address potential sources of bias			Discussion (pp 9, 11)
Study size	10	Explain how the study size was arrived at			Methods (page 6); Figure 1
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Methods (pp 6-7)
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			Methods (pp 6-7)
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	12.1-2: Methods

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Methods (page 5)
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Results (page 8); Figure 1
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			Results (page 8)
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure			Results

		category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			NA
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			NA
Discussion					
Key results	18	Summarise key results with reference to study objectives			Discussion: Page 9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Discussion: Page 9, Page 11
Interpretation	20	Give a cautious overall interpretation of results considering objectives,			Conclusion (Page 11)

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results			Limitations (Page 9)
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			Funding (Page 12)
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Page 12

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

BMJ Open

Development of a computable phenotype to identify a transgender sample for health research purposes: A feasibility study in a large linked provincial healthcare administrative cohort in British Columbia, Canada

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-040928.R1
Article Type:	Original research
Date Submitted by the Author:	09-Nov-2020
Complete List of Authors:	Rich, Ashleigh; The University of British Columbia, School of Population & Public Health; BC Centre for Excellence in HIV/AIDS, Poteat, Tonia; University of North Carolina at Chapel Hill, Department of Social Medicine Koehoorn, Mieke; University of British Columbia, School of Population and Public Health; Li, Jenny; BC Centre for Excellence in HIV/AIDS, Ye, Monica; BC Centre for Excellence in HIV/AIDS, Sereda, Paul; British Columbia Centre for Excellence in HIV/AIDS, Salway, Travis; Simon Fraser University Hogg, R; Simon Fraser University,
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	HIV/AIDS, Health services research, Research methods, Public health
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, Sexual and gender disorders < PSYCHIATRY, HIV & AIDS < INFECTIOUS DISEASES, SOCIAL MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

TITLE PAGE

Title: Development of a computable phenotype to identify a transgender sample for health research purposes: A feasibility study in a large linked provincial healthcare administrative cohort in British Columbia, Canada

Authors: Rich AJ[1,2], Poteat T[3], Koehoorn M[1], Li J[2], Ye M[2], Sereda, P[2], Salway T[4], Hogg RS[2,4]

Affiliations:
1. School of Population and Public Health, University of British Columbia, Vancouver, Canada
2. British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada
3. Department of Social Medicine, University of North Carolina- Chapel Hill, Chapel Hill, USA
4. Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada

Corresponding author:
Ashleigh J Rich
ajrich@mail.ubc.ca

2206 East Mall
Vancouver, BC
V6T 1Z3
Canada

Word count: 4000/4000

ABSTRACT

Objectives: Innovative methods are needed for identification of transgender people in administrative records for health research purposes. This study investigated the feasibility of using transgender-specific healthcare utilization in a Canadian population-based, health records database to develop a computable phenotype (CP) and identify the proportion of transgender people within the HIV-positive population as a public health priority.

Design: The COAST cohort comprises a data linkage between two provincial data sources: The BC Centre for Excellence in HIV/AIDS Drug Treatment Program, which coordinates HIV treatment dispensation across BC; and Population Data BC, a provincial data repository holding individual, longitudinal data for all BC residents (1996-2013).

Setting: British Columbia, Canada.

Participants: COAST participants include 13 907 BC residents living with HIV (≥19 years of age) and a 10% random sample comparison group of the HIV-negative general population (514 952 individuals).

Primary and secondary outcome measures: Healthcare records were used to identify transgender people via a CP algorithm (diagnosis codes + androgen blocker/hormone prescriptions), to examine related diagnoses and prescription concordance, and to validate the CP using an independent provider-report transgender status measure. Demographics and chronic illness burden were also characterized for the transgender sample.

Results: The best-performing CP identified 137 HIV-negative and 51 HIV-positive transgender people (total 188). In validity analyses, the best-performing CP had low sensitivity (27.5%, 95%CI:17.8-39.8), high specificity (99.8%, 95%CI:99.6-99.8), low agreement using Kappa statistics (0.3, 95%CI:0.2-0.5), and moderate positive predictive value (43.2%, 95%CI:28.7-58.9). There was high concordance between exogenous-sex hormone use and transgender-specific diagnoses.

Conclusions: The development of a validated CP opens up new opportunities for identifying transgender people for inclusion in population-based health research using administrative health data, and offers the potential for much-needed and heretofore unavailable evidence on health status, including HIV status, and the healthcare use and needs of transgender people.

KEYWORDS: Transgender Persons, Health Services, Algorithms, Canada

ARTICLE SUMMARY

Strengths and limitations of this study:

- This study demonstrates the feasibility of developing and validating a computable phenotype for identification of a transgender sample, using a population-based representative source population and healthcare records.
- A major contribution of this study is the ascertainment of the population of transgender people living with HIV in the Canadian province of British Columbia, in a universal

healthcare setting, using a computable phenotype, and capacity to estimate the prevalence of transgender status among the population living with HIV in the province.

- Development of a validated transgender computable phenotype algorithm lays the foundation for future investigation of transgender-specific research questions related to general and HIV-specific healthcare use and health outcomes for this key population.

INTRODUCTION

Limited data on transgender people

Transgender people are often overlooked within epidemiological research and population health surveillance due to small sample size, limited research designs, and other institutional and methodological erasures.[1–3] A 2017 review of Medline-indexed literature from 1950 to 2016 found 2405 published articles including transgender people, with almost half published in the last decade.[4] A 2008 United States (US)-based meta-analysis of HIV prevalence among transgender populations found 24 studies of transgender women, and five additional studies of transgender men,[5] though an updated review found 43 primary studies on transgender women and 15 on transgender men published between 2006-2017.[6] Despite this recent increase in transgender health research in general and for HIV specifically, much of the literature has focused on transgender-specific care, mental health and HIV/sexual health,[7,8] leaving the population understudied, in particular in the broader areas of physical health and healthcare utilization.

The erasures or exclusions of transgender persons in health studies may be explained, in part, by methodological challenges. Specific to electronic health record (EHR) data, a 2017 report identified only one transgender person among 38,5820 cancer cases in a Minnesota cancer registry,[9] clearly an undercount given that 0.4% of the US general population and 0.6% (95% credibility intervals: 0.5%-0.7%) of the Minnesota population is estimated to be transgender.[10,11] This highlights the need for improved gender ascertainment and transgender inclusion in research relying on patient records and administrative data. The establishment of best practices for measuring transgender status in survey research, such as the two-step method (measuring sex assigned at birth and current gender identity), points to a way forward for transgender-inclusive population health research.[12,13] However, innovative research methods are needed to identify transgender people in studies that rely on existing data sources (in particular EHR) and that optimize the use of transgender respondents' data in non-transgender specific research.

Computable phenotypes for transgender health research

Previous research in transgender health largely comprises cross-sectional studies, case reports, and qualitative or observational research.[7] Much consists of clinic- or venue-based convenience samples or lack comparison groups.[7,8] The literature is further characterized by inconsistent transgender status measurement,[14] small sample sizes, and focus on the United States (US).[8] In response, researchers have called for advancing transgender health research methods - namely ascertainment of high-quality samples via systematic approaches - including for general population-based and health systems-based studies.[15] One opportunity for the advancement of transgender health research methods is the emerging use of computable phenotypes (CPs)[16] or case ascertainment algorithms, to identify transgender samples in healthcare utilization data. A computable phenotype is an algorithm for identifying a clinical feature, condition, or set of characteristics that can be determined directly from EHR and other ancillary health care data systems (e.g., disease registries, insurance claims data) data.[17] CPs are developed using a combination of data elements (e.g., sociodemographic variables, clinical diagnoses) and value sets (i.e., the selection of a set of relevant values for each data element). Development of CPs using standardized methods and definitions enables identification and inclusion of transgender persons in research, as well as replication of analyses across data

sources, healthcare organizations/sites and studies. CPs have application in clinical care, surveillance, and health research.

Recently, CP and other EHR-based algorithm methods have been applied in a number of settings primarily in the US to identify transgender samples for health research.[14] Specifically, the STRONG study identified a transgender cohort (n = 6,456) using EHR data from Kaiser Permanente health plan members in California and Georgia, for investigation of general and transgender-specific health outcomes.[18] Blosnich et al identified 3,177 people with a "gender identity disorder" diagnosis among military veterans accessing care through the US Veterans Health Administration healthcare system,[19] for examination of mental health and other outcomes. Researchers with the US Centers for Medicare & Medicaid Services identified 4,098 transgender beneficiaries using national Medicare claims data,[20] and researchers at Vanderbilt University identified 234 transgender patients in their university clinic EHR data.[16] While these cohorts represent important opportunities for advancement of transgender health research, these methods have yet to be applied widely outside the US context. This is particularly important as different jurisdictions may vary in medical billing and coding practices, healthcare system patient populations, and representativeness of the general population. Specifically, in Canada, healthcare is delivered through a provincially administered universal healthcare system. As such, research using EHR provides an opportunity to develop methods for population-based, representative estimates of transgender populations within the Canadian context. Coupled with the current absence of gender ascertainment measures in population-based routinely collected data (e.g., census, national government health surveys, etc.) in Canada and many other jurisdictions, this remains an evidence need.

Summary of study rationale

This study investigated the application of emerging transgender health research methods, specifically CPs, in a Canadian context for the first time, testing the feasibility of identification of a transgender sample using EHR data from a provincial healthcare administrative data-linked cohort.

METHODS

Data Sources and Participants

The Comparative Outcomes and Service Utilization Trends Study (COAST)

COAST is a population-based cohort study focused on health services utilization research questions among all people known to be living with HIV (PLWH) in the province of British Columbia (BC) and a 10% random sample comparison group of the HIV-negative general population.[21] The COAST cohort comprises individual-level, longitudinal data from PLWH who have ever accessed HIV treatment in BC between 1996 and 2013, provided by Population Data BC (PopDataBC)[21] via data linkage between two provincial data sources, by personal health number: the Drug Treatment Program (DTP) [22] and the Ministry of Health. PopDataBC provides infrastructure for access to, and linkage of, longitudinal and individual-level administrative health data for all BC residents.[23]. The HIV-negative general population cohort was drawn randomly from the Ministry of Health registry data by PopDataBC. The COAST study has received approval from the University of British Columbia/Providence Health Care Research Ethics Board (#H09-02905) and Simon Fraser University Office of Research Ethics (#2013 s0566). The study complies with the BC Freedom of Information and Protection of

Privacy Act (FIPPA) and did not require informed consent as it is conducted using retrospective administrative and anonymized data for research and statistical purposes only. No patients or public were involved in this study.

Drug Treatment Program

In BC, antiretroviral therapy (ART) is provided to PLWH at no cost to the patient, and distributed through the DTP.[22] The DTP contributed a provider-reported measure of transgender status for COAST.

Ministry of Health

Ministry of Health data available via COAST included insured medical service billing records for outpatient visits,[24,25] hospital (in-patient) visits,[26] prescription medications,[27,28] and vital statistics.[29]

Measures & Analyses

Transgender computable phenotypes

Identification of transgender cases was tested in COAST using International Classification of Disease (ICD) codes (9th and 10th editions) and exogenous sex hormone prescription use. Transgender-specific diagnoses in medical and hospital billing records included the ICD-9 codes 302.5 Trans-sexualism with unspecified history, 302.51 Trans-sexualism with asexual history, 302.52 Trans-sexualism with homosexual history, 302.53 Trans-sexualism with heterosexual history, 302.6 Gender Identity Disorder in children, 302.85 Gender Identity Disorder in adolescents or adults; and ICD-10 codes F64.0 Gender Identity Disorder of childhood, F64.2 Gender Identity Disorder of childhood, F64.8 Other Gender Identity Disorder, and F64.9 Gender Identity Disorder unspecified. The full list of androgen blockers and exogenous sex hormone prescriptions included in analyses is available in the supplementary material.

Concordance

To assess face validity and utility of diagnosis and prescription data over time in CP development (i.e. whether the identified transgender sample had exogenous sex hormone prescription use and other diagnoses patterns consistent with that of transgender populations in other studies), concordance analyses evaluated the presence of at least one included diagnosis and prescription during the COAST study follow-up period with the presence of at least one included diagnosis and prescription in the last study year. Concordance was assessed between transgender-specific diagnoses, exogenous sex hormone and androgen blocker prescriptions, and non-transgender specific diagnoses (ICD-9 259.9 Unspecified Endocrine Disorder and ICD-10 E34.9 Endocrine Disorder, Unspecified [see supplementary material]). Endocrine disorder diagnosis codes are sometimes preferred by medical providers treating transgender people in response to historic exclusions of transgender-specific care from insurance coverage and to combat the stigma of transgender-specific diagnosis codes that have historically been classified as psychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM).[30] Exogenous sex hormone use, while common in transgender populations,[5,31] is not transgender-specific. Cisgender populations also use androgen blocker and sex hormone prescriptions (e.g. estrogen to treat menopausal symptoms in cisgender women, spironolactone is used for hypertension), thus exogenous sex hormone and androgen blocker prescription use

cannot independently identify transgender people. At the same time, not all transgender people use hormones and some access via non-medical sources.[32,33]

Validation

In British Columbia, transgender status is collected in the DTP via a provider-reported sex variable ("Male", "Male to Female", "Female to Male", or "Female"). Patients reported as either "Male to Female" or "Female to Male" were classified as transgender. The provider-reported transgender measure, available for the HIV positive cohort only, was used as a 'gold standard' for CP validation. Sensitivity, specificity, positive predictive value and kappa statistics with corresponding 95% confidence intervals (CI) were calculated for identifying transgender people via the CPs, in the HIV positive cohort only.

Demographics and chronic conditions

To further assess face validity of the transgender CP for future health research, descriptive statistics were calculated for the total transgender sample (both HIV-positive and HIV-negative) using the COAST study key sociodemographic and health variables, specifically laboratory confirmed HIV serostatus (HIV-positive/HIV-negative), baseline age, patient's Health Authority (five provincial regions for the administration of health services that include large urban centres, suburban regions, and rural/remote areas), and chronic illness burden based on standardized case definitions from the BC Ministry of Health [34] and the BC Cancer Agency.[35]

RESULTS

The total COAST cohort included 528 859 people, of which 514 952 were HIV-negative (10% general population random sample) and 13 907 were PLWH (Figure 1).

[Figure 1 here]

Concordance

Of the 237 people who had ever had a transgender-specific diagnosis during the study period, 19.4% also had a recent diagnosis in the last follow-up year (Table 1). None had an unspecified endocrine disorder diagnosis at any time, thus this diagnosis was excluded from all CPs. Of the 237, 79.3% had an exogenous sex hormone or androgen blocker prescription at least once during the study period and 46.4% had one in the last year.

Table 1. Concordance analyses for diagnoses and hormone measures

	N	%
≥ Transgender ICD- ever	237	100
≥ Transgender ICD- recent	46	19.4
Unspecified endocrine disorder use- ever	0	0.0
Unspecified endocrine disorder use- recent	0	0.0
≥ Hormone/blocker use- ever	188	79.3
≥ Hormone/blocker use-recent	110	46.4

Validation

While no one CP consistently performed well across all validation metrics, the CP with the best overall performance across test statistics was based on having received at least one transgender-

specific diagnosis and at least one androgen blocker/exogenous sex hormone prescription over the study follow-up period (Table 2). This CP had high specificity (99.8%, 95% CI: 99.6-99.8), low sensitivity (27.5%, 95% CI: 17.8-39.8), low to moderate Kappa coefficients (0.3, 95% CI: 0.2-0.5) and moderate positive predictive values (43.2%, 95% CI: 28.7-58.9).

Table 2. Validation measures of transgender computable phenotype (CP) with provider-report transgender status measures, in COAST HIV-positive cohort

CP	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	Kappa (95% CI)
≥ 1 transgender ICD- ever	27.5 (17.8, 39.8)	99.7 (99.6, 99.8)	40.4 (26.7, 55.7)	0.3 (0.2, 0.4)
≥ 1 transgender ICD- recent	8.7 (3.6, 18.6)	100.0 (99.9, 100.0)	85.7 (42.0, 99.4)	0.2 (0.1, 0.3)
≥ 1 transgender ICD AND ≥ 1 hormone/blocker use- ever	27.5 (17.8, 39.8)	99.8 (99.6, 99.8)	43.2 (28.7, 58.9)	0.3 (0.2, 0.5)
≥ 1 transgender ICD AND ≥ 1 hormone/blocker use- recent	7.3 (2.7, 16.8)	100.0 (99.9, 100.0)	83.3 (36.5, 99.1)	0.1 (0.0, 0.2)

Transgender phenotype

Applying the best-performing CP, 137 HIV-negative people and 51 HIV-positive people (188 total) were identified as transgender in the respective COAST cohorts (Figure 1).

Demographics and chronic conditions

Demographic characteristics and chronic conditions for the 188 transgender people identified via the best-performing CP are presented in Figures 2 to 4. Transgender people were geographical located throughout BC health regions. The Vancouver Coastal Health Authority region, which includes the largest municipal area in BC, had the highest concentration of transgender people (44.2%) while the Northern Health Authority region - a predominantly rural and remote area of the province - had the lowest (1.6%).[36] The HIV-positive group had a higher median age than the HIV-negative group (35 [Q1, Q3: 30,42] and 30 [Q1, Q3: 19,42], respectively). For the HIV-negative sample, the largest proportion of transgender people were aged 19 to 29 years (44.5%) and the smallest proportion aged 55 years and older (<3%). For the HIV-positive sample, the largest proportion were aged 30 to 34 years (25.5%) and the smallest proportion aged 55 years and older (<2%).

[Figures 2 and 3 here]

Overall, HIV-positive transgender people had a higher prevalence of at least one chronic condition (other than HIV) compared to HIV-negative transgender people (88.2% versus 85.4%, respectively), and of two or more chronic conditions (76.5% versus 52.6%, respectively). Specific chronic disease differences between transgender people living with and without HIV were most notable for a higher prevalence among the HIV-positive cohort of cardiovascular

disease, chronic kidney disease, osteoarthritis, schizophrenia and personality disorders, and chronic liver disease, but a lower prevalence for hypertension.

[Figure 4 here]

DISCUSSION

This study demonstrates the feasibility of identification of a sample of transgender people in a large linked provincial healthcare administrative database, using a CP based on prescriptions and diagnoses. Among a growing number of studies using EHR and CP methods to identify transgender samples for health research purposes, this is the first to do so in Canada., to independently validate the CP using a 'gold standard' of provider-reported transgender status, and the only to use population-based data.

Concordance

There was high concordance between transgender-specific diagnoses and exogenous sex hormone or androgen blocker prescription use in this study. That nearly half of those with at least one transgender-specific diagnosis had been dispensed hormones or blockers in the past year is consistent with findings from US and Canadian studies (48.9% and 43.0%, respectively)[20,32,33] - suggesting face validity for the current CP.

CP development and validation

The best-performing CP overall successfully identified cisgender people who were truly cisgender (specificity) and correctly identified transgender people who were truly transgender (0.2% false positive rate, results not shown). However, the selected CP had relatively low sensitivity, missing approximately 72.5% of 'true' transgender people in COAST, as identified by the gold standard provider-based measure. Though a relatively small proportion of the 'true' transgender sample was identified in this study, the impact on future analyses comparing health outcomes for transgender and cisgender groups is likely negligible, as even the large proportion of 'true' transgender people misclassified as cisgender (approximate n=496) is a very small proportion of the total COAST sample. At worst, this misclassification would bias results related to disparities between transgender and cisgender health toward the null, producing a conservative attenuated effect. Further, as discussed below, gender identity classification will likely greatly improve as transgender care shifts further into the fee-for-service system in BC. As in other Canadian administrative data studies, low sensitivity may be explained in part by provider and system billing preferences using 3-digit ICD diagnosis coding instead of the more specific 4-digit coding, and inconsistencies in the BC billing management system.[37] Despite the low sensitivity, CP development in this study with high specificity offers an advancement for transgender health research. A measure that correctly identifies cases for transgender samples in research with good success translates to better opportunities to include transgender people in health studies and to investigate their health relative to other groups. While future research may lead to improvements in CP development, the CP identified in the current study with good specificity, albeit relatively poor sensitivity, has important utility in advancing opportunities in transgender health research.

The limited agreement between the CP and provider-report transgender status may be due to the widely varying transgender status prevalence depending on study design and ascertainment

measures used.[14] In the BC context, the CP and the DTP measures are assessing transgender status in different ways and for different purposes. In the DTP, transgender status is ascertained in the context of HIV diagnosis and ART prescribing, during which demographics and HIV transmission risk factors are recorded. This differs from recording diagnoses in EHR for those accessing transgender-specific care as utilized in the CP. This may explain the lower PPV for the best-performing CP compared to the CP based on recent transgender diagnoses, suggesting the DTP provider-reported transgender status measure has better coverage for recent cases and the potential for use of recent diagnosis over ever to be beneficial in future CP development. Ultimately, a single CP may not be sufficient for all intended purposes and the best applicable CP (using different types of diagnoses, prescriptions or procedures) may differ depending on the intended healthcare, health research, or health policy application.[17]

There is limited literature on EHR-based studies with the ability to validate an administrative transgender measure using a 'gold standard' comparison measure.[16] The two previous studies that have developed and validated algorithms to identify transgender individuals have both been conducted in non-representative samples in the US, one using Medicare data[38] and one in a university medical center.[16] Similar to the current study, the Medicare study found high specificity when comparing an EHR-based and a two-step survey-based transgender measure. However, the Medicare study found that the EHR measure performed consistently well with high sensitivity and a high Kappa statistic, unlike in the current study. Using chart review as the 'gold standard' for comparison of transgender status, Ehrenfeld et al. found a low false positive rate for their best-performing algorithm (3%), though not as low as the false positive rate in the current study. The overall high levels of agreement for transgender measures in the two previous studies is likely a function of the lack of independence between the 'gold standard' and the CP or algorithm measures. Specifically, only those classified as transgender in the Medicare EHR data were offered survey participation to complete the two-step 'gold standard' survey measure, and only those cases identified as transgender in the university clinic EHR were included in chart review. Thus, previous studies could assess agreement between the two measures, but not robustly validate either. In the current study, the DTP provider-based transgender status measure is independent and thus could be used for robust CP validation.

While not possible to incorporate free-text records in case-finding algorithms in the current study as only structured EHR data is linked through COAST, it is worth noting the opportunities potentiated by use of NLP and machine learning approaches as methods for identifying transgender samples in EHR data as this research area continues to grow. Outside of transgender health, the use of NLP and machine learning to mine unstructured free-text EHR data has demonstrated efficiency in improving case ascertainment algorithm accuracy .[39] As 'gold standard' two-step sex assigned at birth and current gender identity measures of transgender status[12] are slowly being implemented in routinely collected healthcare data sources, in the meantime NLPs to extract free-text data can be used to produce better gold standards against which to measure algorithm performance, as demonstrated by the Medicare study.[38]

Transgender status prevalence & ascertainment

Based on a recent meta-analysis of transgender status prevalence in population-based probability samples,[10] it was expected that an effective CP would identify 0.4% of the general population as transgender, or approximately 54 of the HIV-positive COAST cohort (n=13 907) and 3,098 of

the HIV-negative cohort (n=516 340). Consistent with expectations, the best-performing CP identified 51 PLWH as transgender, equivalent to a transgender status prevalence of 0.4% among PLWH. Contrary to expectations, the best-performing CP identified less than 5% of the number of transgender persons expected in the HIV-negative cohort. This is likely a result of a number of factors including the limitation of CPs to the subset of a population accessing care as noted, and the result of most transgender people in BC receiving care currently outside the main fee-for-service healthcare delivery system. However, it is also consistent with the undercount of transgender populations using diagnostic criteria compared to other methods of ascertainment demonstrated in other studies.[14]

Using the broadest CP algorithm (any transgender-specific diagnosis ever, n=56) and those identified by provider-report together (total n=106), the total transgender PLWH sample would represent as high as 0.88% (range: 0.73-1.1%) of the prevalent HIV infections in BC in 2014.[40] This overrepresentation of transgender people among PLWH is consistent with evidence of a disproportionate HIV burden for transgender populations globally,[5,41,42] as well as in line with the only other available data on the proportion of PLWH who are transgender, from US national surveillance data (2012 data: 1.1%, 95%CI: 0.8-1.4).[43]

Demographics and chronic conditions

Despite moderate to low performance by some validation metrics, particularly low sensitivity, the CP was able to detect meaningful results in the characterization of demographics and chronic condition burden for the transgender sample - supporting CP face validity. The population density and age distribution by HIV-status of transgender people in this study is largely consistent with general population patterns, as well as the larger COAST cohort.[21,36] The overall higher burden of chronic illness for transgender people living with HIV versus without HIV in this study is consistent with elevated chronic illness risk and morbidity among non-transgender PLWH.[44] This higher chronic disease burden is linked to HIV disease processes and related inflammatory immune response.[45] While a small but growing number of studies have begun to investigate the chronic illness burden for transgender populations in other industrialized settings,[16,19,46–48] including using EHR data, findings vary widely due to differences in sampling, study design, setting and measurement.

Limitations

Findings from this study should be interpreted in the context of a few key limitations. CPs are by design only applicable to people accessing healthcare services, often motivated by illness and aided by the ability to access care. As such, this study is limited to those transgender people accessing medical transition care in BC and may only represent 24% to 47% of the total transgender population.[33] This study was also limited by the inability to validate the transgender CP among the HIV-negative COAST cohort, as a 'gold standard' provider-based transgender measure was only available for the HIV-positive cohort. It is possible that the transgender CPs would perform differently in populations living without HIV, particularly as healthcare contact is higher among populations living with HIV. Additionally, this study should be considered in light of the context in which it was conducted, an environment in which transgender healthcare delivery in BC is currently shifting from specialized care settings to the main primary care fee-for-service settings. Given that COAST only includes fee-for-service data, this study was limited by the inability to capture transgender people who access transgender care

outside the fee-for-service system. However, fortunately, as the shift to the fee-for-service system occurs, transgender ascertainment via CPs in BC will likely improve. The administrative data used in this study may also be susceptible to coding error (and coding biases/practices) across conditions and settings,[49] potentially introducing misclassification bias in terms of transgender ascertainment. Finally, chronic condition prevalence data reported in this study should be interpreted with caution, given potential selection bias by serostatus in the COAST cohort; though any such bias likely resulted in conservative estimates of difference by serostatus in this analysis.

CONCLUSION

This study makes a number of important contributions to the literature on innovative methods in transgender health. Major contributions include development and validation of a transgender CP, using a population-based representative source population, in the Canadian context. Another strength is the approximately complete ascertainment of the population of transgender PLWH in BC, and capacity to estimate transgender status prevalence among PLWH. In a current funding environment of limited support for longitudinal transgender health studies in the US and none to date in Canada, this study and the methods employed offer an efficient, replicable and cost-effective way forward in creating electronic cohorts for advancing transgender health research.[15] Moreover, the recent rollback of sexual orientation and gender identity data collection and legal changes in insurance coverage of transgender healthcare in the US potentiate decline in accurate claims coding for gender-affirming care.[30] This highlights the utility of work in this area from other jurisdictions, particularly those with transgender-inclusive universal healthcare systems such as Canada.

Future research should build upon the methods developed in this study and explore complimentary approaches for gender identity ascertainment in administrative and EHR data, such as machine learning approaches, as have been used to develop algorithms based on healthcare utilization data in other research areas. Finally, the current study lays the foundation for future work with the ability to study transgender health and healthcare use patterns over time, with linkage to laboratory data, as well as inclusion of appropriate comparison groups.[15,50]

ACKNOWLEDGEMENTS & DISCLAIMER

The authors thank the COAST study participants, the BC Centre for Excellence in HIV/AIDS, the BC Ministry of Health, BC Vital Statistics Agency, PharmaNet and the institutional data stewards for granting access to the data, and Population Data BC for facilitating the data linkage process. In addition, we would like to thank the COAST core team members and other support staff at these institutions for their administrative assistance with the data access and preparation. All inferences, opinions, and conclusions drawn in this paper are those of the authors, and do not reflect the opinions or policies of the Data Steward(s). There are no conflicts of interest to declare.

COMPETING INTERESTS

None declared.

FUNDING

This work was supported by the Canadian Institutes of Health Research, through an Operation Grant [grant number 130419], a Foundation Award to RSH [grant number 143342], and a Doctoral Research Award to AJR [grant number 152382] and support from the British Columbia Centre for Excellence in HIV/AIDS. The DTP receives funding from the provincial government of British Columbia (PharmaCare). The funders had no role in the study design, analysis, interpretation of the data, drafting of the manuscript or in the decision to submit for publication.

CONTRIBUTORS

AJR led the study from conceptualization to analysis plan to interpretation, drafting of the first manuscript version, revisions and final version. RSH acquired study data and funding. TP, MK, PS, TS, and RSH all contributed to study design, interpretation of results, and reviewed manuscript versions. JL and MY contributed to study analysis and reviewed manuscript versions. All authors provided critical review of first and subsequent manuscript drafts, approved the final version, and agree to be accountable for the work presented.

PATIENT CONSENT FOR PUBLICATION

Not required.

DATA SHARING STATEMENT

The data used for this study are held by the BC Centre for Excellence in HIV/AIDS under the authority of the BC Ministry of Health; as they contain confidential patient health records including HIV serostatus, data are cannot be made available to other parties.

FIGURES LEGENDS

Figure 1. Total transgender sample identified using a computable phenotype with electronic health records

Figure 2. Geographic distribution of transgender people across province, by health authority*

**% of transgender individuals with known health authority (n=182)*

Figure 3. Age distribution of transgender sample, by HIV serostatus

Figure 4. Co-morbidities among transgender sample, by HIV serostatus

REFERENCES

1. Bauer GR. Making sure everyone counts: Considerations for inclusion, identification, and analysis of transgender and transsexual participants in health surveys. What Differ Sex Gend Make Gend Sex Health Res Casebook. Vancouver, Canada: Canadian Institutes of Health Research (CIHR), Institute for Gender and Health (IGH); 2012. p. 59–67.

2. Bauer GR, Hammond R, Travers R, Kaay M, Hohenadel KM, Boyce M. "I don't think this is theoretical; this is our lives": How erasure impacts health care for transgender people. JANAC. 2009;20:348–361.

3. Poteat T, German D, Flynn C. The conflation of gender and sex: Gaps and opportunities in HIV data among transgender women and MSM. Glob Public Health. 2016;11:835–848.

4. Wanta JW, Unger CA. Review of the transgender literature: Where do we go from here? Transgender Health. 2017;2:119–128.

5. Herbst JH, Jacobs ED, Finlayson TJ, McKleroy VS, Neumann MS, Crepaz N, et al. Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a systematic review. AIDS Behav. 2008;12:1–17.

6. Becasen JS, Denard CL, Mullins MM, Higa DH, Sipe TA. Estimating the prevalence of HIV and sexual behaviors among the US transgender population: A systematic review and meta-analysis, 2006–2017. Am J Public Health. 2019;109:e1–8.

7. Marshall Z, Welch V, Thomas J, Brunger F, Swab M, Shemilt I, et al. Documenting research with transgender and gender diverse people: protocol for an evidence map and thematic analysis. Syst Rev. 2017;6:35.

8. Reisner SL, Poteat T, Keatley J, Cabral M, Mothopeng T, Dunham E, et al. Global health burden and needs of transgender populations: a review. The Lancet. 2016;388:412–436.

9. Cathcart-Rake EJ, Lightner DJ, Quevedo FJ, Ertz DC, Jatoi A. Cancer in transgender patients: One case in 385,820 is indicative of a paucity of data. J Oncol Pract. 2017;JOP2017027714.

10. Meerwijk EL, Sevelius JM. Transgender population size in the United States: A meta-regression of population-based probability samples. Am J Public Health. 2017;107:e1–e8.

11. Flores AR, Herman JL, Gates GJ, Brown TNT. How Many Adults Identify as Transgender in the United States? [Internet]. Los Angeles, CA: Williams Institute; 2016 Jun p. 13. Available from: https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf

12. The GenIUSS Group. Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys. Los Angeles, CA: The Williams Institute; 2014.

13. Bauer GR, Braimoh J, Scheim AI, Dharma C. Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. PLOS ONE. 2017;12:e0178043.

14. Collin L, Reisner SL, Tangpricha V, Goodman M. Prevalence of transgender depends on the "case" definition: A systematic review. J Sex Med. 2016;13:613–626.

15. Reisner SL, Deutsch MB, Bhasin S, Bockting W, Brown GR, Feldman J, et al. Advancing methods for US transgender health research. Curr Opin Endocrinol Diabetes Obes. 2016;23:198–207.

16. Ehrenfeld JM, Gottlieb KG, Beach LB, Monahan SE, Fabbri D. Development of a Natural Language Processing Algorithm to Identify and Evaluate Transgender Patients in Electronic Health Record System. Ethn Dis. 2019;29:441–50.

17. Richesson R, Smerek M, Rusincovitch S, Zozus MN, Chaudhuri PS, Hammond WE, et al. Electronic Health Records-Based Phenotyping. In: Uhlenbrauck G, editor. NIH Collab Living Textb Pragmatic Clin Trials. Bethesda, Maryland; 2014.

18. Quinn VP, Nash R, Hunkeler E, Contreras R, Cromwell L, Becerra-Culqui TA, et al. Cohort profile: Study of Transition, Outcomes and Gender (STRONG) to assess health status of transgender people. BMJ Open. 2017;7:e018121.

19. Blosnich JR, Brown GR, Shipherd Phd JC, Kauth M, Piegari RI, Bossarte RM. Prevalence of gender identity disorder and suicide risk among transgender veterans utilizing veterans health administration care. Am J Public Health. 2013;103:e27–32.

20. Proctor K, Haffer SC, Ewald E, Hodge C, James CV. Identifying the transgender population in the Medicare program. Transgender Health. 2016;1:250–265.

21. Eyawo O, Hull MW, Salters K, Samji H, Cescon A, Sereda P, et al. Cohort profile: the Comparative Outcomes And Service Utilization Trends (COAST) Study among people living with and without HIV in British Columbia, Canada. BMJ Open. 2018;8:e019115.

22. Drug Treatment Program [Internet]. BC Cent. Excell. HIVAIDS. 2010 [cited 2019 Sep 27]. Available from: http://www.cfenet.ubc.ca/drug-treatment-program

23. Population Data BC. About PopData [Internet]. Popul. Data BC. 2017 [cited 2019 Sep 27]. Available from: https://www.popdata.bc.ca/aboutus

24. British Columbia Ministry of Health (2014). Medical Services Plan (MSP) Payment Information File [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

25. British Columbia Ministry of Health (2014). Consolidation File (MSP Registration & Premium Billing) [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

26. Canadian Institute for Health Information (2014). Discharge Abstract Database (Hospital Separations) [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

27. British Columbia Ministry of Health (2014). PharmaCare [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

28. British Columbia Ministry of Health (2014). PharmaNet [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

29. British Columbia Vital Statistics Agency (2014). Vital Statistics Deaths [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca

30. Arnold JD, Nelson AE, Loubier EM. Trends in Insurance Coverage for Gender-Affirming Surgeries. JAMA Surg. 2018;153:972.

31. Radix A, Sevelius J, Deutsch MB. Transgender women, hormonal therapy and HIV treatment: a comprehensive review of the literature and recommendations for best practices. J Int AIDS Soc. 2016;19:20810.

32. Rotondi NK, Bauer GR, Scanlon K, Kaay M, Travers R, Travers A. Nonprescribed hormone use and self-performed surgeries: "do-it-yourself" transitions in transgender communities in Ontario, Canada. Am J Public Health. 2013;103:1830–1836.

33. Scheim AI, Bauer GR. Sex and gender diversity among transgender persons in Ontario, Canada: results from a respondent-driven sampling survey. J Sex Res. 2015;52:1–14.

34. Chronic Disease Information Working Group. BC Chronic Disease and Selected Procedure Case Definitions [Internet]. British Columbia, Canada: BC Ministry of Health; 2019 Apr p. 1–132. Available from: https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf

35. British Columbia Cancer Agency. BC Cancer Registry Data [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

36. BC Stats. Sub-provincial Population Estimates [Internet]. 2018 [cited 2019 Feb 1]. Available from: file://localhost/Users/AJRich/Zotero/storage/5VM2LBUN/PopulationEstimates.html

37. Hu W. Diagnostic Codes in MSP Claim Data. Program Monitoring and Information Management Branch, Resource Management Division, Medical Services Plan; 1996 p. 39.

38. Gerth J, Becerra-Culqui T, Bradlyn A, Getahun D, Hunkeler EM, Lash TL, et al. Agreement between medical records and self-reports: Implications for transgender health research. Rev Endocr Metab Disord. 2018;19:263–269.

39. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al. Development of phenotype algorithms using electronic medical records and incorporating natural

language processing. The BMJ [Internet]. 2015 [cited 2020 Oct 22];350. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4707569/

40. BC Centre for Disease Control. HIV in British Columbia: Annual Surveillance Report 2015. Vancouver, Canada: BC Centre for Disease Control; 2017.

41. Poteat T, Scheim A, Xavier J, Reisner S, Baral S. Global epidemiology of HIV infection and related syndemics affecting transgender people. J Acquir Immune Defic Syndr. 2016;72 Suppl 3:S210–9.

42. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV in transgender women: a systematic review and meta-analysis. Lancet Infect Dis. 2013;13:214–222.

43. Centers for Disease Control and Prevention. Behavioral and Clinical Characteristics of Persons Receiving Medical Care for HIV Infection—Medical Monitoring Project, United States, 2012 [Internet]. Washington, D.C.: Centers for Disease Control; 2015 Aug p. 42. Report No.: 12. Available from: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf

44. Grabovac I, Veronese N, Stefanac S, Haider S, Jackson SE, Koyanagi A, et al. Human Immunodeficiency Virus Infection and Diverse Physical Health Outcomes: An Umbrella Review of Meta-analyses of Observational Studies. Clin Infect Dis. 2019;ciz539.

45. Nasi M, De Biasi S, Gibellini L, Bianchini E, Pecorini S, Bacca V, et al. Ageing and inflammation in patients with HIV infection: Ageing, inflammation and HIV infection. Clin Exp Immunol. 2017;187:44–52.

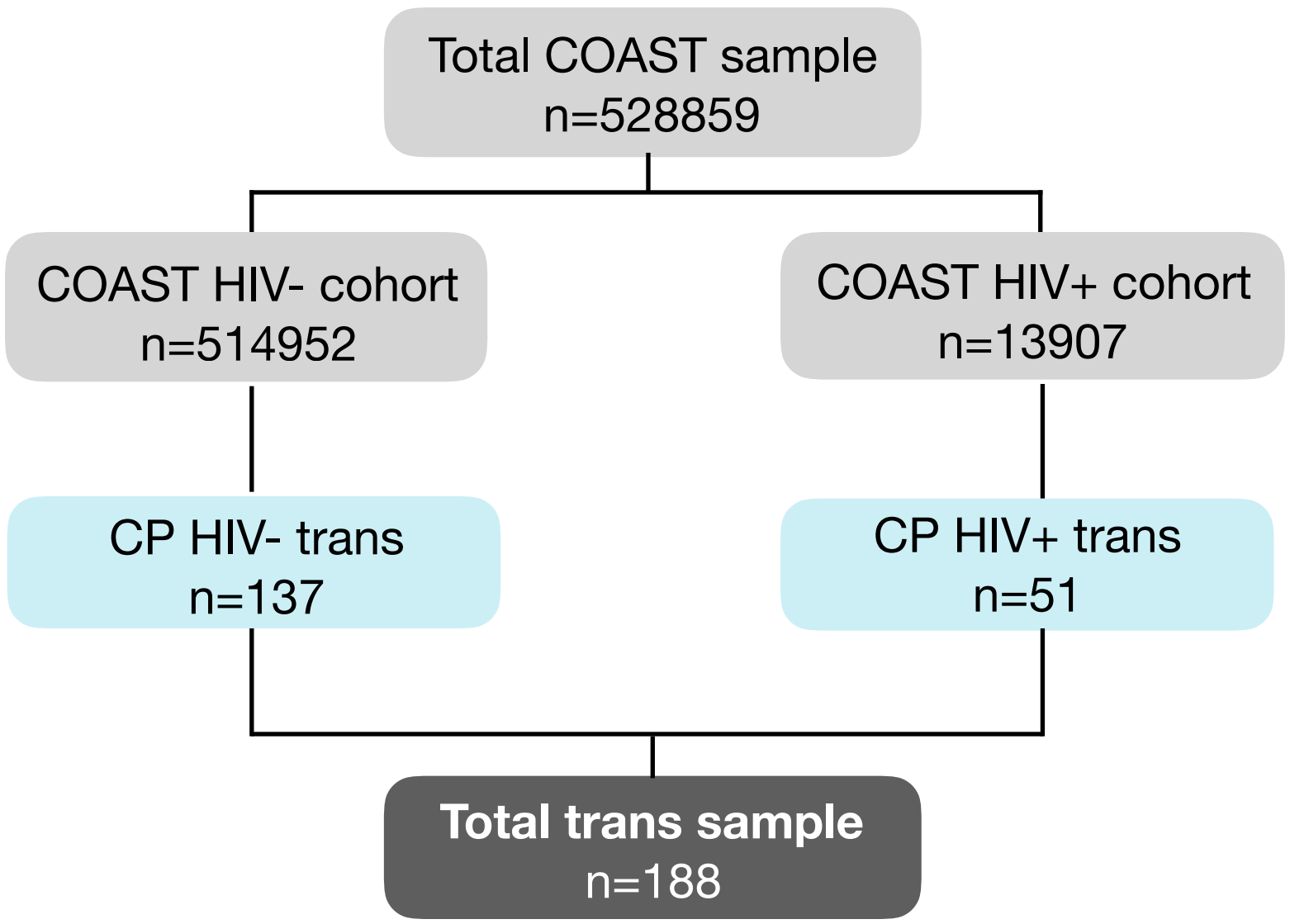
46. Judge C., O'Donovan C., Callaghan G., Gaoatswe G., O'Shea D. Gender dysphoria - prevalence and co-morbidities in an Irish adult population. Front Endocrinol. 2014;5:87.

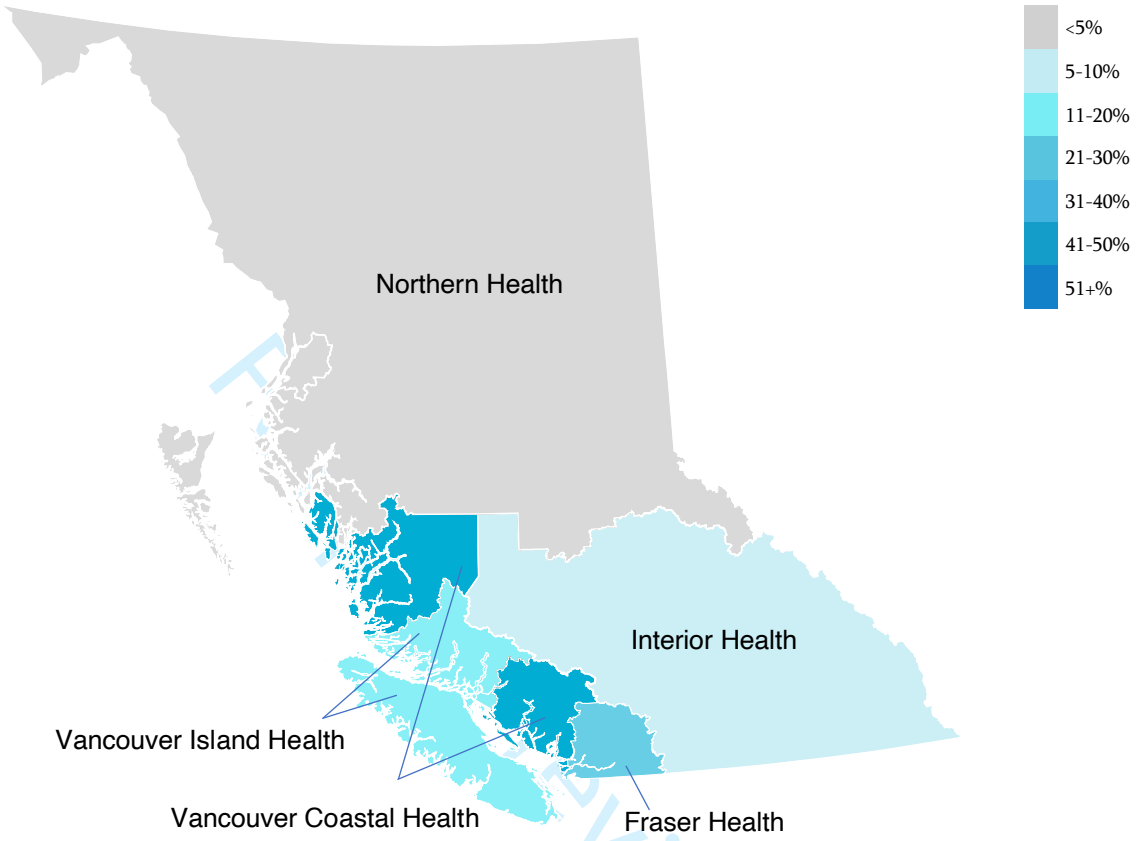
47. Cheung AS, Ooi O, Leemaqz S, Cundill P, Silberstein N, Bretherton I, et al. Sociodemographic and clinical characteristics of transgender adults in Australia. Transgender Health. 2018;3:229–38.

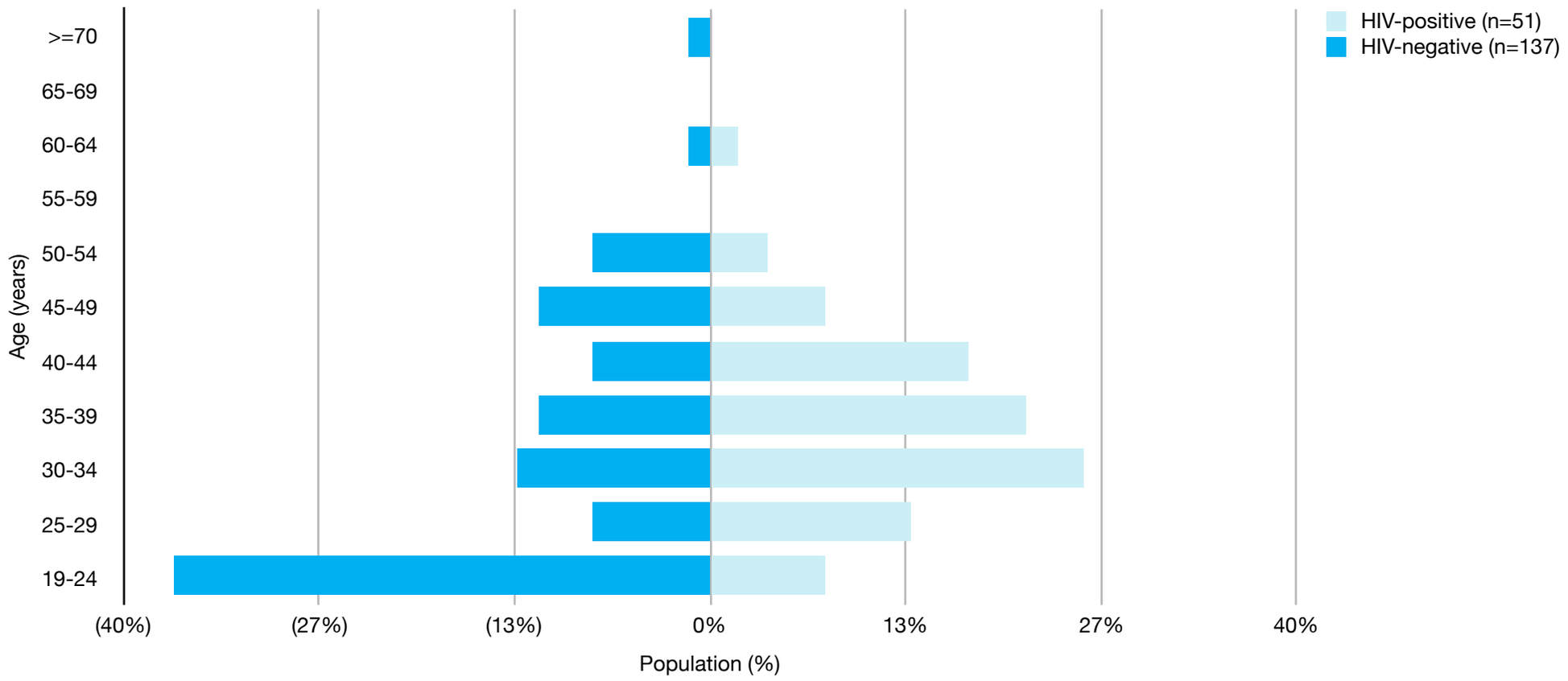
48. Swartz JA, Ducheny K, Holloway T, Stokes L, Willis S, Kuhns LM. A latent class analysis of chronic health conditions among HIV-Positive transgender women of color. AIDS Behav [Internet]. 2019 [cited 2019 Oct 2]; Available from: http://link.springer.com/10.1007/s10461-019-02543-3

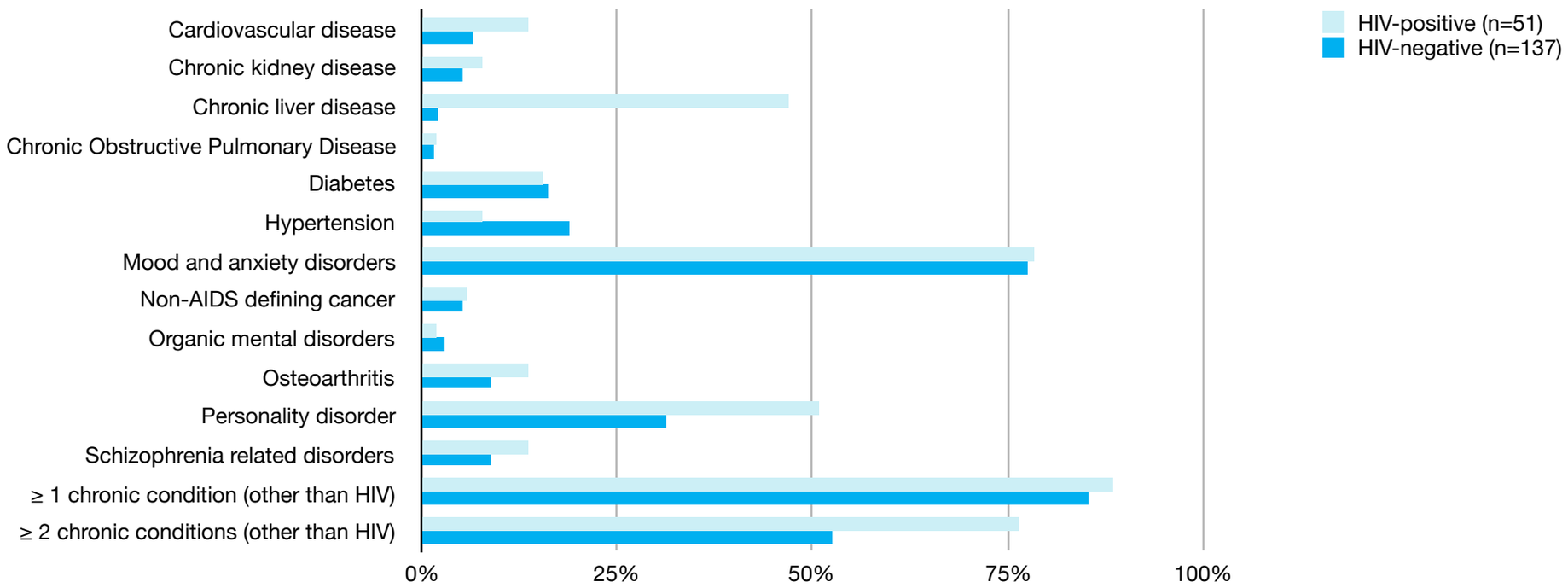
49. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 2011;

50. Reisner SL, Radix A, Deutsch MB. Integrated and gender-affirming transgender clinical care and research. J Acquir Immune Defic Syndr. 2016;72 Suppl 3:S235–42.








Cardiovascular disease
Chronic kidney disease
Chronic liver disease
Chronic Obstructive Pulmonary Disease
Diabetes
Hypertension
Mood and anxiety disorders
Non-AIDS defining cancer
Organic mental disorders
Osteoarthritis
Personality disorder
Schizophrenia related disorders
≥ 1 chronic condition (other than HIV)
≥ 2 chronic conditions (other than HIV)
0%
25%
50%
75%
100%
HIV-positive (n=51)
HIV-negative (n=137)

Supplementary material

Data sources and description of data elements

Data source	Data steward	Description	Data elements provided for this study
Medical Services Plan	British Columbia Ministry of Health	For individuals covered by the provincial universal health insurance plan- Medically necessary services provided by fee-for-service physicians and other healthcare providers, laboratory services, diagnostic procedures, dental/oral surgery	ICD-9 and ICD-10 codes
Consolidation file	British Columbia Ministry of Health	Demographic data on individuals receiving or registered to receive care in BC, pooled from multiple PopData sources	Sociodemographics
Discharge Abstract Database	British Columbia Ministry of Health	Demographic, administrative and clinical data for inpatient hospital discharges and day surgeries	ICD-9 and ICD-10 codes
Vital statistics deaths	British Columbia Vital Statistics Agency	Records of all registered deaths in BC	Death data
Pharmacare	British Columbia Ministry of Health	Data related to prescription drugs dispensed under the BC public drug insurance program.	Prescription drug name/identifier
Pharmanet	British Columbia Ministry of Health Data Stewardship Committee	Data related to prescription drugs dispensed by community and outpatient pharmacies	Prescription drug name/identifier

Supplementary material

Drug Treatment Program and laboratory	British Columbia Centre for Excellence in HIV/AIDS	Antiretroviral therapy use history, laboratory testing, immunological and virologic testing, and demographic data on PLWH who have accessed antiretrovirals in BC	Providers-reported transgender status, laboratory confirmed HIV serostatus

Prescription drugs with drug identification numbers (DIN)s

	Generic Name	DIN
Transfeminine		
Androgen Blockers		
Spironolactone		
	SPIRONOLACTONE	28606
	SPIRONOLACTONE	613215
	SPIRONOLACTONE	285455
	SPIRONOLACTONE	613223
	SPIRONOLACT/HYDROCHLOROTHIAZID	180408
	SPIRONOLACT/HYDROCHLOROTHIAZID	613231
	SPIRONOLACT/HYDROCHLOROTHIAZID	594377
	SPIRONOLACT/HYDROCHLOROTHIAZID	657182
Cyproterone		
	ETHINYL ESTRADIOL/CYPROTERONE	2233542
	NO GENERIC FORMULARY	634514
	CYPROTERONE ACETATE	704431
	CYPROTERONE ACETATE	2229449
	CYPROTERONE ACETATE	2229723
	CYPROTERONE ACETATE	2232872
	CYPROTERONE ACETATE	2245898
	CYPROTERONE ACETATE	704423
Finasteride		
	FINASTERIDE	2010909
	FINASTERIDE	2238213
Dutasteride		
	DUTASTERIDE	2247813
Estrogens		
Estrogen		
	ESTROGENS,CONJUGATED	830240
	ESTROGENS,CONJUGATED	831395

Supplementary material

ETHYNODIOL D-ETHINYL ESTRADIOL	28630
ETHYNODIOL D-ETHINYL ESTRADIOL	469327
NORETHINDRONE-MESTRANOL	22608
NORETHINDRONE-MESTRANOL	22659
NORETHINDRONE A-E ESTRADIOL	297143
NORETHINDRONE A-E ESTRADIOL	315966
NORETHINDRONE-ETHINYL ESTRAD	317047
NORETHINDRONE-ETHINYL ESTRAD	372846
NORETHINDRONE-ETHINYL ESTRAD	373265
NORETHINDRONE-ETHINYL ESTRAD	531006
NORETHINDRONE-ETHINYL ESTRAD	538590
NORETHINDRONE-ETHINYL ESTRAD	602957
NORETHINDRONE-ETHINYL ESTRAD	620947
NORETHINDRONE-ETHINYL ESTRAD	2187086
NORETHINDRONE-ETHINYL ESTRAD	2187108
NORETHINDRONE-ETHINYL ESTRAD	2189054
NORGESTREL-ETHINYL ESTRADIOL	34207
NORGESTREL-ETHINYL ESTRADIOL	300640
LEVONORGESTREL-ETH ESTRA	579386
LEVONORGESTREL-ETH ESTRA	707600
LEVONORGESTREL-ETH ESTRA	782416
LEVONORGESTREL-ETH ESTRA	782432
LEVONORGESTREL-ETH ESTRA	2042320
NORGESTREL-ETHINYL ESTRADIOL	2043033
LEVONORGESTREL-ETH ESTRA	2043726
NORGESTIMATE-ETHINYL ESTRADIOL	2258560
NORETHINDRONE-MESTRANOL	30333
NORETHINDRONE-MESTRANOL	30341
LEVONORGESTREL-ETH ESTRA	2236974
ETHYNODIOL D-ETHINYL ESTRADIOL	471526
NORETHINDRONE-ETHINYL ESTRAD	340731
NORETHINDRONE-MESTRANOL	340758
NORETHINDRONE A-E ESTRADIOL	343838
NORETHINDRONE A-E ESTRADIOL	353027
NORETHINDRONE-ETHINYL ESTRAD	372838
NORETHINDRONE-ETHINYL ESTRAD	373273
NORETHINDRONE-ETHINYL ESTRAD	531014
NORETHINDRONE-ETHINYL ESTRAD	602965
NORETHINDRONE-ETHINYL ESTRAD	695734
NORETHINDRONE-ETHINYL ESTRAD	2187094

Supplementary material

NORETHINDRONE-ETHINYL ESTRAD	2187116
NORETHINDRONE-ETHINYL ESTRAD	2189062
ETHINYL ESTRADIOL/NORETH AC	2242531
NORGESTREL-ETHINYL ESTRADIOL	340766
NORGESTREL-ETHINYL ESTRADIOL	342815
LEVONORGESTREL-ETH ESTRA	586609
LEVONORGESTREL-ETH ESTRA	707503
LEVONORGESTREL-ETH ESTRA	782424
LEVONORGESTREL-ETH ESTRA	782440
LEVONORGESTREL-ETH ESTRA	2042339
NORGESTREL-ETHINYL ESTRADIOL	2043041
LEVONORGESTREL-ETH ESTRA	2043734
NORGESTIMATE-ETHINYL ESTRADIOL	2258587
LEVONORGESTREL-ETH ESTRA	2236975
NORGESTIMATE-ETHINYL ESTRADIOL	1968440
NORGESTIMATE-ETHINYL ESTRADIOL	2028700
NORGESTIMATE-ETHINYL ESTRADIOL	1992872
NORGESTIMATE-ETHINYL ESTRADIOL	2029421
DESOGESTREL-ETHINYL ESTRADIOL	2042487
DESOGESTREL-ETHINYL ESTRADIOL	2042541
DESOGESTREL-ETHINYL ESTRADIOL	2042479
DESOGESTREL-ETHINYL ESTRADIOL	2042533
ESTRADIOL/NORETH AC	2241835
ESTRADIOL/NORETH AC	2241837
LEVONORGESTREL	2241674
ESTROGEN,CON/M-PROGEST ACET	2242878
ESTROGEN,CON/M-PROGEST ACET	2242879
ESTRADIOL/NORETH AC	2243529
ESTRADIOL/NORETH AC	2243530
ETHINYL ESTRADIOL/DROSPIRENONE	2261723
ETHINYL ESTRADIOL/DROSPIRENONE	2261731
ETONOGESTREL/ETHINYL ESTRADIOL	2253186
ETHINYL ESTRADIOL/NORELGEST	2248297
DIENESTROL	441295
DIETHYLSTILBESTROL	3360
DIETHYLSTILBESTROL	2091461
DIETHYLSTILBESTROL	2091488
ESTRADIOL	464791
ESTRADIOL	2148587
ESTRADIOL	464805

Supplementary material

ESTRADIOL	2148595
ESTRADIOL VALERATE	29238
ESTRADIOL	756849
ESTRADIOL	2237807
ESTRADIOL	2243722
ESTRADIOL	2245676
ESTRADIOL	756857
ESTRADIOL	2204428
ESTRADIOL	2231509
ESTRADIOL	2237808
ESTRADIOL	2243724
ESTRADIOL	2244000
ESTRADIOL	2246967
ESTRADIOL	756792
ESTRADIOL	2204444
ESTRADIOL	2231510
ESTRADIOL	2244002
ESTRADIOL	2246969
ESTRADIOL	2168898
ESTRADIOL	2204436
ESTRADIOL	2244001
ESTRADIOL	2246968
ESTRADIOL	2225190
ESTRADIOL	2204401
ESTRADIOL	2238704
ESTRADIOL	2243999
ESTRADIOL	2241332
ESTRADIOL	2247499
ESTRADIOL	2247500
ESTROGENS,CONJUGATED	2569
ESTROGENS,CONJUGATED	2043394
ESTROGENS,CONJUGATED	2230891
ESTROGENS,CONJUGATED	2239654
ESTROGENS,CONJUGATED	2577
ESTROGENS,CONJUGATED	265470
ESTROGENS,CONJUGATED	587281
ESTROGENS,CONJUGATED	2043408
ESTROGENS,CONJUGATED	2089
ESTROGENS,CONJUGATED	2043440
ESTROGENS,CONJUGATED	403466

Supplementary material

	ESTROGENS,CONJUGATED	2043416
	ESTROGENS,CONJUGATED	2230892
	ESTROGENS,CONJUGATED	2239655
	ESTROGENS,CONJUGATED	2585
	ESTROGENS,CONJUGATED	265489
	ESTROGENS,CONJUGATED	587303
	ESTROGENS,CONJUGATED	2043424
	ESTROGENS,CONJUGATED	2043432
	ESTROGENS,CONJUGATED	2043386
	ME-TESTOSTERONE/ESTROGEN,CON	53538
	ESTROPIPATE	282685
	ESTROPIPATE	2089769
	ESTROPIPATE	282677
	ESTROPIPATE	2089777
	ESTROPIPATE	2089793
	ESTRADIOL/NORETH AC	2108186
	NORGESTIMATE-ETHINYL ESTRADIOL	2229064
	NORGESTIMATE-ETHINYL ESTRADIOL	2229218
	NORGESTIMATE-ETHINYL ESTRADIOL	2229226
	ETHINYL ESTRADIOL/CYPROTERONE	2233542
	ETHINYL ESTRADIOL/NORELGEST	2246340
	NO GENERIC FORMULARY	66124057
	NO GENERIC FORMULARY	66124058
	NO GENERIC FORMULARY	66124060
	NO GENERIC FORMULARY	66124061
	NO GENERIC FORMULARY	66124062
	NO GENERIC FORMULARY	66124063
	NO GENERIC FORMULARY	66124064
Progestogens		
Progesterone		
	PROGESTERONE,MICRONIZED	2241013
	MEDROXYPROGESTERONE ACET	30848
	MEDROXYPROGESTERONE ACET	30856
	MEDROXYPROGESTERONE ACET	585092
	NO GENERIC FORMULARY	66123240
	MEDROXYPROGESTERONE ACET	708917
	MEDROXYPROGESTERONE ACET	2148552
	MEDROXYPROGESTERONE ACET	2221284
	MEDROXYPROGESTERONE ACET	2229838
	MEDROXYPROGESTERONE ACET	2244726

Supplementary material

	MEDROXYPROGESTERONE ACET	2246627
	MEDROXYPROGESTERONE ACET	30937
	MEDROXYPROGESTERONE ACET	2010739
	MEDROXYPROGESTERONE ACET	2148560
	MEDROXYPROGESTERONE ACET	2221292
	MEDROXYPROGESTERONE ACET	2229839
	MEDROXYPROGESTERONE ACET	2244727
	MEDROXYPROGESTERONE ACET	2246628
	MEDROXYPROGESTERONE ACET	729973
	MEDROXYPROGESTERONE ACET	2010933
	MEDROXYPROGESTERONE ACET	2148579
	MEDROXYPROGESTERONE ACET	2221306
	MEDROXYPROGESTERONE ACET	2229840
	MEDROXYPROGESTERONE ACET	2246629
	MEDROXYPROGESTERONE ACET	30945
	MEDROXYPROGESTERONE ACET	2267640
	NORETHINDRONE	37605
	PROGESTERONE,MICRONIZED	2166704
	PROGESTERONE	739952
	PROGESTERONE	1977652
	PROGESTERONE	2128470
	LEVONORGESTREL	2243005
Transmasculine		
Testosterone		
	TESTOSTERONE	2249499
	TESTOSTERONE CYPIONATE	30783
	TESTOSTERONE PROPIONATE	1977571
	TESTOSTERONE CYPIONATE	1977601
	TESTOSTERONE CYPIONATE	2220318
	TESTOSTERONE CYPIONATE	2246063
	TESTOSTERONE ENANTHATE	29246
	TESTOSTERONE ENANTHATE	716936
	TESTOSTERONE ENANTHATE	739944
	TESTOSTERONE UNDECANOATE	782327
	TESTOSTERONE ENANTHATE/ESTRAD	108278
	TESTOSTERONE ENANTHATE/ESTRAD	2061031
	TESTOSTERONE	2239653
	TESTOSTERONE	2245346
	TESTOSTERONE	2245345
	TESTOSTERONE	2245972

Supplementary material

Chronic condition case definitions

Chronic condition		Case definition	Codes
Cardiovascular disease*	Acute myocardial infarction	1 or more hospitalizations with relevant ICD codes	ICD-10: I21 Acute myocardial infarction I22 Subsequent myocardial infarction ICD-9: 410 Acute myocardial infarction
	Ischemic heart disease	At least one of the following: 2 medical visits with Angina ICD-9 code 413 plus 1 heart disease prescription in 365 days; or 1 specialist visit with Angina ICD-9 code 413 plus 1 prescription in 365 days; or 2 medical visits with two ICD9 codes 410, 411, 412, 413, 414 in 365 days; or 1 CCI/CCP CABG,PCI/PCTA procedure code; or 1 hospitalization with relevant ICD code.	ICD-10: I20 Angina pectoris I21 Acute myocardial infarction I 22 Subsequent myocardial infarction I23 Certain current complications following acute myocardial infarction I24 Other acute ischaemic heart diseases I25 Chronic ischaemic heart disease ICD-9: 410 Acute myocardial infarction 411 Other acute and subacute forms of ischaemic heart disease 412 Old myocardial infarction 413 Angina pectoris 414 Other forms of chronic ischaemic heart disease
	Chronic heart failure	1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	ICD-10: I50 Heart failure ICD-9: 428 Heart failure
	Stroke- hospital	1 or more hospitalizations with relevant ICD codes	ICD-10: H34.1 Central retinal artery occlusion I60 Subarachnoid hemorrhage I61 Intracerebral haemorrhage I63 Cerebral infarction (exclude I63.6 Cerebral infarction due to

Supplementary material

			cerebral venous thrombosis, nonpyogenic) I64 Stroke, not specified as haemorrhage or infarction 362.3 Retinal vascular occlusion 430 Subarachnoid hemorrhage 431 Intracerebral hemorrhage 433.x1 Occlusion and stenosis of precerebral arteries 434.x Occlusion cerebral arteries 436 Acute but ill-defined cerebrovascular disease *Excludes any traumatic brain injury*
	Transient ischemic attack	1 or more hospitalizations with relevant ICD codes	ICD-10: H34.0 Transient retinal artery occlusion G45.0 Vertebro-basilar artery syndrome G45.1 Carotid artery syndrome (hemispheric) G45.2 Multiple and bilateral precerebral artery syndromes G45.3 Amaurosis fugax G45.8 Other transient cerebral ischemic attacks and related syndromes G45.9 Transient cerebral ischemic attack, unspecified ICD-9: 435 Transient cerebral ischemia *Excludes any traumatic brain injury*
Chronic kidney disease*		1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	ICD-10: N01 Rapidly progressive nephritic syndrome N03 Chronic nephritic syndrome N04 Nephrotic syndrome N05 Unspecified nephritic syndrome N06 Isolated proteinuria with specified morphological lesion N07 Hereditary nephropathy, not elsewhere classified

Supplementary material

		N18 Chronic kidney disease N19 Unspecified kidney failure N26 Unspecified contracted kidney N27 Small kidney of unknown cause ICD-9: 581 Nephrotic syndrome 582 Chronic glomerulonephritis 583 Nephritis and nephropathy, not specified as acute or chronic 585 Chronic renal failure 586 Renal failure, unspecified 587 Renal sclerosis, unspecified 589 Small kidney of unknown cause
Chronic liver disease	1 or more hospitalization or medical visit with relevant diagnosis within 365 days	ICD-9: 571.0 Alcoholic fatty liver 571.2 Alcoholic cirrhosis of liver 571.3 Alcoholic liver damage, unspecified 571.4 Chronic hepatitis 571.5 Cirrhosis of liver without mention of alcohol 571.6 Billiary cirrhosis 571.8 Other chronic nonalcoholic liver disease 571.9 Unspecified chronic liver disease without mention of alcohol 070.3 Viral hepatitis B without mention of hepatic coma 070.30 Viral hepatitis B without mention of hepatic coma, acute or unspecified, without mention of hepatitis delta 070.31 Viral hepatitis B without mention of hepatic coma, acute or unspecified, with hepatitis delta 070.32 Viral hepatitis B without mention of hepatic coma, chronic, without mention of hepatitis delta 070.33 Viral hepatitis B without mention of hepatic coma, chronic, with hepatitis delta

Supplementary material

		070.52 Hepatitis delta without mention of active Hepatitis B disease or hepatic coma V02.61 Hepatitis B carrier 070.42 Hepatitis delta without mention of active Hepatitis B disease with hepatic coma 070.54 Chronic hepatitis C without mention of hepatic coma V02.62 Hepatitis C carrier
Chronic Obstructive Pulmonary Disease*	1 or more hospitalization or 2 or more medical visits within 365 days	ICD-10: J41 Simple and mucopurulent chronic bronchitis J42 Unspecified chronic bronchitis J43 Emphysema J44 Other chronic obstructive pulmonary disease ICD-9: 491 Chronic bronchitis 492 Emphysema 496 Chronic airways obstruction, not elsewhere classified
Diabetes Mellitus*	At least 1 of the following: 1 hospitalization or 2 medical visits in 365 days with relevant ICD codes; or 2 or more insulin prescriptions in 365 days; or 2 or more oral antihyperglycemic (not including metformin) prescriptions in 365 days; or 1 insulin and 1 oral antihyperglycemic (including metformin) in 365 days; or 2 metformin prescriptions and 1 medical visit in one year with relevant ICD codes.	ICD-10: E10 Type 1 diabetes mellitus E11 Type 2 diabetes mellitus E13 Other specified diabetes mellitus E14 Unspecified diabetes mellitus ICD-9: 250 Diabetes mellitus

Supplementary material

	Excludes gestational diabetes.	
Hypertension*	1 or more hospitalizations or 2 or more medical visits within 2 years with relevant ICD codes. *Excludes gestational hypertension.*	ICD-10: I10 Essential (primary) hypertension I11 Hypertensive heart disease I12 Hypertensive renal disease I13 Hypertensive heart and renal disease I15 Secondary hypertension ICD-9: 401 Essential hypertension 402 Hypertensive heart disease 403 Hypertensive renal disease 404 Hypertensive heart and renal disease 405 Secondary hypertension
Mood and anxiety disorders*	1 or more hospitalizations with a relevant ICD code or 2 or more medical visits with a relevant code within 2 years	ICD-10: F30 Manic episode F31 Bipolar affective disorder F32 Depressive episode F33 Recurrent depressive disorder F34 Persistent mood [affective] disorders F38 Other mood [affective] disorders F39 Unspecified mood [affective] disorder F40 Phobic anxiety disorders F41 Other anxiety disorders F42 Obsessive-compulsive disorder F43 Reaction to severe stress, and adjustment disorders F44 Dissociative (conversion) disorders F45 Somatoform disorders F48 Other neurotic disorders F68 Other disorders of adult personality & behavior ICD-9:

Supplementary material

			296 Affective psychoses 300 Neurotic disorders 311 Depressive disorder, not elsewhere classified MSP DX Code: 50B Anxiety/Depression
Non-AIDS defining cancer†	All prevalent cancer cases were included, with the exception of AIDS defining malignancies (Kaposi's sarcoma, non-Hodgkin's lymphoma, invasive cervical cancer)		*Cancer case definition details available from the British Columbia Cancer Agency: http://www.bccancer.bc.ca/health-info/types-of-cancer*
Organic mental disorders		1 or more medical visits or hospitalizations with relevant diagnoses within 365 days	ICD-9: 290.x Dementias 294.x Other organic psychotic conditions 331.x Alzheimer's ICD-10: F00.x Dementia in Alzheimer's disease F01.x Vascular Dementia F02.x Dementia in other disease classified elsewhere F03.x Unspecified dementia F04 Amnestic disorder due to physiological condition F06 Other mental disorders due to known physiological condition F09 Unspecified mental disorder due to known physiological condition G30 Alzheimer's disease with early onset
Osteoarthritis*		1 or more hospitalization or 2 or more medical visits in 365 days with a relevant ICD code	ICD-10: M15 Polyarthrosis M16 Coxarthrosis [arthrosis of hip] M17 Gonarthrosis [arthrosis of knee] M18 Arthrosis of first carpometacarpal joint M19 Other arthrosis

Supplementary material

		ICD-9: 715 Osteoarthrosis and allied disorders
Personality disorder	1 or more hospitalizations or medical visits with a relevant diagnosis within 365 days	ICD-9: 301.x Personality disorders ICD-10: F60.x Specified personality disorders F62 Enduring personality changes, not attributable to brain damage and disease F68.1 Intentional production or feigning of symptoms or disabilities, either physical or psychological F68.8 Other specified disorders or adult personality and behaviour F69 Unspecified disorder or adult personality and behaviour
Schizophrenia related disorder	1 or more medical visit or hospitalizations with relevant diagnoses within 365 days	ICD-9: 295.x Schizophrenic disorders 297.0 Paranoid state, simple 297.1 Delusional disorder 297.2 Paraphrenia 297.3 Shared psychotic disorder ICD-10: F20.x Paranoid schizophrenia F21.x Schizotypal disorder F23.2 Acute schizophrenia-like psychotic disorder F25.x Schizoaffective disorders

* Case definition adapted from British Columbia Ministry of Health version 2017, April 4 2019 update

† Case-definition adapted from British Columbia Cancer Agency

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	1.1-3: Title page
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			Introduction (pp 4-5)
Objectives	3	State specific objectives, including any prespecified hypotheses			Introduction (page 5)
Methods					
Study Design	4	Present key elements of study design early in the paper			Methods (page 5); Supplementary Material
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			Methods (page 5)

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	6.1: Methods (pp 6-7); Supplementary Material 6.2: Methods (pp 6-7) 6.3: Figure 1
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	7.1: Methods (pp 6-7); Supplementary Material
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			Methods (pp 6-7)

Bias	9	Describe any efforts to address potential sources of bias			Discussion (pp 9, 11)
Study size	10	Explain how the study size was arrived at			Methods (page 6); Figure 1
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Methods (pp 6-7)
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			Methods (pp 6-7)
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	12.1-2: Methods

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Methods (page 5)
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Results (page 8); Figure 1
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			Results (page 8)
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure			Results

		category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			NA
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			NA
Discussion					
Key results	18	Summarise key results with reference to study objectives			Discussion: Page 9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Discussion: Page 9, Page 11
Interpretation	20	Give a cautious overall interpretation of results considering objectives,			Conclusion (Page 11)

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results			Limitations (Page 9)
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			Funding (Page 12)
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Page 12

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

BMJ Open

Development of a computable phenotype to identify a transgender sample for health research purposes: A feasibility study in a large linked provincial healthcare administrative cohort in British Columbia, Canada

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-040928.R2
Article Type:	Original research
Date Submitted by the Author:	06-Feb-2021
Complete List of Authors:	Rich, Ashleigh; The University of British Columbia, School of Population & Public Health; BC Centre for Excellence in HIV/AIDS, Poteat, Tonia; University of North Carolina at Chapel Hill, Department of Social Medicine Koehoorn, Mieke; University of British Columbia, School of Population and Public Health; Li, Jenny; BC Centre for Excellence in HIV/AIDS, Ye, Monica; BC Centre for Excellence in HIV/AIDS, Sereda, Paul; British Columbia Centre for Excellence in HIV/AIDS, Salway, Travis; Simon Fraser University Hogg, R; Simon Fraser University,
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	HIV/AIDS, Health services research, Research methods, Public health
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, Sexual and gender disorders < PSYCHIATRY, HIV & AIDS < INFECTIOUS DISEASES, SOCIAL MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

TITLE PAGE

Title: Development of a computable phenotype to identify a transgender sample for health research purposes: A feasibility study in a large linked provincial healthcare administrative cohort in British Columbia, Canada

Authors: Rich AJ[1,2], Poteat T[3], Koehoorn M[1], Li J[2], Ye M[2], Sereda, P[2], Salway T[4], Hogg RS[2,4]

Affiliations:
1. School of Population and Public Health, University of British Columbia, Vancouver, Canada
2. British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada
3. Department of Social Medicine, University of North Carolina- Chapel Hill, Chapel Hill, USA
4. Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada

Corresponding author:
Ashleigh J Rich
ajrich@mail.ubc.ca

2206 East Mall
Vancouver, BC
V6T 1Z3
Canada

ABSTRACT

Objectives: Innovative methods are needed for identification of transgender people in administrative records for health research purposes. This study investigated the feasibility of using transgender-specific healthcare utilization in a Canadian population-based, health records database to develop a computable phenotype (CP) and identify the proportion of transgender people within the HIV-positive population as a public health priority.

Design: The COAST cohort comprises a data linkage between two provincial data sources: The BC Centre for Excellence in HIV/AIDS Drug Treatment Program, which coordinates HIV treatment dispensation across BC; and Population Data BC, a provincial data repository holding individual, longitudinal data for all BC residents (1996-2013).

Setting: British Columbia, Canada.

Participants: COAST participants include 13 907 BC residents living with HIV (≥19 years of age) and a 10% random sample comparison group of the HIV-negative general population (514 952 individuals).

Primary and secondary outcome measures: Healthcare records were used to identify transgender people via a CP algorithm (diagnosis codes + androgen blocker/hormone prescriptions), to examine related diagnoses and prescription concordance, and to validate the CP using an independent provider-report transgender status measure. Demographics and chronic illness burden were also characterized for the transgender sample.

Results: The best-performing CP identified 137 HIV-negative and 51 HIV-positive transgender people (total 188). In validity analyses, the best-performing CP had low sensitivity (27.5%, 95%CI:17.8-39.8), high specificity (99.8%, 95%CI:99.6-99.8), low agreement using Kappa statistics (0.3, 95%CI:0.2-0.5), and moderate positive predictive value (43.2%, 95%CI:28.7-58.9). There was high concordance between exogenous-sex hormone use and transgender-specific diagnoses.

Conclusions: The development of a validated CP opens up new opportunities for identifying transgender people for inclusion in population-based health research using administrative health data, and offers the potential for much-needed and heretofore unavailable evidence on health status, including HIV status, and the healthcare use and needs of transgender people.

KEYWORDS: Transgender Persons, Health Services, Algorithms, Canada

ARTICLE SUMMARY

Strengths and limitations of this study:

- This study demonstrates the feasibility of developing and validating a computable phenotype for identification of a transgender sample, using a population-based representative source population and healthcare records.
- A major contribution of this study is the ascertainment of the population of transgender people living with HIV in the Canadian province of British Columbia, in a universal

healthcare setting, using a computable phenotype, and capacity to estimate the prevalence of transgender status among the population living with HIV in the province.

- Development of a validated transgender computable phenotype algorithm lays the foundation for future investigation of transgender-specific research questions related to general and HIV-specific healthcare use and health outcomes for this key population.

INTRODUCTION

Limited data on transgender people

Transgender people are often overlooked within epidemiological research and population health surveillance due to small sample size, limited research designs, and other institutional and methodological erasures.[1–3] A 2017 review of Medline-indexed literature from 1950 to 2016 found 2405 published articles including transgender people, with almost half published in the last decade.[4] A 2008 United States (US)-based meta-analysis of HIV prevalence among transgender populations found 24 studies of transgender women, and five additional studies of transgender men,[5] though an updated review found 43 primary studies on transgender women and 15 on transgender men published between 2006-2017.[6] Despite this recent increase in transgender health research in general and for HIV specifically, much of the literature has focused on transgender-specific care, mental health and HIV/sexual health,[7,8] leaving the population understudied, in particular in the broader areas of physical health and healthcare utilization.

The erasures or exclusions of transgender persons in health studies may be explained, in part, by methodological challenges. Specific to electronic health record (EHR) data, a 2017 report identified only one transgender person among 38,5820 cancer cases in a Minnesota cancer registry,[9] clearly an undercount given that 0.4% of the US general population and 0.6% (95% credibility intervals: 0.5%-0.7%) of the Minnesota population is estimated to be transgender.[10,11] This highlights the need for improved gender ascertainment and transgender inclusion in research relying on patient records and administrative data. The establishment of best practices for measuring transgender status in survey research, such as the two-step method (measuring sex assigned at birth and current gender identity), points to a way forward for transgender-inclusive population health research.[12,13] However, innovative research methods are needed to identify transgender people in studies that rely on existing data sources (in particular EHR) and that optimize the use of transgender respondents' data in non-transgender specific research.

Computable phenotypes for transgender health research

Previous research in transgender health largely comprises cross-sectional studies, case reports, and qualitative or observational research.[7] Much consists of clinic- or venue-based convenience samples or lack comparison groups.[7,8] The literature is further characterized by inconsistent transgender status measurement,[14] small sample sizes, and focus on the United States (US).[8] In response, researchers have called for advancing transgender health research methods - namely ascertainment of high-quality samples via systematic approaches - including for general population-based and health systems-based studies.[15] One opportunity for the advancement of transgender health research methods is the emerging use of computable phenotypes (CPs)[16] or case ascertainment algorithms, to identify transgender samples in healthcare utilization data. A computable phenotype is an algorithm for identifying a clinical feature, condition, or set of characteristics that can be determined directly from EHR and other ancillary health care data systems (e.g., disease registries, insurance claims data) data.[17] CPs are developed using a combination of data elements (e.g., sociodemographic variables, clinical diagnoses) and value sets (i.e., the selection of a set of relevant values for each data element). Development of CPs using standardized methods and definitions enables identification and inclusion of transgender persons in research, as well as replication of analyses across data

sources, healthcare organizations/sites and studies. CPs have application in clinical care, surveillance, and health research.

Recently, CP and other EHR-based algorithm methods have been applied in a number of settings primarily in the US to identify transgender samples for health research.[14] Specifically, the STRONG study identified a transgender cohort (n = 6,456) using EHR data from Kaiser Permanente health plan members in California and Georgia, for investigation of general and transgender-specific health outcomes.[18] Blosnich et al identified 3,177 people with a "gender identity disorder" diagnosis among military veterans accessing care through the US Veterans Health Administration healthcare system,[19] for examination of mental health and other outcomes. Researchers with the US Centers for Medicare & Medicaid Services identified 4,098 transgender beneficiaries using national Medicare claims data,[20] and researchers at Vanderbilt University identified 234 transgender patients in their university clinic EHR data.[16] While these cohorts represent important opportunities for advancement of transgender health research, these methods have yet to be applied widely outside the US context. This is particularly important as different jurisdictions may vary in medical billing and coding practices, healthcare system patient populations, and representativeness of the general population. Specifically, in Canada, healthcare is delivered through a provincially administered universal healthcare system. As such, research using EHR provides an opportunity to develop methods for population-based, representative estimates of transgender populations within the Canadian context. Coupled with the current absence of gender ascertainment measures in population-based routinely collected data (e.g., census, national government health surveys, etc.) in Canada and many other jurisdictions, this remains an evidence need.

Summary of study rationale

This study investigated the application of emerging transgender health research methods, specifically CPs, in a Canadian context for the first time, testing the feasibility of identification of a transgender sample using EHR data from a provincial healthcare administrative data-linked cohort.

METHODS

Data Sources and Participants

The Comparative Outcomes and Service Utilization Trends Study (COAST)

COAST is a population-based cohort study focused on health services utilization research questions among all people known to be living with HIV (PLWH) in the province of British Columbia (BC) and a 10% random sample comparison group of the HIV-negative general population.[21] The COAST cohort comprises individual-level, longitudinal data from PLWH who have ever accessed HIV treatment in BC between 1996 and 2013, provided by Population Data BC (PopDataBC)[21] via data linkage between two provincial data sources, by personal health number: the Drug Treatment Program (DTP) [22] and the Ministry of Health. PopDataBC provides infrastructure for access to, and linkage of, longitudinal and individual-level administrative health data for all BC residents.[23]. The HIV-negative general population cohort was drawn randomly from the Ministry of Health registry data by PopDataBC. The COAST study has received approval from the University of British Columbia/Providence Health Care Research Ethics Board (#H09-02905) and Simon Fraser University Office of Research Ethics (#2013 s0566). The study complies with the BC Freedom of Information and Protection of

Privacy Act (FIPPA) and did not require informed consent as it is conducted using retrospective administrative and anonymized data for research and statistical purposes only. No patients or public were involved in this study.

Drug Treatment Program

In BC, antiretroviral therapy (ART) is provided to PLWH at no cost to the patient, and distributed through the DTP.[22] The DTP contributed a provider-reported measure of transgender status for COAST.

Ministry of Health

Ministry of Health data available via COAST included insured medical service billing records for outpatient visits,[24,25] hospital (in-patient) visits,[26] prescription medications,[27,28] and vital statistics.[29]

Measures & Analyses

Transgender computable phenotypes

Identification of transgender cases was tested in COAST using International Classification of Disease (ICD) codes (9th and 10th editions) and exogenous sex hormone prescription use. Transgender-specific diagnoses in medical and hospital billing records included the ICD-9 codes 302.5 Trans-sexualism with unspecified history, 302.51 Trans-sexualism with asexual history, 302.52 Trans-sexualism with homosexual history, 302.53 Trans-sexualism with heterosexual history, 302.6 Gender Identity Disorder in children, 302.85 Gender Identity Disorder in adolescents or adults; and ICD-10 codes F64.0 Gender Identity Disorder of childhood, F64.2 Gender Identity Disorder of childhood, F64.8 Other Gender Identity Disorder, and F64.9 Gender Identity Disorder unspecified. The full list of androgen blockers and exogenous sex hormone prescriptions included in analyses is available in the supplementary material.

Concordance

To assess face validity and utility of diagnosis and prescription data over time in CP development (i.e. whether the identified transgender sample had exogenous sex hormone prescription use and other diagnoses patterns consistent with that of transgender populations in other studies), concordance analyses evaluated the presence of at least one included diagnosis and prescription during the COAST study follow-up period with the presence of at least one included diagnosis and prescription in the last study year. Concordance was assessed between transgender-specific diagnoses, exogenous sex hormone and androgen blocker prescriptions, and non-transgender specific diagnoses (ICD-9 259.9 Unspecified Endocrine Disorder and ICD-10 E34.9 Endocrine Disorder, Unspecified [see supplementary material]). Endocrine disorder diagnosis codes are sometimes preferred by medical providers treating transgender people in response to historic exclusions of transgender-specific care from insurance coverage and to combat the stigma of transgender-specific diagnosis codes that have historically been classified as psychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM).[30] Exogenous sex hormone use, while common in transgender populations,[5,31] is not transgender-specific. Cisgender populations also use androgen blocker and sex hormone prescriptions (e.g. estrogen to treat menopausal symptoms in cisgender women, spironolactone is used for hypertension), thus exogenous sex hormone and androgen blocker prescription use

cannot independently identify transgender people. At the same time, not all transgender people use hormones and some access via non-medical sources.[32,33]

Validation

In British Columbia, transgender status is collected in the DTP via a provider-reported sex variable ("Male", "Male to Female", "Female to Male", or "Female"). Patients reported as either "Male to Female" or "Female to Male" were classified as transgender. The provider-reported transgender measure, available for the HIV positive cohort only, was used as a 'gold standard' for CP validation. Sensitivity, specificity, positive predictive value and kappa statistics with corresponding 95% confidence intervals (CI) were calculated for identifying transgender people via the CPs, in the HIV positive cohort only. Follow-up time (mean and range) for each CP group was also produced.

Demographics and chronic conditions

To further assess face validity of the transgender CP for future health research, descriptive statistics were calculated for the transgender sample produced via application of the best performing CP from the validation analysis to both the COAST HIV-positive and HIV-negative cohorts. Descriptive statistics included COAST study key sociodemographic and health variables, specifically laboratory confirmed HIV serostatus (HIV-positive/HIV-negative), baseline age, patient's Health Authority (five provincial regions for the administration of health services that include large urban centres, suburban regions, and rural/remote areas), and chronic illness burden based on standardized case definitions from the BC Ministry of Health [34] and the BC Cancer Agency.[35]

RESULTS

The total COAST cohort included 528 859 people, of which 514 952 were HIV-negative (10% general population random sample) and 13 907 were PLWH (Figure 1).

[Figure 1 here]

Concordance

Of the 237 people who had ever had a transgender-specific diagnosis during the study period, 19.4% also had a recent diagnosis in the last follow-up year (Table 1). None had an unspecified endocrine disorder diagnosis at any time; thus, this diagnosis was excluded from all CPs. Of the 237, 79.3% had an exogenous sex hormone or androgen blocker prescription at least once during the study period and 46.4% had one in the last year.

Table 1. Concordance analyses for diagnoses and hormone measures

	N	%
≥ Transgender ICD- ever	237	100
≥ Transgender ICD- recent	46	19.4
Unspecified endocrine disorder use- ever	0	0.0
Unspecified endocrine disorder use- recent	0	0.0
≥ Hormone/blocker use- ever	188	79.3
≥ Hormone/blocker use-recent	110	46.4

Validation

While no one CP consistently performed well across all validation metrics, the CP with the best overall performance across test statistics was based on having received at least one transgender-specific diagnosis and at least one androgen blocker/exogenous sex hormone prescription over the study follow-up period (Table 2). This CP had high specificity (99.8%, 95% CI: 99.6-99.8), low sensitivity (27.5%, 95% CI: 17.8-39.8), low to moderate Kappa coefficients (0.3, 95% CI: 0.2-0.5) and moderate positive predictive values (43.2%, 95% CI: 28.7-58.9). This CP also had the second longest mean follow-up time (mean: 136.3, range: 21.0-203.0), similar overall to the other CP groups (mean: 136.5, range: 21.0-203.0; mean: 117.1, range: 24.0-198.0; mean: 130.4, range: 69.0-198.0; respectively).

Table 2. Validation measures of transgender computable phenotype (CP) with provider-report transgender status measures, in COAST HIV-positive cohort

CP	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	Kappa (95% CI)
≥ 1 transgender ICD- ever	27.5 (17.8, 39.8)	99.7 (99.6, 99.8)	40.4 (26.7, 55.7)	0.3 (0.2, 0.4)
≥ 1 transgender ICD- recent	8.7 (3.6, 18.6)	100.0 (99.9, 100.0)	85.7 (42.0, 99.4)	0.2 (0.1, 0.3)
≥ 1 transgender ICD AND ≥ 1 hormone/blocker use- ever	27.5 (17.8, 39.8)	99.8 (99.6, 99.8)	43.2 (28.7, 58.9)	0.3 (0.2, 0.5)
≥ 1 transgender ICD AND ≥ 1 hormone/blocker use- recent	7.3 (2.7, 16.8)	100.0 (99.9, 100.0)	83.3 (36.5, 99.1)	0.1 (0.0, 0.2)

Transgender phenotype

Applying the best-performing CP, 137 HIV-negative people and 51 HIV-positive people (188 total) were identified as transgender in the respective COAST cohorts (Figure 1).

Demographics and chronic conditions

Demographic characteristics and chronic conditions for the 188 transgender people identified via the best-performing CP are presented in Figures 2 to 4. Transgender people were geographical located throughout BC health regions. The Vancouver Coastal Health Authority region, which includes the largest municipal area in BC, had the highest concentration of transgender people (44.2%) while the Northern Health Authority region - a predominantly rural and remote area of the province - had the lowest (1.6%).[36] The HIV-positive group had a higher median age than the HIV-negative group (35 [Q1, Q3: 30,42] and 30 [Q1, Q3: 19,42], respectively). For the HIV-negative sample, the largest proportion of transgender people were aged 19 to 29 years (44.5%) and the smallest proportion aged 55 years and older (<3%). For the HIV-positive sample, the largest proportion were aged 30 to 34 years (25.5%) and the smallest proportion aged 55 years and older (<2%).

[Figures 2 and 3 here]

Overall, HIV-positive transgender people had a higher prevalence of at least one chronic condition (other than HIV) compared to HIV-negative transgender people (88.2% versus 85.4%, respectively), and of two or more chronic conditions (76.5% versus 52.6%, respectively). Specific chronic disease differences between transgender people living with and without HIV were most notable for a higher prevalence among the HIV-positive cohort of cardiovascular disease, chronic kidney disease, osteoarthritis, schizophrenia and personality disorders, and chronic liver disease, but a lower prevalence for hypertension.

[Figure 4 here]

DISCUSSION

This study demonstrates the feasibility of identification of a sample of transgender people in a large linked provincial healthcare administrative database, using a CP based on prescriptions and diagnoses. Among a growing number of studies using EHR and CP methods to identify transgender samples for health research purposes, this is the first to do so in Canada., to independently validate the CP using a 'gold standard' of provider-reported transgender status, and the only to use population-based data.

Concordance

There was high concordance between transgender-specific diagnoses and exogenous sex hormone or androgen blocker prescription use in this study. That nearly half of those with at least one transgender-specific diagnosis had been dispensed hormones or blockers in the past year is consistent with findings from US and Canadian studies (48.9% and 43.0%, respectively)[20,32,33] - suggesting face validity for the current CP.

CP development and validation

The best-performing CP overall successfully identified cisgender people who were truly cisgender (specificity) and correctly identified transgender people who were truly transgender (0.2% false positive rate, results not shown). However, the selected CP had relatively low sensitivity, missing approximately 72.5% of 'true' transgender people in COAST, as identified by the gold standard provider-based measure. Though a relatively small proportion of the 'true' transgender sample was identified in this study, the impact on future analyses comparing health outcomes for transgender and cisgender groups is likely negligible, as even the large proportion of 'true' transgender people misclassified as cisgender (approximate n=496) is a very small proportion of the total COAST sample. At worst, this misclassification would bias results related to disparities between transgender and cisgender health toward the null, producing a conservative attenuated effect in COAST, and other such administrative datasets. Further, as discussed below, gender identity classification will likely greatly improve as transgender care shifts further into the fee-for-service system in BC. As in other Canadian administrative data studies, low sensitivity may be explained in part by provider and system billing preferences using 3-digit ICD diagnosis coding instead of the more specific 4-digit coding, and inconsistencies in the BC billing management system.[37] Despite the low sensitivity, CP development in this study with high specificity offers an advancement for transgender health research. A measure that correctly identifies cases for transgender samples in research with good success translates to better

opportunities to include transgender people in health studies and to investigate their health relative to other groups. While future research may lead to improvements in CP development, the CP identified in the current study with good specificity, albeit relatively poor sensitivity, has important utility in advancing opportunities in transgender health research. Additionally, while differential follow-up time can affect algorithm performance, the similar mean and range follow-up time for all CPs in this study suggests that differential follow-up time was not an important source of bias in this study.

The limited agreement between the CP and provider-report transgender status may be due to the widely varying transgender status prevalence depending on study design and ascertainment measures used.[14] In the BC context, the CP and the DTP measures are assessing transgender status in different ways and for different purposes. In the DTP, transgender status is ascertained in the context of HIV diagnosis and ART prescribing, during which demographics and HIV transmission risk factors are recorded. This differs from recording diagnoses in EHR for those accessing transgender-specific care as utilized in the CP. This may explain the lower PPV for the best-performing CP compared to the CP based on recent transgender diagnoses, suggesting the DTP provider-reported transgender status measure has better coverage for recent cases and the potential for use of recent diagnosis over ever to be beneficial in future CP development. Ultimately, a single CP may not be sufficient for all intended purposes and the best applicable CP (using different types of diagnoses, prescriptions or procedures) may differ depending on the intended healthcare, health research, or health policy application.[17]

There is limited literature on EHR-based studies with the ability to validate an administrative transgender measure using a 'gold standard' comparison measure.[16] The two previous studies that have developed and validated algorithms to identify transgender individuals have both been conducted in non-representative samples in the US, one using Medicare data[38] and one in a university medical center.[16] Similar to the current study, the Medicare study found high specificity when comparing an EHR-based and a two-step survey-based transgender measure. However, the Medicare study found that the EHR measure performed consistently well with high sensitivity and a high Kappa statistic, unlike in the current study. Using chart review as the 'gold standard' for comparison of transgender status, Ehrenfeld et al. found a low false positive rate for their best-performing algorithm (3%), though not as low as the false positive rate in the current study. The overall high levels of agreement for transgender measures in the two previous studies is likely a function of the lack of independence between the 'gold standard' and the CP or algorithm measures. Specifically, only those classified as transgender in the Medicare EHR data were offered survey participation to complete the two-step 'gold standard' survey measure, and only those cases identified as transgender in the university clinic EHR were included in chart review. Thus, previous studies could assess agreement between the two measures, but not robustly validate either. In the current study, the DTP provider-based transgender status measure is independent and thus could be used for robust CP validation.

While not possible to incorporate free-text records in case-finding algorithms in the current study as only structured EHR data is linked through COAST, it is worth noting the opportunities potentiated by use of NLP and machine learning approaches as methods for identifying transgender samples in EHR data as this research area continues to grow. Outside of transgender health, the use of NLP and machine learning to mine unstructured free-text EHR data has

demonstrated efficiency in improving case ascertainment algorithm accuracy .[39] As 'gold standard' two-step sex assigned at birth and current gender identity measures of transgender status[12] are slowly being implemented in routinely collected healthcare data sources, in the meantime NLPs to extract free-text data can be used to produce better gold standards against which to measure algorithm performance, as demonstrated by the Medicare study.[38]

Transgender status prevalence & ascertainment

Based on a recent meta-analysis of transgender status prevalence in population-based probability samples,[10] it was expected that an effective CP would identify 0.4% of the general population as transgender, or approximately 54 of the HIV-positive COAST cohort (n=13 907) and 3,098 of the HIV-negative cohort (n=516 340). Consistent with expectations, the best-performing CP identified 51 PLWH as transgender, equivalent to a transgender status prevalence of 0.4% among PLWH. Contrary to expectations, the best-performing CP identified less than 5% of the number of transgender persons expected in the HIV-negative cohort. This is likely a result of a number of factors including the limitation of CPs to the subset of a population accessing care as noted, and the result of most transgender people in BC receiving care currently outside the main fee-for-service healthcare delivery system. However, it is also consistent with the undercount of transgender populations using diagnostic criteria compared to other methods of ascertainment demonstrated in other studies.[14]

Using the broadest CP algorithm (any transgender-specific diagnosis ever, n=56) and those identified by provider-report together (total n=106), the total transgender PLWH sample would represent as high as 0.88% (range: 0.73-1.1%) of the prevalent HIV infections in BC in 2014.[40] This overrepresentation of transgender people among PLWH is consistent with evidence of a disproportionate HIV burden for transgender populations globally,[5,41,42] as well as in line with the only other available data on the proportion of PLWH who are transgender, from US national surveillance data (2012 data: 1.1%, 95%CI: 0.8-1.4).[43]

Demographics and chronic conditions

Despite moderate to low performance by some validation metrics, particularly low sensitivity, the CP was able to detect meaningful results in the characterization of demographics and chronic condition burden for the transgender sample - supporting CP face validity. The population density and age distribution by HIV-status of transgender people in this study is largely consistent with general population patterns, as well as the larger COAST cohort.[21,36] The overall higher burden of chronic illness for transgender people living with HIV versus without HIV in this study is consistent with elevated chronic illness risk and morbidity among non-transgender PLWH.[44] This higher chronic disease burden is linked to HIV disease processes and related inflammatory immune response.[45] While a small but growing number of studies have begun to investigate the chronic illness burden for transgender populations in other industrialized settings,[16,19,46–48] including using EHR data, findings vary widely due to differences in sampling, study design, setting and measurement.

Limitations

Findings from this study should be interpreted in the context of a few key limitations. CPs are by design only applicable to people accessing healthcare services, often motivated by illness and aided by the ability to access care. As such, this study is limited to those transgender people

accessing medical transition care in BC and may only represent 24% to 47% of the total transgender population.[33] This study was also limited by the inability to validate the transgender CP among the HIV-negative COAST cohort, as a 'gold standard' provider-based transgender measure was only available for the HIV-positive cohort. It is possible that the transgender CPs would perform differently in populations living without HIV, particularly as healthcare contact is higher among populations living with HIV. Additionally, this study should be considered in light of the context in which it was conducted, an environment in which transgender healthcare delivery in BC is currently shifting from specialized care settings to the main primary care fee-for-service settings. Given that COAST only includes fee-for-service data, this study was limited by the inability to capture transgender people who access transgender care outside the fee-for-service system. However, fortunately, as the shift to the fee-for-service system occurs, transgender ascertainment via CPs in BC will likely improve. The administrative data used in this study may also be susceptible to coding error (and coding biases/practices) across conditions and settings,[49] potentially introducing misclassification bias in terms of transgender ascertainment. Finally, chronic condition prevalence data reported in this study should be interpreted with caution, given potential selection bias by serostatus in the COAST cohort; though any such bias likely resulted in conservative estimates of difference by serostatus in this analysis.

CONCLUSION

This study makes a number of important contributions to the literature on innovative methods in transgender health. Major contributions include development and validation of a transgender CP, using a population-based representative source population, in the Canadian context. Another strength is the approximately complete ascertainment of the population of transgender PLWH in BC, and capacity to estimate transgender status prevalence among PLWH. In a current funding environment of limited support for longitudinal transgender health studies in the US and none to date in Canada, this study and the methods employed offer an efficient, replicable and cost-effective way forward in creating electronic cohorts for advancing transgender health research.[15] Moreover, the recent rollback of sexual orientation and gender identity data collection and legal changes in insurance coverage of transgender healthcare in the US potentiate decline in accurate claims coding for gender-affirming care.[30] This highlights the utility of work in this area from other jurisdictions, particularly those with transgender-inclusive universal healthcare systems such as Canada.

Future research should build upon the methods developed in this study and explore complimentary approaches for gender identity ascertainment in administrative and EHR data, such as machine learning approaches, as have been used to develop algorithms based on healthcare utilization data in other research areas. Finally, the current study lays the foundation for future work with the ability to study transgender health and healthcare use patterns over time, with linkage to laboratory data, as well as inclusion of appropriate comparison groups.[15,50]

ACKNOWLEDGEMENTS & DISCLAIMER

The authors thank the COAST study participants, the BC Centre for Excellence in HIV/AIDS, the BC Ministry of Health, BC Vital Statistics Agency, PharmaNet and the institutional data stewards for granting access to the data, and Population Data BC for facilitating the data linkage process. In addition, we would like to thank the COAST core team members and other support

staff at these institutions for their administrative assistance with the data access and preparation. All inferences, opinions, and conclusions drawn in this paper are those of the authors, and do not reflect the opinions or policies of the Data Steward(s). There are no conflicts of interest to declare.

COMPETING INTERESTS

None declared.

FUNDING

This work was supported by the Canadian Institutes of Health Research, through an Operation Grant [grant number 130419], a Foundation Award to RSH [grant number 143342], and a Doctoral Research Award to AJR [grant number 152382] and support from the British Columbia Centre for Excellence in HIV/AIDS. The DTP receives funding from the provincial government of British Columbia (PharmaCare). The funders had no role in the study design, analysis, interpretation of the data, drafting of the manuscript or in the decision to submit for publication.

CONTRIBUTORS

AJR led the study from conceptualization to analysis plan to interpretation, drafting of the first manuscript version, revisions and final version. RSH acquired study data and funding. TP, MK, PS, TS, and RSH all contributed to study design, interpretation of results, and reviewed manuscript versions. JL and MY contributed to study analysis and reviewed manuscript versions. All authors provided critical review of first and subsequent manuscript drafts, approved the final version, and agree to be accountable for the work presented.

PATIENT AND PUBLIC INVOLVEMENT

No patients involved.

DATA SHARING STATEMENT

The data used for this study are held by the BC Centre for Excellence in HIV/AIDS under the authority of the BC Ministry of Health; as they contain confidential patient health records including HIV serostatus, data cannot be made available to other parties.

FIGURES LEGENDS

Figure 1. Total transgender sample identified using a computable phenotype with electronic health records

Figure 2. Geographic distribution of transgender people across province, by health authority*

**% of transgender individuals with known health authority (n=182)*

Figure 3. Age distribution of transgender sample, by HIV serostatus

Figure 4. Co-morbidities among transgender sample, by HIV serostatus

REFERENCES

1. Bauer GR. Making sure everyone counts: Considerations for inclusion, identification, and analysis of transgender and transsexual participants in health surveys. What Differ Sex Gend Make Gend Sex Health Res Casebook. Vancouver, Canada: Canadian Institutes of Health Research (CIHR), Institute for Gender and Health (IGH); 2012. p. 59–67.

2. Bauer GR, Hammond R, Travers R, Kaay M, Hohenadel KM, Boyce M. "I don't think this is theoretical; this is our lives": How erasure impacts health care for transgender people. JANAC. 2009;20:348–361.

3. Poteat T, German D, Flynn C. The conflation of gender and sex: Gaps and opportunities in HIV data among transgender women and MSM. Glob Public Health. 2016;11:835–848.

4. Wanta JW, Unger CA. Review of the transgender literature: Where do we go from here? Transgender Health. 2017;2:119–128.

5. Herbst JH, Jacobs ED, Finlayson TJ, McKleroy VS, Neumann MS, Crepaz N, et al. Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a systematic review. AIDS Behav. 2008;12:1–17.

6. Becasen JS, Denard CL, Mullins MM, Higa DH, Sipe TA. Estimating the prevalence of HIV and sexual behaviors among the US transgender population: A systematic review and meta-analysis, 2006–2017. Am J Public Health. 2019;109:e1–8.

7. Marshall Z, Welch V, Thomas J, Brunger F, Swab M, Shemilt I, et al. Documenting research with transgender and gender diverse people: protocol for an evidence map and thematic analysis. Syst Rev. 2017;6:35.

8. Reisner SL, Poteat T, Keatley J, Cabral M, Mothopeng T, Dunham E, et al. Global health burden and needs of transgender populations: a review. The Lancet. 2016;388:412–436.

9. Cathcart-Rake EJ, Lightner DJ, Quevedo FJ, Ertz DC, Jatoi A. Cancer in transgender patients: One case in 385,820 is indicative of a paucity of data. J Oncol Pract. 2017;JOP2017027714.

10. Meerwijk EL, Sevelius JM. Transgender population size in the United States: A meta-regression of population-based probability samples. Am J Public Health. 2017;107:e1–e8.

11. Flores AR, Herman JL, Gates GJ, Brown TNT. How Many Adults Identify as Transgender in the United States? [Internet]. Los Angeles, CA: Williams Institute; 2016 Jun p. 13. Available from: https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf

12. The GenIUSS Group. Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys. Los Angeles, CA: The Williams Institute; 2014.

13. Bauer GR, Braimoh J, Scheim AI, Dharma C. Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. PLOS ONE. 2017;12:e0178043.

14. Collin L, Reisner SL, Tangpricha V, Goodman M. Prevalence of transgender depends on the "case" definition: A systematic review. J Sex Med. 2016;13:613–626.

15. Reisner SL, Deutsch MB, Bhasin S, Bockting W, Brown GR, Feldman J, et al. Advancing methods for US transgender health research. Curr Opin Endocrinol Diabetes Obes. 2016;23:198–207.

16. Ehrenfeld JM, Gottlieb KG, Beach LB, Monahan SE, Fabbri D. Development of a Natural Language Processing Algorithm to Identify and Evaluate Transgender Patients in Electronic Health Record System. Ethn Dis. 2019;29:441–50.

17. Richesson R, Smerek M, Rusincovitch S, Zozus MN, Chaudhuri PS, Hammond WE, et al. Electronic Health Records-Based Phenotyping. In: Uhlenbrauck G, editor. NIH Collab Living Textb Pragmatic Clin Trials. Bethesda, Maryland; 2014.

18. Quinn VP, Nash R, Hunkeler E, Contreras R, Cromwell L, Becerra-Culqui TA, et al. Cohort profile: Study of Transition, Outcomes and Gender (STRONG) to assess health status of transgender people. BMJ Open. 2017;7:e018121.

19. Blosnich JR, Brown GR, Shipherd Phd JC, Kauth M, Piegari RI, Bossarte RM. Prevalence of gender identity disorder and suicide risk among transgender veterans utilizing veterans health administration care. Am J Public Health. 2013;103:e27–32.

20. Proctor K, Haffer SC, Ewald E, Hodge C, James CV. Identifying the transgender population in the Medicare program. Transgender Health. 2016;1:250–265.

21. Eyawo O, Hull MW, Salters K, Samji H, Cescon A, Sereda P, et al. Cohort profile: the Comparative Outcomes And Service Utilization Trends (COAST) Study among people living with and without HIV in British Columbia, Canada. BMJ Open. 2018;8:e019115.

22. Drug Treatment Program [Internet]. BC Cent. Excell. HIVAIDS. 2010 [cited 2019 Sep 27]. Available from: http://www.cfenet.ubc.ca/drug-treatment-program

23. Population Data BC. About PopData [Internet]. Popul. Data BC. 2017 [cited 2019 Sep 27]. Available from: https://www.popdata.bc.ca/aboutus

24. British Columbia Ministry of Health (2014). Medical Services Plan (MSP) Payment Information File [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

25. British Columbia Ministry of Health (2014). Consolidation File (MSP Registration & Premium Billing) [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

26. Canadian Institute for Health Information (2014). Discharge Abstract Database (Hospital Separations) [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

27. British Columbia Ministry of Health (2014). PharmaCare [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

28. British Columbia Ministry of Health (2014). PharmaNet [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

29. British Columbia Vital Statistics Agency (2014). Vital Statistics Deaths [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca

30. Arnold JD, Nelson AE, Loubier EM. Trends in Insurance Coverage for Gender-Affirming Surgeries. JAMA Surg. 2018;153:972.

31. Radix A, Sevelius J, Deutsch MB. Transgender women, hormonal therapy and HIV treatment: a comprehensive review of the literature and recommendations for best practices. J Int AIDS Soc. 2016;19:20810.

32. Rotondi NK, Bauer GR, Scanlon K, Kaay M, Travers R, Travers A. Nonprescribed hormone use and self-performed surgeries: "do-it-yourself" transitions in transgender communities in Ontario, Canada. Am J Public Health. 2013;103:1830–1836.

33. Scheim AI, Bauer GR. Sex and gender diversity among transgender persons in Ontario, Canada: results from a respondent-driven sampling survey. J Sex Res. 2015;52:1–14.

34. Chronic Disease Information Working Group. BC Chronic Disease and Selected Procedure Case Definitions [Internet]. British Columbia, Canada: BC Ministry of Health; 2019 Apr p. 1–132. Available from: https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf

35. British Columbia Cancer Agency. BC Cancer Registry Data [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

36. BC Stats. Sub-provincial Population Estimates [Internet]. 2018 [cited 2019 Feb 1]. Available from: file://localhost/Users/AJRich/Zotero/storage/5VM2LBUN/PopulationEstimates.html

37. Hu W. Diagnostic Codes in MSP Claim Data. Program Monitoring and Information Management Branch, Resource Management Division, Medical Services Plan; 1996 p. 39.

38. Gerth J, Becerra-Culqui T, Bradlyn A, Getahun D, Hunkeler EM, Lash TL, et al. Agreement between medical records and self-reports: Implications for transgender health research. Rev Endocr Metab Disord. 2018;19:263–269.

39. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al. Development of phenotype algorithms using electronic medical records and incorporating natural

language processing. The BMJ [Internet]. 2015 [cited 2020 Oct 22];350. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4707569/

40. BC Centre for Disease Control. HIV in British Columbia: Annual Surveillance Report 2015. Vancouver, Canada: BC Centre for Disease Control; 2017.

41. Poteat T, Scheim A, Xavier J, Reisner S, Baral S. Global epidemiology of HIV infection and related syndemics affecting transgender people. J Acquir Immune Defic Syndr. 2016;72 Suppl 3:S210–9.

42. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV in transgender women: a systematic review and meta-analysis. Lancet Infect Dis. 2013;13:214–222.

43. Centers for Disease Control and Prevention. Behavioral and Clinical Characteristics of Persons Receiving Medical Care for HIV Infection—Medical Monitoring Project, United States, 2012 [Internet]. Washington, D.C.: Centers for Disease Control; 2015 Aug p. 42. Report No.: 12. Available from: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf

44. Grabovac I, Veronese N, Stefanac S, Haider S, Jackson SE, Koyanagi A, et al. Human Immunodeficiency Virus Infection and Diverse Physical Health Outcomes: An Umbrella Review of Meta-analyses of Observational Studies. Clin Infect Dis. 2019;ciz539.

45. Nasi M, De Biasi S, Gibellini L, Bianchini E, Pecorini S, Bacca V, et al. Ageing and inflammation in patients with HIV infection: Ageing, inflammation and HIV infection. Clin Exp Immunol. 2017;187:44–52.

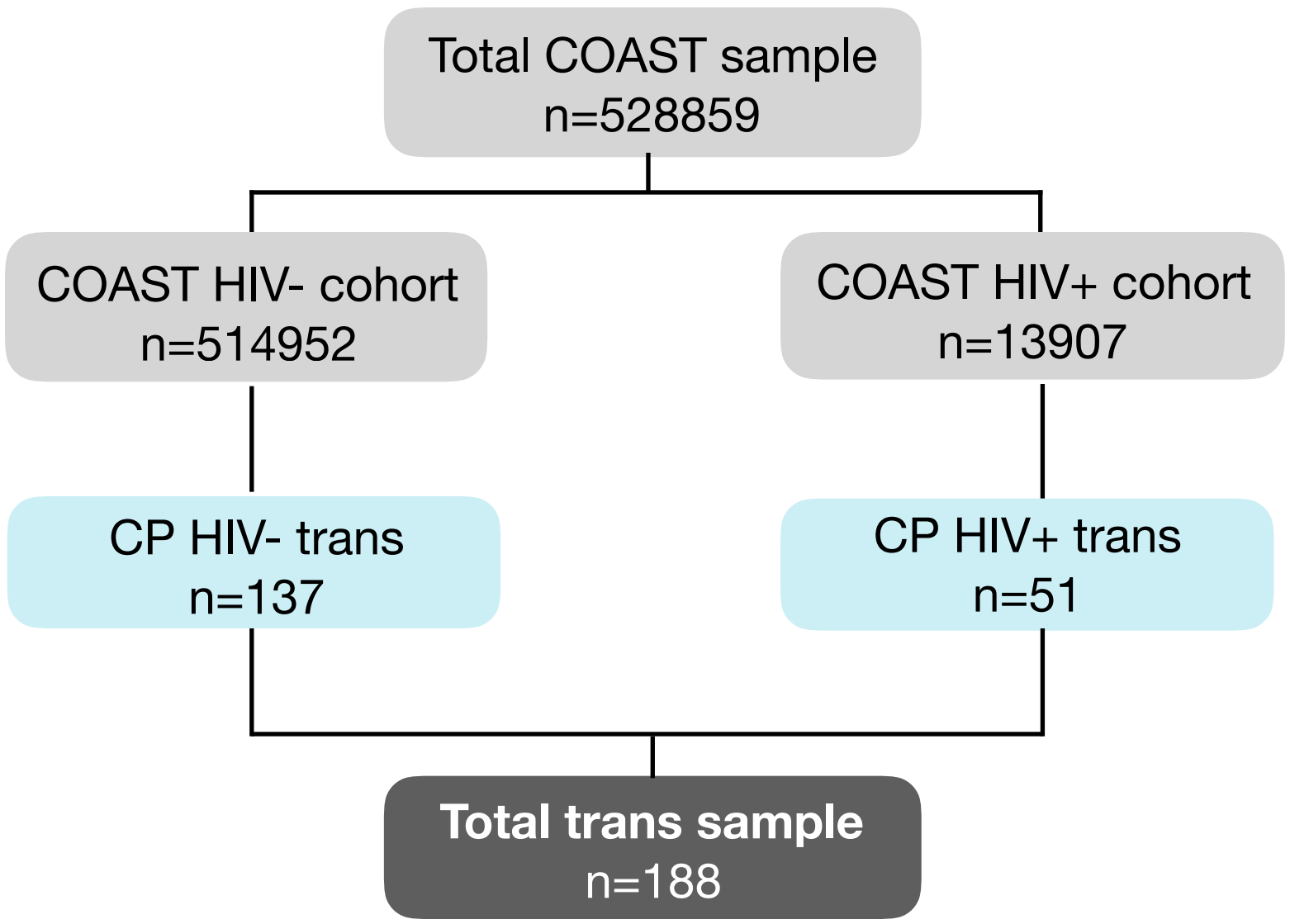
46. Judge C., O'Donovan C., Callaghan G., Gaoatswe G., O'Shea D. Gender dysphoria - prevalence and co-morbidities in an Irish adult population. Front Endocrinol. 2014;5:87.

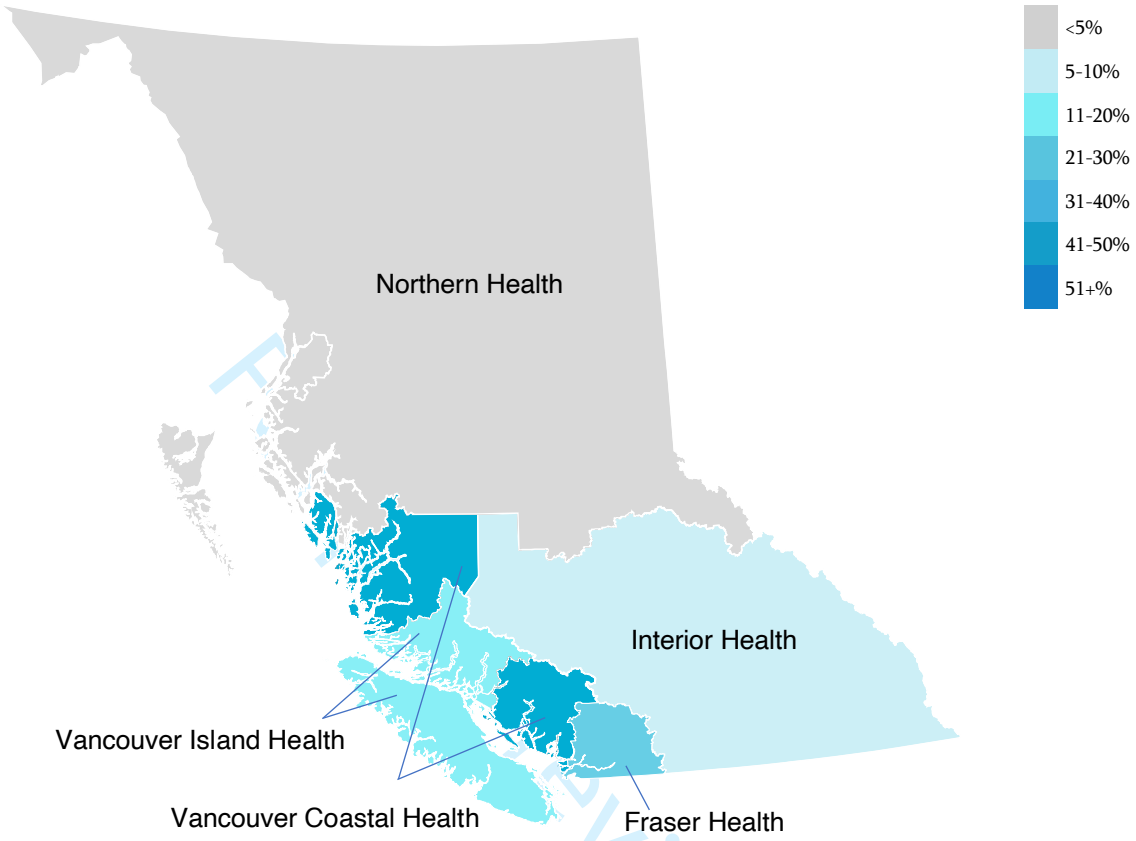
47. Cheung AS, Ooi O, Leemaqz S, Cundill P, Silberstein N, Bretherton I, et al. Sociodemographic and clinical characteristics of transgender adults in Australia. Transgender Health. 2018;3:229–38.

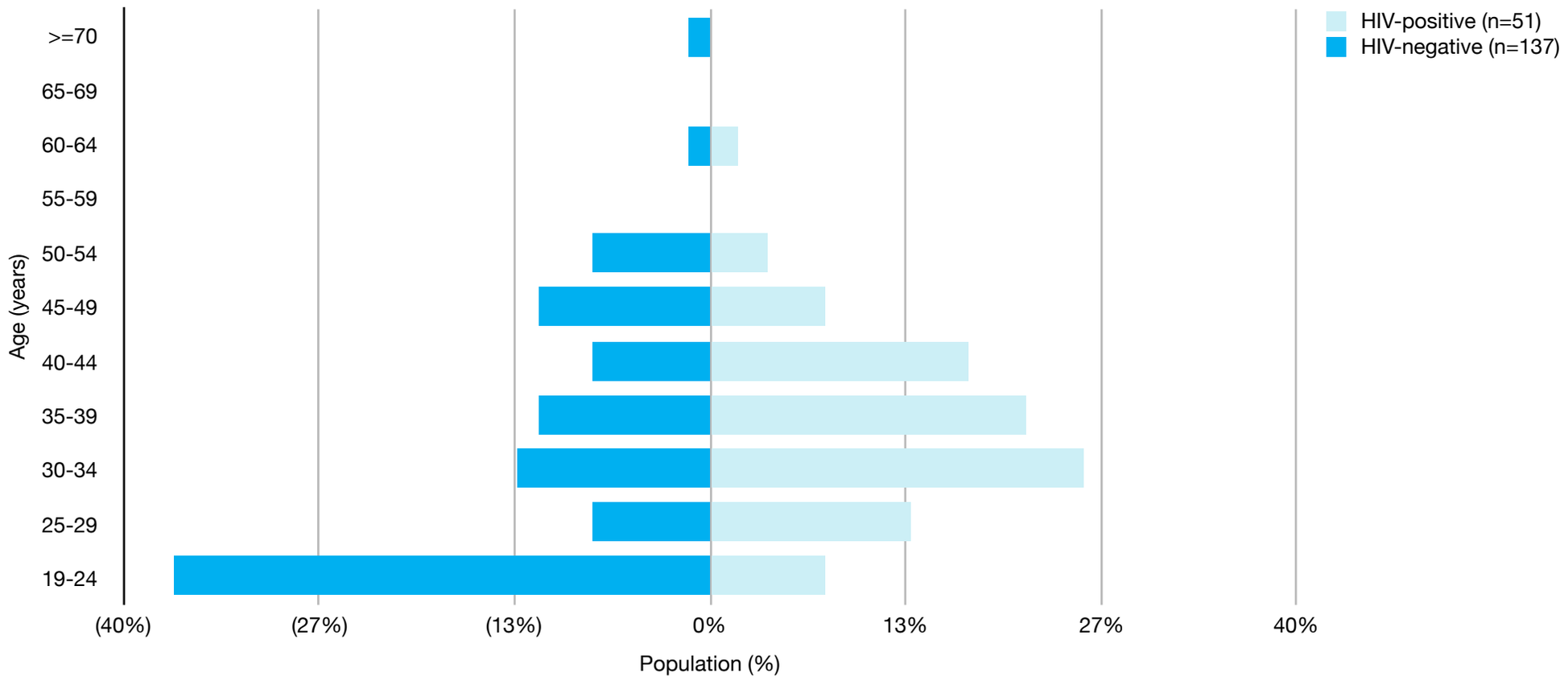
48. Swartz JA, Ducheny K, Holloway T, Stokes L, Willis S, Kuhns LM. A latent class analysis of chronic health conditions among HIV-Positive transgender women of color. AIDS Behav [Internet]. 2019 [cited 2019 Oct 2]; Available from: http://link.springer.com/10.1007/s10461-019-02543-3

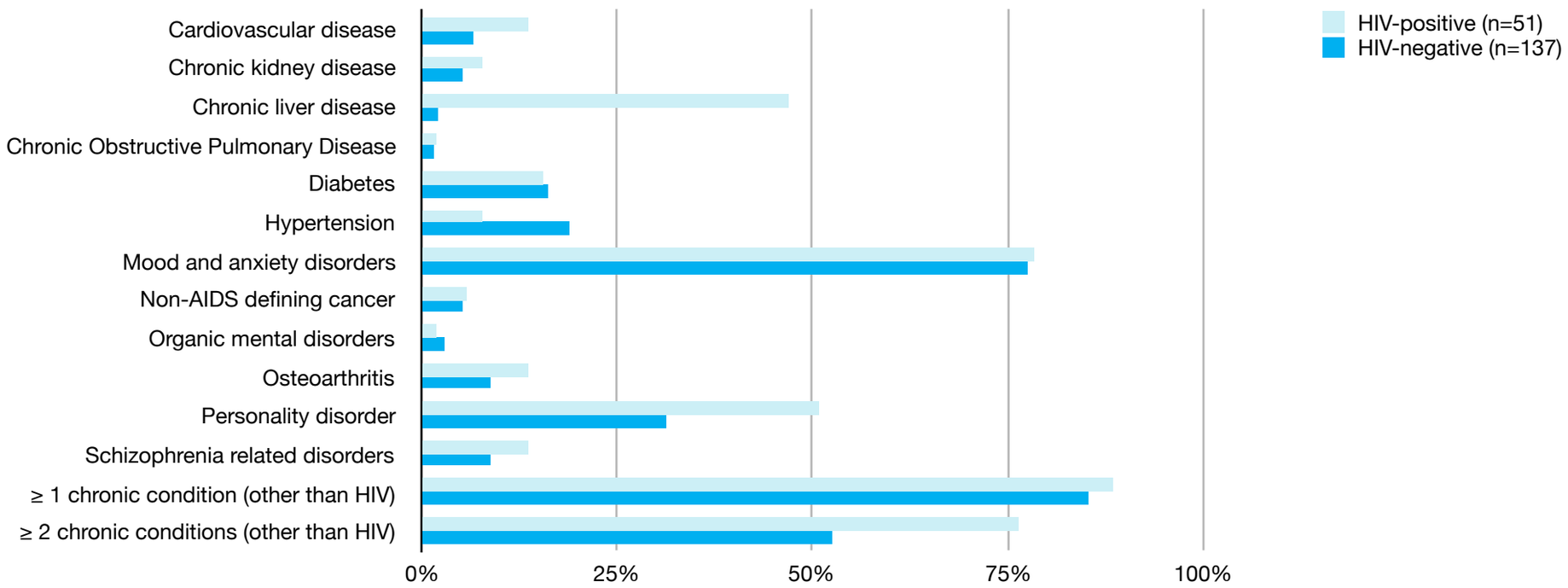
49. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 2011;

50. Reisner SL, Radix A, Deutsch MB. Integrated and gender-affirming transgender clinical care and research. J Acquir Immune Defic Syndr. 2016;72 Suppl 3:S235–42.








Cardiovascular disease
Chronic kidney disease
Chronic liver disease
Chronic Obstructive Pulmonary Disease
Diabetes
Hypertension
Mood and anxiety disorders
Non-AIDS defining cancer
Organic mental disorders
Osteoarthritis
Personality disorder
Schizophrenia related disorders
≥ 1 chronic condition (other than HIV)
≥ 2 chronic conditions (other than HIV)
0%
25%
50%
75%
100%
HIV-positive (n=51)
HIV-negative (n=137)

Supplementary material

Data sources and description of data elements

Data source	Data steward	Description	Data elements provided for this study
Medical Services Plan	British Columbia Ministry of Health	For individuals covered by the provincial universal health insurance plan- Medically necessary services provided by fee-for-service physicians and other healthcare providers, laboratory services, diagnostic procedures, dental/oral surgery	ICD-9 and ICD-10 codes
Consolidation file	British Columbia Ministry of Health	Demographic data on individuals receiving or registered to receive care in BC, pooled from multiple PopData sources	Sociodemographics
Discharge Abstract Database	British Columbia Ministry of Health	Demographic, administrative and clinical data for inpatient hospital discharges and day surgeries	ICD-9 and ICD-10 codes
Vital statistics deaths	British Columbia Vital Statistics Agency	Records of all registered deaths in BC	Death data
Pharmacare	British Columbia Ministry of Health	Data related to prescription drugs dispensed under the BC public drug insurance program.	Prescription drug name/identifier
Pharmanet	British Columbia Ministry of Health Data Stewardship Committee	Data related to prescription drugs dispensed by community and outpatient pharmacies	Prescription drug name/identifier

Supplementary material

Drug Treatment Program and laboratory	British Columbia Centre for Excellence in HIV/AIDS	Antiretroviral therapy use history, laboratory testing, immunological and virologic testing, and demographic data on PLWH who have accessed antiretrovirals in BC	Providers-reported transgender status, laboratory confirmed HIV serostatus

Prescription drugs with drug identification numbers (DIN)s

	Generic Name	DIN
Transfeminine		
Androgen Blockers		
Spironolactone		
	SPIRONOLACTONE	28606
	SPIRONOLACTONE	613215
	SPIRONOLACTONE	285455
	SPIRONOLACTONE	613223
	SPIRONOLACT/HYDROCHLOROTHIAZID	180408
	SPIRONOLACT/HYDROCHLOROTHIAZID	613231
	SPIRONOLACT/HYDROCHLOROTHIAZID	594377
	SPIRONOLACT/HYDROCHLOROTHIAZID	657182
Cyproterone		
	ETHINYL ESTRADIOL/CYPROTERONE	2233542
	NO GENERIC FORMULARY	634514
	CYPROTERONE ACETATE	704431
	CYPROTERONE ACETATE	2229449
	CYPROTERONE ACETATE	2229723
	CYPROTERONE ACETATE	2232872
	CYPROTERONE ACETATE	2245898
	CYPROTERONE ACETATE	704423
Finasteride		
	FINASTERIDE	2010909
	FINASTERIDE	2238213
Dutasteride		
	DUTASTERIDE	2247813
Estrogens		
Estrogen		
	ESTROGENS,CONJUGATED	830240
	ESTROGENS,CONJUGATED	831395

Supplementary material

ETHYNODIOL D-ETHINYL ESTRADIOL	28630
ETHYNODIOL D-ETHINYL ESTRADIOL	469327
NORETHINDRONE-MESTRANOL	22608
NORETHINDRONE-MESTRANOL	22659
NORETHINDRONE A-E ESTRADIOL	297143
NORETHINDRONE A-E ESTRADIOL	315966
NORETHINDRONE-ETHINYL ESTRAD	317047
NORETHINDRONE-ETHINYL ESTRAD	372846
NORETHINDRONE-ETHINYL ESTRAD	373265
NORETHINDRONE-ETHINYL ESTRAD	531006
NORETHINDRONE-ETHINYL ESTRAD	538590
NORETHINDRONE-ETHINYL ESTRAD	602957
NORETHINDRONE-ETHINYL ESTRAD	620947
NORETHINDRONE-ETHINYL ESTRAD	2187086
NORETHINDRONE-ETHINYL ESTRAD	2187108
NORETHINDRONE-ETHINYL ESTRAD	2189054
NORGESTREL-ETHINYL ESTRADIOL	34207
NORGESTREL-ETHINYL ESTRADIOL	300640
LEVONORGESTREL-ETH ESTRA	579386
LEVONORGESTREL-ETH ESTRA	707600
LEVONORGESTREL-ETH ESTRA	782416
LEVONORGESTREL-ETH ESTRA	782432
LEVONORGESTREL-ETH ESTRA	2042320
NORGESTREL-ETHINYL ESTRADIOL	2043033
LEVONORGESTREL-ETH ESTRA	2043726
NORGESTIMATE-ETHINYL ESTRADIOL	2258560
NORETHINDRONE-MESTRANOL	30333
NORETHINDRONE-MESTRANOL	30341
LEVONORGESTREL-ETH ESTRA	2236974
ETHYNODIOL D-ETHINYL ESTRADIOL	471526
NORETHINDRONE-ETHINYL ESTRAD	340731
NORETHINDRONE-MESTRANOL	340758
NORETHINDRONE A-E ESTRADIOL	343838
NORETHINDRONE A-E ESTRADIOL	353027
NORETHINDRONE-ETHINYL ESTRAD	372838
NORETHINDRONE-ETHINYL ESTRAD	373273
NORETHINDRONE-ETHINYL ESTRAD	531014
NORETHINDRONE-ETHINYL ESTRAD	602965
NORETHINDRONE-ETHINYL ESTRAD	695734
NORETHINDRONE-ETHINYL ESTRAD	2187094

Supplementary material

NORETHINDRONE-ETHINYL ESTRAD	2187116
NORETHINDRONE-ETHINYL ESTRAD	2189062
ETHINYL ESTRADIOL/NORETH AC	2242531
NORGESTREL-ETHINYL ESTRADIOL	340766
NORGESTREL-ETHINYL ESTRADIOL	342815
LEVONORGESTREL-ETH ESTRA	586609
LEVONORGESTREL-ETH ESTRA	707503
LEVONORGESTREL-ETH ESTRA	782424
LEVONORGESTREL-ETH ESTRA	782440
LEVONORGESTREL-ETH ESTRA	2042339
NORGESTREL-ETHINYL ESTRADIOL	2043041
LEVONORGESTREL-ETH ESTRA	2043734
NORGESTIMATE-ETHINYL ESTRADIOL	2258587
LEVONORGESTREL-ETH ESTRA	2236975
NORGESTIMATE-ETHINYL ESTRADIOL	1968440
NORGESTIMATE-ETHINYL ESTRADIOL	2028700
NORGESTIMATE-ETHINYL ESTRADIOL	1992872
NORGESTIMATE-ETHINYL ESTRADIOL	2029421
DESOGESTREL-ETHINYL ESTRADIOL	2042487
DESOGESTREL-ETHINYL ESTRADIOL	2042541
DESOGESTREL-ETHINYL ESTRADIOL	2042479
DESOGESTREL-ETHINYL ESTRADIOL	2042533
ESTRADIOL/NORETH AC	2241835
ESTRADIOL/NORETH AC	2241837
LEVONORGESTREL	2241674
ESTROGEN,CON/M-PROGEST ACET	2242878
ESTROGEN,CON/M-PROGEST ACET	2242879
ESTRADIOL/NORETH AC	2243529
ESTRADIOL/NORETH AC	2243530
ETHINYL ESTRADIOL/DROSPIRENONE	2261723
ETHINYL ESTRADIOL/DROSPIRENONE	2261731
ETONOGESTREL/ETHINYL ESTRADIOL	2253186
ETHINYL ESTRADIOL/NORELGEST	2248297
DIENESTROL	441295
DIETHYLSTILBESTROL	3360
DIETHYLSTILBESTROL	2091461
DIETHYLSTILBESTROL	2091488
ESTRADIOL	464791
ESTRADIOL	2148587
ESTRADIOL	464805

Supplementary material

ESTRADIOL	2148595
ESTRADIOL VALERATE	29238
ESTRADIOL	756849
ESTRADIOL	2237807
ESTRADIOL	2243722
ESTRADIOL	2245676
ESTRADIOL	756857
ESTRADIOL	2204428
ESTRADIOL	2231509
ESTRADIOL	2237808
ESTRADIOL	2243724
ESTRADIOL	2244000
ESTRADIOL	2246967
ESTRADIOL	756792
ESTRADIOL	2204444
ESTRADIOL	2231510
ESTRADIOL	2244002
ESTRADIOL	2246969
ESTRADIOL	2168898
ESTRADIOL	2204436
ESTRADIOL	2244001
ESTRADIOL	2246968
ESTRADIOL	2225190
ESTRADIOL	2204401
ESTRADIOL	2238704
ESTRADIOL	2243999
ESTRADIOL	2241332
ESTRADIOL	2247499
ESTRADIOL	2247500
ESTROGENS,CONJUGATED	2569
ESTROGENS,CONJUGATED	2043394
ESTROGENS,CONJUGATED	2230891
ESTROGENS,CONJUGATED	2239654
ESTROGENS,CONJUGATED	2577
ESTROGENS,CONJUGATED	265470
ESTROGENS,CONJUGATED	587281
ESTROGENS,CONJUGATED	2043408
ESTROGENS,CONJUGATED	2089
ESTROGENS,CONJUGATED	2043440
ESTROGENS,CONJUGATED	403466

Supplementary material

	ESTROGENS,CONJUGATED	2043416
	ESTROGENS,CONJUGATED	2230892
	ESTROGENS,CONJUGATED	2239655
	ESTROGENS,CONJUGATED	2585
	ESTROGENS,CONJUGATED	265489
	ESTROGENS,CONJUGATED	587303
	ESTROGENS,CONJUGATED	2043424
	ESTROGENS,CONJUGATED	2043432
	ESTROGENS,CONJUGATED	2043386
	ME-TESTOSTERONE/ESTROGEN,CON	53538
	ESTROPIPATE	282685
	ESTROPIPATE	2089769
	ESTROPIPATE	282677
	ESTROPIPATE	2089777
	ESTROPIPATE	2089793
	ESTRADIOL/NORETH AC	2108186
	NORGESTIMATE-ETHINYL ESTRADIOL	2229064
	NORGESTIMATE-ETHINYL ESTRADIOL	2229218
	NORGESTIMATE-ETHINYL ESTRADIOL	2229226
	ETHINYL ESTRADIOL/CYPROTERONE	2233542
	ETHINYL ESTRADIOL/NORELGEST	2246340
	NO GENERIC FORMULARY	66124057
	NO GENERIC FORMULARY	66124058
	NO GENERIC FORMULARY	66124060
	NO GENERIC FORMULARY	66124061
	NO GENERIC FORMULARY	66124062
	NO GENERIC FORMULARY	66124063
	NO GENERIC FORMULARY	66124064
Progestogens		
Progesterone		
	PROGESTERONE,MICRONIZED	2241013
	MEDROXYPROGESTERONE ACET	30848
	MEDROXYPROGESTERONE ACET	30856
	MEDROXYPROGESTERONE ACET	585092
	NO GENERIC FORMULARY	66123240
	MEDROXYPROGESTERONE ACET	708917
	MEDROXYPROGESTERONE ACET	2148552
	MEDROXYPROGESTERONE ACET	2221284
	MEDROXYPROGESTERONE ACET	2229838
	MEDROXYPROGESTERONE ACET	2244726

Supplementary material

	MEDROXYPROGESTERONE ACET	2246627
	MEDROXYPROGESTERONE ACET	30937
	MEDROXYPROGESTERONE ACET	2010739
	MEDROXYPROGESTERONE ACET	2148560
	MEDROXYPROGESTERONE ACET	2221292
	MEDROXYPROGESTERONE ACET	2229839
	MEDROXYPROGESTERONE ACET	2244727
	MEDROXYPROGESTERONE ACET	2246628
	MEDROXYPROGESTERONE ACET	729973
	MEDROXYPROGESTERONE ACET	2010933
	MEDROXYPROGESTERONE ACET	2148579
	MEDROXYPROGESTERONE ACET	2221306
	MEDROXYPROGESTERONE ACET	2229840
	MEDROXYPROGESTERONE ACET	2246629
	MEDROXYPROGESTERONE ACET	30945
	MEDROXYPROGESTERONE ACET	2267640
	NORETHINDRONE	37605
	PROGESTERONE,MICRONIZED	2166704
	PROGESTERONE	739952
	PROGESTERONE	1977652
	PROGESTERONE	2128470
	LEVONORGESTREL	2243005
Transmasculine		
Testosterone		
	TESTOSTERONE	2249499
	TESTOSTERONE CYPIONATE	30783
	TESTOSTERONE PROPIONATE	1977571
	TESTOSTERONE CYPIONATE	1977601
	TESTOSTERONE CYPIONATE	2220318
	TESTOSTERONE CYPIONATE	2246063
	TESTOSTERONE ENANTHATE	29246
	TESTOSTERONE ENANTHATE	716936
	TESTOSTERONE ENANTHATE	739944
	TESTOSTERONE UNDECANOATE	782327
	TESTOSTERONE ENANTHATE/ESTRAD	108278
	TESTOSTERONE ENANTHATE/ESTRAD	2061031
	TESTOSTERONE	2239653
	TESTOSTERONE	2245346
	TESTOSTERONE	2245345
	TESTOSTERONE	2245972

Supplementary material

Chronic condition case definitions

Chronic condition		Case definition	Codes
Cardiovascular disease*	Acute myocardial infarction	1 or more hospitalizations with relevant ICD codes	ICD-10: I21 Acute myocardial infarction I22 Subsequent myocardial infarction ICD-9: 410 Acute myocardial infarction
	Ischemic heart disease	At least one of the following: 2 medical visits with Angina ICD-9 code 413 plus 1 heart disease prescription in 365 days; or 1 specialist visit with Angina ICD-9 code 413 plus 1 prescription in 365 days; or 2 medical visits with two ICD9 codes 410, 411, 412, 413, 414 in 365 days; or 1 CCI/CCP CABG,PCI/PCTA procedure code; or 1 hospitalization with relevant ICD code.	ICD-10: I20 Angina pectoris I21 Acute myocardial infarction I 22 Subsequent myocardial infarction I23 Certain current complications following acute myocardial infarction I24 Other acute ischaemic heart diseases I25 Chronic ischaemic heart disease ICD-9: 410 Acute myocardial infarction 411 Other acute and subacute forms of ischaemic heart disease 412 Old myocardial infarction 413 Angina pectoris 414 Other forms of chronic ischaemic heart disease
	Chronic heart failure	1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	ICD-10: I50 Heart failure ICD-9: 428 Heart failure
	Stroke- hospital	1 or more hospitalizations with relevant ICD codes	ICD-10: H34.1 Central retinal artery occlusion I60 Subarachnoid hemorrhage I61 Intracerebral haemorrhage I63 Cerebral infarction (exclude I63.6 Cerebral infarction due to

Supplementary material

			cerebral venous thrombosis, nonpyogenic) I64 Stroke, not specified as haemorrhage or infarction 362.3 Retinal vascular occlusion 430 Subarachnoid hemorrhage 431 Intracerebral hemorrhage 433.x1 Occlusion and stenosis of precerebral arteries 434.x Occlusion cerebral arteries 436 Acute but ill-defined cerebrovascular disease *Excludes any traumatic brain injury*
	Transient ischemic attack	1 or more hospitalizations with relevant ICD codes	ICD-10: H34.0 Transient retinal artery occlusion G45.0 Vertebro-basilar artery syndrome G45.1 Carotid artery syndrome (hemispheric) G45.2 Multiple and bilateral precerebral artery syndromes G45.3 Amaurosis fugax G45.8 Other transient cerebral ischemic attacks and related syndromes G45.9 Transient cerebral ischemic attack, unspecified ICD-9: 435 Transient cerebral ischemia *Excludes any traumatic brain injury*
Chronic kidney disease*		1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	ICD-10: N01 Rapidly progressive nephritic syndrome N03 Chronic nephritic syndrome N04 Nephrotic syndrome N05 Unspecified nephritic syndrome N06 Isolated proteinuria with specified morphological lesion N07 Hereditary nephropathy, not elsewhere classified

Supplementary material

		N18 Chronic kidney disease N19 Unspecified kidney failure N26 Unspecified contracted kidney N27 Small kidney of unknown cause ICD-9: 581 Nephrotic syndrome 582 Chronic glomerulonephritis 583 Nephritis and nephropathy, not specified as acute or chronic 585 Chronic renal failure 586 Renal failure, unspecified 587 Renal sclerosis, unspecified 589 Small kidney of unknown cause
Chronic liver disease	1 or more hospitalization or medical visit with relevant diagnosis within 365 days	ICD-9: 571.0 Alcoholic fatty liver 571.2 Alcoholic cirrhosis of liver 571.3 Alcoholic liver damage, unspecified 571.4 Chronic hepatitis 571.5 Cirrhosis of liver without mention of alcohol 571.6 Billiary cirrhosis 571.8 Other chronic nonalcoholic liver disease 571.9 Unspecified chronic liver disease without mention of alcohol 070.3 Viral hepatitis B without mention of hepatic coma 070.30 Viral hepatitis B without mention of hepatic coma, acute or unspecified, without mention of hepatitis delta 070.31 Viral hepatitis B without mention of hepatic coma, acute or unspecified, with hepatitis delta 070.32 Viral hepatitis B without mention of hepatic coma, chronic, without mention of hepatitis delta 070.33 Viral hepatitis B without mention of hepatic coma, chronic, with hepatitis delta

Supplementary material

		070.52 Hepatitis delta without mention of active Hepatitis B disease or hepatic coma V02.61 Hepatitis B carrier 070.42 Hepatitis delta without mention of active Hepatitis B disease with hepatic coma 070.54 Chronic hepatitis C without mention of hepatic coma V02.62 Hepatitis C carrier
Chronic Obstructive Pulmonary Disease*	1 or more hospitalization or 2 or more medical visits within 365 days	ICD-10: J41 Simple and mucopurulent chronic bronchitis J42 Unspecified chronic bronchitis J43 Emphysema J44 Other chronic obstructive pulmonary disease ICD-9: 491 Chronic bronchitis 492 Emphysema 496 Chronic airways obstruction, not elsewhere classified
Diabetes Mellitus*	At least 1 of the following: 1 hospitalization or 2 medical visits in 365 days with relevant ICD codes; or 2 or more insulin prescriptions in 365 days; or 2 or more oral antihyperglycemic (not including metformin) prescriptions in 365 days; or 1 insulin and 1 oral antihyperglycemic (including metformin) in 365 days; or 2 metformin prescriptions and 1 medical visit in one year with relevant ICD codes.	ICD-10: E10 Type 1 diabetes mellitus E11 Type 2 diabetes mellitus E13 Other specified diabetes mellitus E14 Unspecified diabetes mellitus ICD-9: 250 Diabetes mellitus

Supplementary material

	Excludes gestational diabetes.	
Hypertension*	1 or more hospitalizations or 2 or more medical visits within 2 years with relevant ICD codes. *Excludes gestational hypertension.*	ICD-10: I10 Essential (primary) hypertension I11 Hypertensive heart disease I12 Hypertensive renal disease I13 Hypertensive heart and renal disease I15 Secondary hypertension ICD-9: 401 Essential hypertension 402 Hypertensive heart disease 403 Hypertensive renal disease 404 Hypertensive heart and renal disease 405 Secondary hypertension
Mood and anxiety disorders*	1 or more hospitalizations with a relevant ICD code or 2 or more medical visits with a relevant code within 2 years	ICD-10: F30 Manic episode F31 Bipolar affective disorder F32 Depressive episode F33 Recurrent depressive disorder F34 Persistent mood [affective] disorders F38 Other mood [affective] disorders F39 Unspecified mood [affective] disorder F40 Phobic anxiety disorders F41 Other anxiety disorders F42 Obsessive-compulsive disorder F43 Reaction to severe stress, and adjustment disorders F44 Dissociative (conversion) disorders F45 Somatoform disorders F48 Other neurotic disorders F68 Other disorders of adult personality & behavior ICD-9:

Supplementary material

			296 Affective psychoses 300 Neurotic disorders 311 Depressive disorder, not elsewhere classified MSP DX Code: 50B Anxiety/Depression
Non-AIDS defining cancer†	All prevalent cancer cases were included, with the exception of AIDS defining malignancies (Kaposi's sarcoma, non-Hodgkin's lymphoma, invasive cervical cancer)		*Cancer case definition details available from the British Columbia Cancer Agency: http://www.bccancer.bc.ca/health-info/types-of-cancer*
Organic mental disorders		1 or more medical visits or hospitalizations with relevant diagnoses within 365 days	ICD-9: 290.x Dementias 294.x Other organic psychotic conditions 331.x Alzheimer's ICD-10: F00.x Dementia in Alzheimer's disease F01.x Vascular Dementia F02.x Dementia in other disease classified elsewhere F03.x Unspecified dementia F04 Amnestic disorder due to physiological condition F06 Other mental disorders due to known physiological condition F09 Unspecified mental disorder due to known physiological condition G30 Alzheimer's disease with early onset
Osteoarthritis*		1 or more hospitalization or 2 or more medical visits in 365 days with a relevant ICD code	ICD-10: M15 Polyarthrosis M16 Coxarthrosis [arthrosis of hip] M17 Gonarthrosis [arthrosis of knee] M18 Arthrosis of first carpometacarpal joint M19 Other arthrosis

Supplementary material

		ICD-9: 715 Osteoarthrosis and allied disorders
Personality disorder	1 or more hospitalizations or medical visits with a relevant diagnosis within 365 days	ICD-9: 301.x Personality disorders ICD-10: F60.x Specified personality disorders F62 Enduring personality changes, not attributable to brain damage and disease F68.1 Intentional production or feigning of symptoms or disabilities, either physical or psychological F68.8 Other specified disorders or adult personality and behaviour F69 Unspecified disorder or adult personality and behaviour
Schizophrenia related disorder	1 or more medical visit or hospitalizations with relevant diagnoses within 365 days	ICD-9: 295.x Schizophrenic disorders 297.0 Paranoid state, simple 297.1 Delusional disorder 297.2 Paraphrenia 297.3 Shared psychotic disorder ICD-10: F20.x Paranoid schizophrenia F21.x Schizotypal disorder F23.2 Acute schizophrenia-like psychotic disorder F25.x Schizoaffective disorders

* Case definition adapted from British Columbia Ministry of Health version 2017, April 4 2019 update

† Case-definition adapted from British Columbia Cancer Agency

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	1.1-3: Title page
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			Introduction (pp 4-5)
Objectives	3	State specific objectives, including any prespecified hypotheses			Introduction (page 5)
Methods					
Study Design	4	Present key elements of study design early in the paper			Methods (page 5); Supplementary Material
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			Methods (page 5)

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	6.1: Methods (pp 6-7); Supplementary Material 6.2: Methods (pp 6-7) 6.3: Figure 1
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	7.1: Methods (pp 6-7); Supplementary Material
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			Methods (pp 6-7)

Bias	9	Describe any efforts to address potential sources of bias			Discussion (pp 9, 11)
Study size	10	Explain how the study size was arrived at			Methods (page 6); Figure 1
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Methods (pp 6-7)
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			Methods (pp 6-7)
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	12.1-2: Methods

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Methods (page 5)
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Results (page 8); Figure 1
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			Results (page 8)
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure			Results

		category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			NA
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			NA
Discussion					
Key results	18	Summarise key results with reference to study objectives			Discussion: Page 9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Discussion: Page 9, Page 11
Interpretation	20	Give a cautious overall interpretation of results considering objectives,			Conclusion (Page 11)

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results			Limitations (Page 9)
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			Funding (Page 12)
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Page 12

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.

BMJ Open

Development of a computable phenotype to identify a transgender sample for health research purposes: A feasibility study in a large linked provincial healthcare administrative cohort in British Columbia, Canada

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2020-040928.R3
Article Type:	Original research
Date Submitted by the Author:	03-Mar-2021
Complete List of Authors:	Rich, Ashleigh; The University of British Columbia, School of Population & Public Health; BC Centre for Excellence in HIV/AIDS, Poteat, Tonia; University of North Carolina at Chapel Hill, Department of Social Medicine Koehoorn, Mieke; University of British Columbia, School of Population and Public Health; Li, Jenny; BC Centre for Excellence in HIV/AIDS, Ye, Monica; BC Centre for Excellence in HIV/AIDS, Sereda, Paul; British Columbia Centre for Excellence in HIV/AIDS, Salway, Travis; Simon Fraser University Hogg, R; Simon Fraser University,
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	HIV/AIDS, Health services research, Research methods, Public health
Keywords:	EPIDEMIOLOGY, PUBLIC HEALTH, Sexual and gender disorders < PSYCHIATRY, HIV & AIDS < INFECTIOUS DISEASES, SOCIAL MEDICINE

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

TITLE PAGE

Title: Development of a computable phenotype to identify a transgender sample for health research purposes: A feasibility study in a large linked provincial healthcare administrative cohort in British Columbia, Canada

Authors: Rich AJ[1,2], Poteat T[3], Koehoorn M[1], Li J[2], Ye M[2], Sereda, P[2], Salway T[4], Hogg RS[2,4]

Affiliations:
1. School of Population and Public Health, University of British Columbia, Vancouver, Canada
2. British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada
3. Department of Social Medicine, University of North Carolina- Chapel Hill, Chapel Hill, USA
4. Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada

Corresponding author:
Ashleigh J Rich
ajrich@mail.ubc.ca

2206 East Mall
Vancouver, BC
V6T 1Z3
Canada

ABSTRACT

Objectives: Innovative methods are needed for identification of transgender people in administrative records for health research purposes. This study investigated the feasibility of using transgender-specific healthcare utilization in a Canadian population-based, health records database to develop a computable phenotype (CP) and identify the proportion of transgender people within the HIV-positive population as a public health priority.

Design: The COAST cohort comprises a data linkage between two provincial data sources: The BC Centre for Excellence in HIV/AIDS Drug Treatment Program, which coordinates HIV treatment dispensation across BC; and Population Data BC, a provincial data repository holding individual, longitudinal data for all BC residents (1996-2013).

Setting: British Columbia, Canada.

Participants: COAST participants include 13 907 BC residents living with HIV (≥19 years of age) and a 10% random sample comparison group of the HIV-negative general population (514 952 individuals).

Primary and secondary outcome measures: Healthcare records were used to identify transgender people via a CP algorithm (diagnosis codes + androgen blocker/hormone prescriptions), to examine related diagnoses and prescription concordance, and to validate the CP using an independent provider-report transgender status measure. Demographics and chronic illness burden were also characterized for the transgender sample.

Results: The best-performing CP identified 137 HIV-negative and 51 HIV-positive transgender people (total 188). In validity analyses, the best-performing CP had low sensitivity (27.5%, 95%CI:17.8-39.8), high specificity (99.8%, 95%CI:99.6-99.8), low agreement using Kappa statistics (0.3, 95%CI:0.2-0.5), and moderate positive predictive value (43.2%, 95%CI:28.7-58.9). There was high concordance between exogenous-sex hormone use and transgender-specific diagnoses.

Conclusions: The development of a validated CP opens up new opportunities for identifying transgender people for inclusion in population-based health research using administrative health data, and offers the potential for much-needed and heretofore unavailable evidence on health status, including HIV status, and the healthcare use and needs of transgender people.

KEYWORDS: Transgender Persons, Health Services, Algorithms, Canada

ARTICLE SUMMARY

Strengths and limitations of this study:

- This study demonstrates the feasibility of developing and validating a computable phenotype for identification of a transgender sample, using a population-based representative source population and healthcare records.
- A major contribution of this study is the ascertainment of the population of transgender people living with HIV in the Canadian province of British Columbia, in a universal

healthcare setting, using a computable phenotype, and capacity to estimate the prevalence of transgender status among the population living with HIV in the province.

- Development of a validated computable phenotype algorithm using diagnosis and prescription data to identify transgender samples in administrative data without other gender identity ascertainment measures lays the foundation for future investigation of transgender-specific research questions related to general and HIV-specific healthcare use and health outcomes for this key population.
- While administrative data is an invaluable resource for answering important health and healthcare utilization research questions, this study is limited to those transgender people accessing medical transition care in BC and may not represent the transgender population as a whole.

INTRODUCTION

Limited data on transgender people

Transgender people are often overlooked within epidemiological research and population health surveillance due to small sample size, limited research designs, and other institutional and methodological erasures.[1–3] A 2017 review of Medline-indexed literature from 1950 to 2016 found 2405 published articles including transgender people, with almost half published in the last decade.[4] A 2008 United States (US)-based meta-analysis of HIV prevalence among transgender populations found 24 studies of transgender women, and five additional studies of transgender men,[5] though an updated review found 43 primary studies on transgender women and 15 on transgender men published between 2006-2017.[6] Despite this recent increase in transgender health research in general and for HIV specifically, much of the literature has focused on transgender-specific care, mental health and HIV/sexual health,[7,8] leaving the population understudied, in particular in the broader areas of physical health and healthcare utilization.

The erasures or exclusions of transgender persons in health studies may be explained, in part, by methodological challenges. Specific to electronic health record (EHR) data, a 2017 report identified only one transgender person among 38,5820 cancer cases in a Minnesota cancer registry,[9] clearly an undercount given that 0.4% of the US general population and 0.6% (95% credibility intervals: 0.5%-0.7%) of the Minnesota population is estimated to be transgender.[10,11] This highlights the need for improved gender ascertainment and transgender inclusion in research relying on patient records and administrative data. The establishment of best practices for measuring transgender status in survey research, such as the two-step method (measuring sex assigned at birth and current gender identity), points to a way forward for transgender-inclusive population health research.[12,13] However, innovative research methods are needed to identify transgender people in studies that rely on existing data sources (in particular EHR) and that optimize the use of transgender respondents' data in non-transgender specific research.

Computable phenotypes for transgender health research

Previous research in transgender health largely comprises cross-sectional studies, case reports, and qualitative or observational research.[7] Much consists of clinic- or venue-based convenience samples or lack comparison groups.[7,8] The literature is further characterized by inconsistent transgender status measurement,[14] small sample sizes, and focus on the United States (US).[8] In response, researchers have called for advancing transgender health research methods - namely ascertainment of high-quality samples via systematic approaches - including for general population-based and health systems-based studies.[15] One opportunity for the advancement of transgender health research methods is the emerging use of computable phenotypes (CPs)[16] or case ascertainment algorithms, to identify transgender samples in healthcare utilization data. A computable phenotype is an algorithm for identifying a clinical feature, condition, or set of characteristics that can be determined directly from EHR and other ancillary health care data systems (e.g., disease registries, insurance claims data) data.[17] CPs are developed using a combination of data elements (e.g., sociodemographic variables, clinical diagnoses) and value sets (i.e., the selection of a set of relevant values for each data element). Development of CPs using standardized methods and definitions enables identification and inclusion of transgender persons in research, as well as replication of analyses across data

sources, healthcare organizations/sites and studies. CPs have application in clinical care, surveillance, and health research.

Recently, CP and other EHR-based algorithm methods have been applied in a number of settings primarily in the US to identify transgender samples for health research.[14] Specifically, the STRONG study identified a transgender cohort (n = 6,456) using EHR data from Kaiser Permanente health plan members in California and Georgia, for investigation of general and transgender-specific health outcomes.[18] Blosnich et al identified 3,177 people with a "gender identity disorder" diagnosis among military veterans accessing care through the US Veterans Health Administration healthcare system,[19] for examination of mental health and other outcomes. Researchers with the US Centers for Medicare & Medicaid Services identified 4,098 transgender beneficiaries using national Medicare claims data,[20] and researchers at Vanderbilt University identified 234 transgender patients in their university clinic EHR data.[16] While these cohorts represent important opportunities for advancement of transgender health research, these methods have yet to be applied widely outside the US context. This is particularly important as different jurisdictions may vary in medical billing and coding practices, healthcare system patient populations, and representativeness of the general population. Specifically, in Canada, healthcare is delivered through a provincially administered universal healthcare system. As such, research using EHR provides an opportunity to develop methods for population-based, representative estimates of transgender populations within the Canadian context. Coupled with the current absence of gender ascertainment measures in population-based routinely collected data (e.g., census, national government health surveys, etc.) in Canada and many other jurisdictions, this remains an evidence need.

Summary of study rationale

This study investigated the application of emerging transgender health research methods, specifically CPs, in a Canadian context for the first time, testing the feasibility of identification of a transgender sample using EHR data from a provincial healthcare administrative data-linked cohort.

METHODS

Data Sources and Participants

The Comparative Outcomes and Service Utilization Trends Study (COAST)

COAST is a population-based cohort study focused on health services utilization research questions among all people known to be living with HIV (PLWH) in the province of British Columbia (BC) and a 10% random sample comparison group of the HIV-negative general population.[21] The COAST cohort comprises individual-level, longitudinal data from PLWH who have ever accessed HIV treatment in BC between 1996 and 2013, provided by Population Data BC (PopDataBC)[21] via data linkage between two provincial data sources, by personal health number: the Drug Treatment Program (DTP) [22] and the Ministry of Health. PopDataBC provides infrastructure for access to, and linkage of, longitudinal and individual-level administrative health data for all BC residents.[23]. The HIV-negative general population cohort was drawn randomly from the Ministry of Health registry data by PopDataBC. The COAST study has received approval from the University of British Columbia/Providence Health Care Research Ethics Board (#H09-02905) and Simon Fraser University Office of Research Ethics (#2013 s0566). The study complies with the BC Freedom of Information and Protection of

Privacy Act (FIPPA) and did not require informed consent as it is conducted using retrospective administrative and anonymized data for research and statistical purposes only. No patients or public were involved in this study.

Drug Treatment Program

In BC, antiretroviral therapy (ART) is provided to PLWH at no cost to the patient, and distributed through the DTP.[22] The DTP contributed a provider-reported measure of transgender status for COAST.

Ministry of Health

Ministry of Health data available via COAST included insured medical service billing records for outpatient visits,[24,25] hospital (in-patient) visits,[26] prescription medications,[27,28] and vital statistics.[29]

Measures & Analyses

Transgender computable phenotypes

Identification of transgender cases was tested in COAST using International Classification of Disease (ICD) codes (9th and 10th editions) and exogenous sex hormone prescription use. Transgender-specific diagnoses in medical and hospital billing records included the ICD-9 codes 302.5 Trans-sexualism with unspecified history, 302.51 Trans-sexualism with asexual history, 302.52 Trans-sexualism with homosexual history, 302.53 Trans-sexualism with heterosexual history, 302.6 Gender Identity Disorder in children, 302.85 Gender Identity Disorder in adolescents or adults; and ICD-10 codes F64.0 Gender Identity Disorder of childhood, F64.2 Gender Identity Disorder of childhood, F64.8 Other Gender Identity Disorder, and F64.9 Gender Identity Disorder unspecified. The full list of androgen blockers and exogenous sex hormone prescriptions included in analyses is available in the supplementary material.

Concordance

To assess face validity and utility of diagnosis and prescription data over time in CP development (i.e. whether the identified transgender sample had exogenous sex hormone prescription use and other diagnoses patterns consistent with that of transgender populations in other studies), concordance analyses evaluated the presence of at least one included diagnosis and prescription during the COAST study follow-up period with the presence of at least one included diagnosis and prescription in the last study year. Concordance was assessed between transgender-specific diagnoses, exogenous sex hormone and androgen blocker prescriptions, and non-transgender specific diagnoses (ICD-9 259.9 Unspecified Endocrine Disorder and ICD-10 E34.9 Endocrine Disorder, Unspecified [see supplementary material]). Endocrine disorder diagnosis codes are sometimes preferred by medical providers treating transgender people in response to historic exclusions of transgender-specific care from insurance coverage and to combat the stigma of transgender-specific diagnosis codes that have historically been classified as psychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM).[30] Exogenous sex hormone use, while common in transgender populations,[5,31] is not transgender-specific. Cisgender populations also use androgen blocker and sex hormone prescriptions (e.g. estrogen to treat menopausal symptoms in cisgender women, spironolactone is used for hypertension), thus exogenous sex hormone and androgen blocker prescription use

cannot independently identify transgender people. At the same time, not all transgender people use hormones and some access via non-medical sources.[32,33]

Validation

In British Columbia, transgender status is collected in the DTP via a provider-reported sex variable ("Male", "Male to Female", "Female to Male", or "Female"). Patients reported as either "Male to Female" or "Female to Male" were classified as transgender. The provider-reported transgender measure, available for the HIV positive cohort only, was used as a 'gold standard' for CP validation. Sensitivity, specificity, positive predictive value and kappa statistics with corresponding 95% confidence intervals (CI) were calculated for identifying transgender people via the CPs, in the HIV positive cohort only. Follow-up time (mean and range) for each CP group was also produced.

Demographics and chronic conditions

To further assess face validity of the transgender CP for future health research, descriptive statistics were calculated for the transgender sample produced via application of the best performing CP from the validation analysis to both the COAST HIV-positive and HIV-negative cohorts. Descriptive statistics included COAST study key sociodemographic and health variables, specifically laboratory confirmed HIV serostatus (HIV-positive/HIV-negative), baseline age, patient's Health Authority (five provincial regions for the administration of health services that include large urban centres, suburban regions, and rural/remote areas), and chronic illness burden based on standardized case definitions from the BC Ministry of Health [34] and the BC Cancer Agency.[35]

RESULTS

The total COAST cohort included 528 859 people, of which 514 952 were HIV-negative (10% general population random sample) and 13 907 were PLWH (Figure 1).

[Figure 1 here]

Concordance

Of the 237 people who had ever had a transgender-specific diagnosis during the study period, 19.4% also had a recent diagnosis in the last follow-up year (Table 1). None had an unspecified endocrine disorder diagnosis at any time; thus, this diagnosis was excluded from all CPs. Of the 237, 79.3% had an exogenous sex hormone or androgen blocker prescription at least once during the study period and 46.4% had one in the last year.

Table 1. Concordance analyses for diagnoses and hormone measures

	N	%
≥ Transgender ICD- ever	237	100
≥ Transgender ICD- recent	46	19.4
Unspecified endocrine disorder use- ever	0	0.0
Unspecified endocrine disorder use- recent	0	0.0
≥ Hormone/blocker use- ever	188	79.3
≥ Hormone/blocker use-recent	110	46.4

Validation

While no one CP consistently performed well across all validation metrics, the CP with the best overall performance across test statistics was based on having received at least one transgender-specific diagnosis and at least one androgen blocker/exogenous sex hormone prescription over the study follow-up period (Table 2). This CP had high specificity (99.8%, 95% CI: 99.6-99.8), low sensitivity (27.5%, 95% CI: 17.8-39.8), low to moderate Kappa coefficients (0.3, 95% CI: 0.2-0.5) and moderate positive predictive values (43.2%, 95% CI: 28.7-58.9). This CP also had the second longest mean follow-up time (mean: 136.3, range: 21.0-203.0), similar overall to the other CP groups (mean: 136.5, range: 21.0-203.0; mean: 117.1, range: 24.0-198.0; mean: 130.4, range: 69.0-198.0; respectively).

Table 2. Validation measures of transgender computable phenotype (CP) with provider-report transgender status measures, in COAST HIV-positive cohort

CP	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	Kappa (95% CI)
≥ 1 transgender ICD- ever	27.5 (17.8, 39.8)	99.7 (99.6, 99.8)	40.4 (26.7, 55.7)	0.3 (0.2, 0.4)
≥ 1 transgender ICD- recent	8.7 (3.6, 18.6)	100.0 (99.9, 100.0)	85.7 (42.0, 99.4)	0.2 (0.1, 0.3)
≥ 1 transgender ICD AND ≥ 1 hormone/blocker use- ever	27.5 (17.8, 39.8)	99.8 (99.6, 99.8)	43.2 (28.7, 58.9)	0.3 (0.2, 0.5)
≥ 1 transgender ICD AND ≥ 1 hormone/blocker use- recent	7.3 (2.7, 16.8)	100.0 (99.9, 100.0)	83.3 (36.5, 99.1)	0.1 (0.0, 0.2)

Transgender phenotype

Applying the best-performing CP, 137 HIV-negative people and 51 HIV-positive people (188 total) were identified as transgender in the respective COAST cohorts (Figure 1).

Demographics and chronic conditions

Demographic characteristics and chronic conditions for the 188 transgender people identified via the best-performing CP are presented in Figures 2 to 4. Transgender people were geographical located throughout BC health regions. The Vancouver Coastal Health Authority region, which includes the largest municipal area in BC, had the highest concentration of transgender people (44.2%) while the Northern Health Authority region - a predominantly rural and remote area of the province - had the lowest (1.6%).[36] The HIV-positive group had a higher median age than the HIV-negative group (35 [Q1, Q3: 30,42] and 30 [Q1, Q3: 19,42], respectively). For the HIV-negative sample, the largest proportion of transgender people were aged 19 to 29 years (44.5%) and the smallest proportion aged 55 years and older (<3%). For the HIV-positive sample, the largest proportion were aged 30 to 34 years (25.5%) and the smallest proportion aged 55 years and older (<2%).

[Figures 2 and 3 here]

Overall, HIV-positive transgender people had a higher prevalence of at least one chronic condition (other than HIV) compared to HIV-negative transgender people (88.2% versus 85.4%, respectively), and of two or more chronic conditions (76.5% versus 52.6%, respectively). Specific chronic disease differences between transgender people living with and without HIV were most notable for a higher prevalence among the HIV-positive cohort of cardiovascular disease, chronic kidney disease, osteoarthritis, schizophrenia and personality disorders, and chronic liver disease, but a lower prevalence for hypertension.

[Figure 4 here]

DISCUSSION

This study demonstrates the feasibility of identification of a sample of transgender people in a large linked provincial healthcare administrative database, using a CP based on prescriptions and diagnoses. Among a growing number of studies using EHR and CP methods to identify transgender samples for health research purposes, this is the first to do so in Canada., to independently validate the CP using a 'gold standard' of provider-reported transgender status, and the only to use population-based data.

Concordance

There was high concordance between transgender-specific diagnoses and exogenous sex hormone or androgen blocker prescription use in this study. That nearly half of those with at least one transgender-specific diagnosis had been dispensed hormones or blockers in the past year is consistent with findings from US and Canadian studies (48.9% and 43.0%, respectively)[20,32,33] - suggesting face validity for the current CP.

CP development and validation

The best-performing CP overall successfully identified cisgender people who were truly cisgender (specificity) and correctly identified transgender people who were truly transgender (0.2% false positive rate, results not shown). However, the selected CP had relatively low sensitivity, missing approximately 72.5% of 'true' transgender people in COAST, as identified by the gold standard provider-based measure. Though a relatively small proportion of the 'true' transgender sample was identified in this study, the impact on future analyses comparing health outcomes for transgender and cisgender groups is likely negligible, as even the large proportion of 'true' transgender people misclassified as cisgender (approximate n=496) is a very small proportion of the total COAST sample. At worst, this misclassification would bias results related to disparities between transgender and cisgender health toward the null, producing a conservative attenuated effect in COAST, and other such administrative datasets. Further, as discussed below, gender identity classification will likely greatly improve as transgender care shifts further into the fee-for-service system in BC. As in other Canadian administrative data studies, low sensitivity may be explained in part by provider and system billing preferences using 3-digit ICD diagnosis coding instead of the more specific 4-digit coding, and inconsistencies in the BC billing management system.[37] Despite the low sensitivity, CP development in this study with high specificity offers an advancement for transgender health research. A measure that correctly identifies cases for transgender samples in research with good success translates to better

opportunities to include transgender people in health studies and to investigate their health relative to other groups. While future research may lead to improvements in CP development, the CP identified in the current study with good specificity, albeit relatively poor sensitivity, has important utility in advancing opportunities in transgender health research. Additionally, while differential follow-up time can affect algorithm performance, the similar mean and range follow-up time for all CPs in this study suggests that differential follow-up time was not an important source of bias in this study.

The limited agreement between the CP and provider-report transgender status may be due to the widely varying transgender status prevalence depending on study design and ascertainment measures used.[14] In the BC context, the CP and the DTP measures are assessing transgender status in different ways and for different purposes. In the DTP, transgender status is ascertained in the context of HIV diagnosis and ART prescribing, during which demographics and HIV transmission risk factors are recorded. This differs from recording diagnoses in EHR for those accessing transgender-specific care as utilized in the CP. This may explain the lower PPV for the best-performing CP compared to the CP based on recent transgender diagnoses, suggesting the DTP provider-reported transgender status measure has better coverage for recent cases and the potential for use of recent diagnosis over ever to be beneficial in future CP development. Ultimately, a single CP may not be sufficient for all intended purposes and the best applicable CP (using different types of diagnoses, prescriptions or procedures) may differ depending on the intended healthcare, health research, or health policy application.[17]

There is limited literature on EHR-based studies with the ability to validate an administrative transgender measure using a 'gold standard' comparison measure.[16] The two previous studies that have developed and validated algorithms to identify transgender individuals have both been conducted in non-representative samples in the US, one using Medicare data[38] and one in a university medical center.[16] Similar to the current study, the Medicare study found high specificity when comparing an EHR-based and a two-step survey-based transgender measure. However, the Medicare study found that the EHR measure performed consistently well with high sensitivity and a high Kappa statistic, unlike in the current study. Using chart review as the 'gold standard' for comparison of transgender status, Ehrenfeld et al. found a low false positive rate for their best-performing algorithm (3%), though not as low as the false positive rate in the current study. The overall high levels of agreement for transgender measures in the two previous studies is likely a function of the lack of independence between the 'gold standard' and the CP or algorithm measures. Specifically, only those classified as transgender in the Medicare EHR data were offered survey participation to complete the two-step 'gold standard' survey measure, and only those cases identified as transgender in the university clinic EHR were included in chart review. Thus, previous studies could assess agreement between the two measures, but not robustly validate either. In the current study, the DTP provider-based transgender status measure is independent and thus could be used for robust CP validation.

While not possible to incorporate free-text records in case-finding algorithms in the current study as only structured EHR data is linked through COAST, it is worth noting the opportunities potentiated by use of NLP and machine learning approaches as methods for identifying transgender samples in EHR data as this research area continues to grow. Outside of transgender health, the use of NLP and machine learning to mine unstructured free-text EHR data has

demonstrated efficiency in improving case ascertainment algorithm accuracy .[39] As 'gold standard' two-step sex assigned at birth and current gender identity measures of transgender status[12] are slowly being implemented in routinely collected healthcare data sources, in the meantime NLPs to extract free-text data can be used to produce better gold standards against which to measure algorithm performance, as demonstrated by the Medicare study.[38]

Transgender status prevalence & ascertainment

Based on a recent meta-analysis of transgender status prevalence in population-based probability samples,[10] it was expected that an effective CP would identify 0.4% of the general population as transgender, or approximately 54 of the HIV-positive COAST cohort (n=13 907) and 3,098 of the HIV-negative cohort (n=516 340). Consistent with expectations, the best-performing CP identified 51 PLWH as transgender, equivalent to a transgender status prevalence of 0.4% among PLWH. Contrary to expectations, the best-performing CP identified less than 5% of the number of transgender persons expected in the HIV-negative cohort. This is likely a result of a number of factors including the limitation of CPs to the subset of a population accessing care as noted, and the result of most transgender people in BC receiving care currently outside the main fee-for-service healthcare delivery system. However, it is also consistent with the undercount of transgender populations using diagnostic criteria compared to other methods of ascertainment demonstrated in other studies.[14]

Using the broadest CP algorithm (any transgender-specific diagnosis ever, n=56) and those identified by provider-report together (total n=106), the total transgender PLWH sample would represent as high as 0.88% (range: 0.73-1.1%) of the prevalent HIV infections in BC in 2014.[40] This overrepresentation of transgender people among PLWH is consistent with evidence of a disproportionate HIV burden for transgender populations globally,[5,41,42] as well as in line with the only other available data on the proportion of PLWH who are transgender, from US national surveillance data (2012 data: 1.1%, 95%CI: 0.8-1.4).[43]

Demographics and chronic conditions

Despite moderate to low performance by some validation metrics, particularly low sensitivity, the CP was able to detect meaningful results in the characterization of demographics and chronic condition burden for the transgender sample - supporting CP face validity. The population density and age distribution by HIV-status of transgender people in this study is largely consistent with general population patterns, as well as the larger COAST cohort.[21,36] The overall higher burden of chronic illness for transgender people living with HIV versus without HIV in this study is consistent with elevated chronic illness risk and morbidity among non-transgender PLWH.[44] This higher chronic disease burden is linked to HIV disease processes and related inflammatory immune response.[45] While a small but growing number of studies have begun to investigate the chronic illness burden for transgender populations in other industrialized settings,[16,19,46–48] including using EHR data, findings vary widely due to differences in sampling, study design, setting and measurement.

Limitations

Findings from this study should be interpreted in the context of a few key limitations. CPs are by design only applicable to people accessing healthcare services, often motivated by illness and aided by the ability to access care. As such, this study is limited to those transgender people

accessing medical transition care in BC and may only represent 24% to 47% of the total transgender population.[33] This study was also limited by the inability to validate the transgender CP among the HIV-negative COAST cohort, as a 'gold standard' provider-based transgender measure was only available for the HIV-positive cohort. It is possible that the transgender CPs would perform differently in populations living without HIV, particularly as healthcare contact is higher among populations living with HIV. Additionally, this study should be considered in light of the context in which it was conducted, an environment in which transgender healthcare delivery in BC is currently shifting from specialized care settings to the main primary care fee-for-service settings. Given that COAST only includes fee-for-service data, this study was limited by the inability to capture transgender people who access transgender care outside the fee-for-service system. However, fortunately, as the shift to the fee-for-service system occurs, transgender ascertainment via CPs in BC will likely improve. The administrative data used in this study may also be susceptible to coding error (and coding biases/practices) across conditions and settings,[49] potentially introducing misclassification bias in terms of transgender ascertainment. Finally, chronic condition prevalence data reported in this study should be interpreted with caution, given potential selection bias by serostatus in the COAST cohort; though any such bias likely resulted in conservative estimates of difference by serostatus in this analysis.

CONCLUSION

This study makes a number of important contributions to the literature on innovative methods in transgender health. Major contributions include development and validation of a transgender CP, using a population-based representative source population, in the Canadian context. Another strength is the approximately complete ascertainment of the population of transgender PLWH in BC, and capacity to estimate transgender status prevalence among PLWH. In a current funding environment of limited support for longitudinal transgender health studies in the US and none to date in Canada, this study and the methods employed offer an efficient, replicable and cost-effective way forward in creating electronic cohorts for advancing transgender health research.[15] Moreover, the recent rollback of sexual orientation and gender identity data collection and legal changes in insurance coverage of transgender healthcare in the US potentiate decline in accurate claims coding for gender-affirming care.[30] This highlights the utility of work in this area from other jurisdictions, particularly those with transgender-inclusive universal healthcare systems such as Canada.

Future research should build upon the methods developed in this study and explore complimentary approaches for gender identity ascertainment in administrative and EHR data, such as machine learning approaches, as have been used to develop algorithms based on healthcare utilization data in other research areas. Finally, the current study lays the foundation for future work with the ability to study transgender health and healthcare use patterns over time, with linkage to laboratory data, as well as inclusion of appropriate comparison groups.[15,50]

ACKNOWLEDGEMENTS & DISCLAIMER

The authors thank the COAST study participants, the BC Centre for Excellence in HIV/AIDS, the BC Ministry of Health, BC Vital Statistics Agency, PharmaNet and the institutional data stewards for granting access to the data, and Population Data BC for facilitating the data linkage process. In addition, we would like to thank the COAST core team members and other support

staff at these institutions for their administrative assistance with the data access and preparation. All inferences, opinions, and conclusions drawn in this paper are those of the authors, and do not reflect the opinions or policies of the Data Steward(s). There are no conflicts of interest to declare.

COMPETING INTERESTS

None declared.

FUNDING

This work was supported by the Canadian Institutes of Health Research, through an Operation Grant [grant number 130419], a Foundation Award to RSH [grant number 143342], and a Doctoral Research Award to AJR [grant number 152382] and support from the British Columbia Centre for Excellence in HIV/AIDS. The DTP receives funding from the provincial government of British Columbia (PharmaCare). The funders had no role in the study design, analysis, interpretation of the data, drafting of the manuscript or in the decision to submit for publication.

CONTRIBUTORS

AJR led the study from conceptualization to analysis plan to interpretation, drafting of the first manuscript version, revisions and final version. RSH acquired study data and funding. TP, MK, PS, TS, and RSH all contributed to study design, interpretation of results, and reviewed manuscript versions. JL and MY contributed to study analysis and reviewed manuscript versions. All authors provided critical review of first and subsequent manuscript drafts, approved the final version, and agree to be accountable for the work presented.

PATIENT AND PUBLIC INVOLVEMENT

No patients involved.

DATA SHARING STATEMENT

The data used for this study are held by the BC Centre for Excellence in HIV/AIDS under the authority of the BC Ministry of Health; as they contain confidential patient health records including HIV serostatus, data cannot be made available to other parties.

FIGURES LEGENDS

Figure 1. Total transgender sample identified using a computable phenotype with electronic health records

Figure 2. Geographic distribution of transgender people across province, by health authority*

**% of transgender individuals with known health authority (n=182)*

Figure 3. Age distribution of transgender sample, by HIV serostatus

Figure 4. Co-morbidities among transgender sample, by HIV serostatus

REFERENCES

1. Bauer GR. Making sure everyone counts: Considerations for inclusion, identification, and analysis of transgender and transsexual participants in health surveys. What Differ Sex Gend Make Gend Sex Health Res Casebook. Vancouver, Canada: Canadian Institutes of Health Research (CIHR), Institute for Gender and Health (IGH); 2012. p. 59–67.

2. Bauer GR, Hammond R, Travers R, Kaay M, Hohenadel KM, Boyce M. "I don't think this is theoretical; this is our lives": How erasure impacts health care for transgender people. JANAC. 2009;20:348–361.

3. Poteat T, German D, Flynn C. The conflation of gender and sex: Gaps and opportunities in HIV data among transgender women and MSM. Glob Public Health. 2016;11:835–848.

4. Wanta JW, Unger CA. Review of the transgender literature: Where do we go from here? Transgender Health. 2017;2:119–128.

5. Herbst JH, Jacobs ED, Finlayson TJ, McKleroy VS, Neumann MS, Crepaz N, et al. Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a systematic review. AIDS Behav. 2008;12:1–17.

6. Becasen JS, Denard CL, Mullins MM, Higa DH, Sipe TA. Estimating the prevalence of HIV and sexual behaviors among the US transgender population: A systematic review and meta-analysis, 2006–2017. Am J Public Health. 2019;109:e1–8.

7. Marshall Z, Welch V, Thomas J, Brunger F, Swab M, Shemilt I, et al. Documenting research with transgender and gender diverse people: protocol for an evidence map and thematic analysis. Syst Rev. 2017;6:35.

8. Reisner SL, Poteat T, Keatley J, Cabral M, Mothopeng T, Dunham E, et al. Global health burden and needs of transgender populations: a review. The Lancet. 2016;388:412–436.

9. Cathcart-Rake EJ, Lightner DJ, Quevedo FJ, Ertz DC, Jatoi A. Cancer in transgender patients: One case in 385,820 is indicative of a paucity of data. J Oncol Pract. 2017;JOP2017027714.

10. Meerwijk EL, Sevelius JM. Transgender population size in the United States: A meta-regression of population-based probability samples. Am J Public Health. 2017;107:e1–e8.

11. Flores AR, Herman JL, Gates GJ, Brown TNT. How Many Adults Identify as Transgender in the United States? [Internet]. Los Angeles, CA: Williams Institute; 2016 Jun p. 13. Available from: https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf

12. The GenIUSS Group. Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys. Los Angeles, CA: The Williams Institute; 2014.

13. Bauer GR, Braimoh J, Scheim AI, Dharma C. Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. PLOS ONE. 2017;12:e0178043.

14. Collin L, Reisner SL, Tangpricha V, Goodman M. Prevalence of transgender depends on the "case" definition: A systematic review. J Sex Med. 2016;13:613–626.

15. Reisner SL, Deutsch MB, Bhasin S, Bockting W, Brown GR, Feldman J, et al. Advancing methods for US transgender health research. Curr Opin Endocrinol Diabetes Obes. 2016;23:198–207.

16. Ehrenfeld JM, Gottlieb KG, Beach LB, Monahan SE, Fabbri D. Development of a Natural Language Processing Algorithm to Identify and Evaluate Transgender Patients in Electronic Health Record System. Ethn Dis. 2019;29:441–50.

17. Richesson R, Smerek M, Rusincovitch S, Zozus MN, Chaudhuri PS, Hammond WE, et al. Electronic Health Records-Based Phenotyping. In: Uhlenbrauck G, editor. NIH Collab Living Textb Pragmatic Clin Trials. Bethesda, Maryland; 2014.

18. Quinn VP, Nash R, Hunkeler E, Contreras R, Cromwell L, Becerra-Culqui TA, et al. Cohort profile: Study of Transition, Outcomes and Gender (STRONG) to assess health status of transgender people. BMJ Open. 2017;7:e018121.

19. Blosnich JR, Brown GR, Shipherd Phd JC, Kauth M, Piegari RI, Bossarte RM. Prevalence of gender identity disorder and suicide risk among transgender veterans utilizing veterans health administration care. Am J Public Health. 2013;103:e27–32.

20. Proctor K, Haffer SC, Ewald E, Hodge C, James CV. Identifying the transgender population in the Medicare program. Transgender Health. 2016;1:250–265.

21. Eyawo O, Hull MW, Salters K, Samji H, Cescon A, Sereda P, et al. Cohort profile: the Comparative Outcomes And Service Utilization Trends (COAST) Study among people living with and without HIV in British Columbia, Canada. BMJ Open. 2018;8:e019115.

22. Drug Treatment Program [Internet]. BC Cent. Excell. HIVAIDS. 2010 [cited 2019 Sep 27]. Available from: http://www.cfenet.ubc.ca/drug-treatment-program

23. Population Data BC. About PopData [Internet]. Popul. Data BC. 2017 [cited 2019 Sep 27]. Available from: https://www.popdata.bc.ca/aboutus

24. British Columbia Ministry of Health (2014). Medical Services Plan (MSP) Payment Information File [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

25. British Columbia Ministry of Health (2014). Consolidation File (MSP Registration & Premium Billing) [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

26. Canadian Institute for Health Information (2014). Discharge Abstract Database (Hospital Separations) [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

27. British Columbia Ministry of Health (2014). PharmaCare [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

28. British Columbia Ministry of Health (2014). PharmaNet [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

29. British Columbia Vital Statistics Agency (2014). Vital Statistics Deaths [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca

30. Arnold JD, Nelson AE, Loubier EM. Trends in Insurance Coverage for Gender-Affirming Surgeries. JAMA Surg. 2018;153:972.

31. Radix A, Sevelius J, Deutsch MB. Transgender women, hormonal therapy and HIV treatment: a comprehensive review of the literature and recommendations for best practices. J Int AIDS Soc. 2016;19:20810.

32. Rotondi NK, Bauer GR, Scanlon K, Kaay M, Travers R, Travers A. Nonprescribed hormone use and self-performed surgeries: "do-it-yourself" transitions in transgender communities in Ontario, Canada. Am J Public Health. 2013;103:1830–1836.

33. Scheim AI, Bauer GR. Sex and gender diversity among transgender persons in Ontario, Canada: results from a respondent-driven sampling survey. J Sex Res. 2015;52:1–14.

34. Chronic Disease Information Working Group. BC Chronic Disease and Selected Procedure Case Definitions [Internet]. British Columbia, Canada: BC Ministry of Health; 2019 Apr p. 1–132. Available from: https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf

35. British Columbia Cancer Agency. BC Cancer Registry Data [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from: http://www.popdata.bc.ca/data

36. BC Stats. Sub-provincial Population Estimates [Internet]. 2018 [cited 2019 Feb 1]. Available from: file://localhost/Users/AJRich/Zotero/storage/5VM2LBUN/PopulationEstimates.html

37. Hu W. Diagnostic Codes in MSP Claim Data. Program Monitoring and Information Management Branch, Resource Management Division, Medical Services Plan; 1996 p. 39.

38. Gerth J, Becerra-Culqui T, Bradlyn A, Getahun D, Hunkeler EM, Lash TL, et al. Agreement between medical records and self-reports: Implications for transgender health research. Rev Endocr Metab Disord. 2018;19:263–269.

39. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al. Development of phenotype algorithms using electronic medical records and incorporating natural

language processing. The BMJ [Internet]. 2015 [cited 2020 Oct 22];350. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4707569/

40. BC Centre for Disease Control. HIV in British Columbia: Annual Surveillance Report 2015. Vancouver, Canada: BC Centre for Disease Control; 2017.

41. Poteat T, Scheim A, Xavier J, Reisner S, Baral S. Global epidemiology of HIV infection and related syndemics affecting transgender people. J Acquir Immune Defic Syndr. 2016;72 Suppl 3:S210–9.

42. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV in transgender women: a systematic review and meta-analysis. Lancet Infect Dis. 2013;13:214–222.

43. Centers for Disease Control and Prevention. Behavioral and Clinical Characteristics of Persons Receiving Medical Care for HIV Infection—Medical Monitoring Project, United States, 2012 [Internet]. Washington, D.C.: Centers for Disease Control; 2015 Aug p. 42. Report No.: 12. Available from: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf

44. Grabovac I, Veronese N, Stefanac S, Haider S, Jackson SE, Koyanagi A, et al. Human Immunodeficiency Virus Infection and Diverse Physical Health Outcomes: An Umbrella Review of Meta-analyses of Observational Studies. Clin Infect Dis. 2019;ciz539.

45. Nasi M, De Biasi S, Gibellini L, Bianchini E, Pecorini S, Bacca V, et al. Ageing and inflammation in patients with HIV infection: Ageing, inflammation and HIV infection. Clin Exp Immunol. 2017;187:44–52.

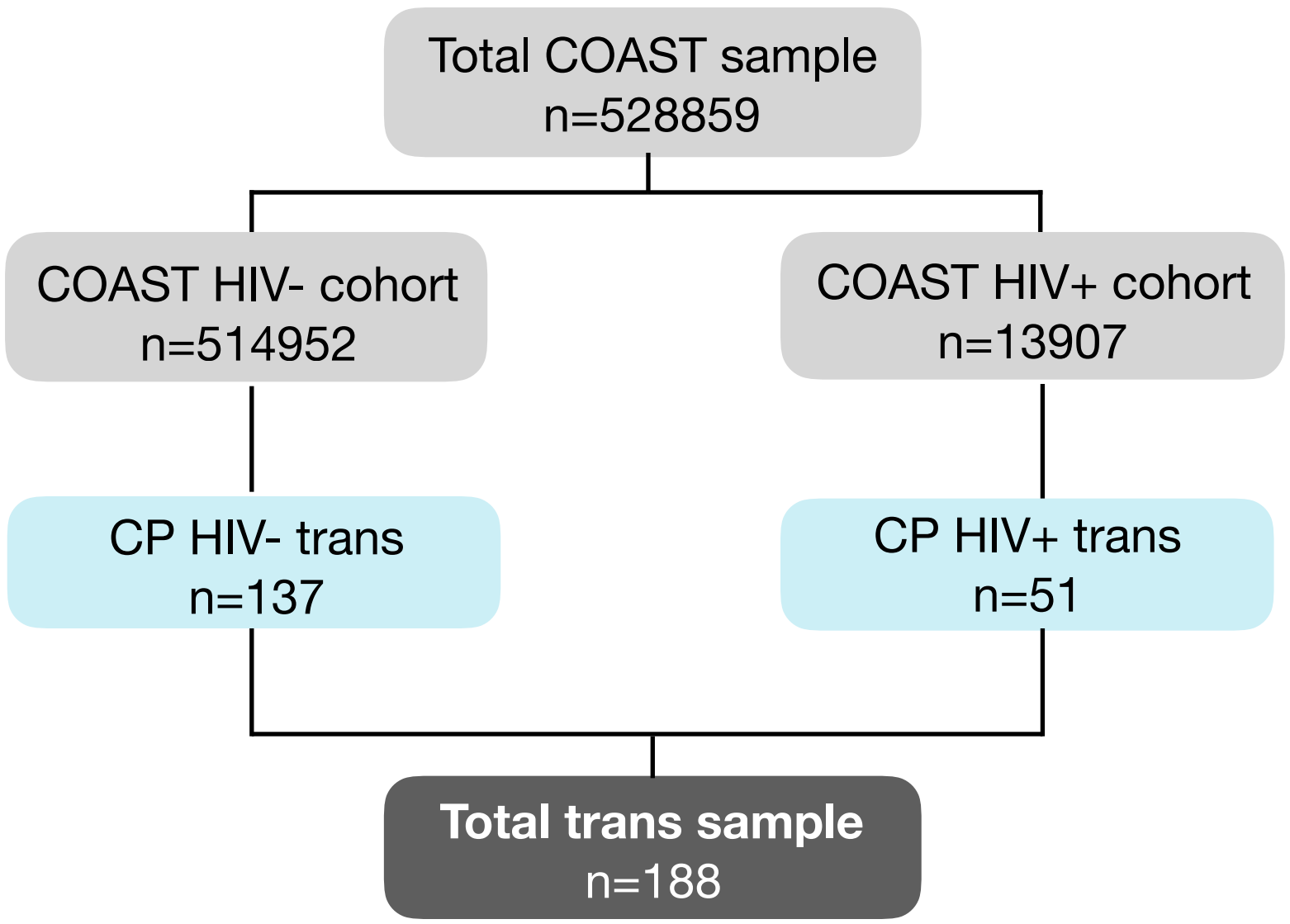
46. Judge C., O'Donovan C., Callaghan G., Gaoatswe G., O'Shea D. Gender dysphoria - prevalence and co-morbidities in an Irish adult population. Front Endocrinol. 2014;5:87.

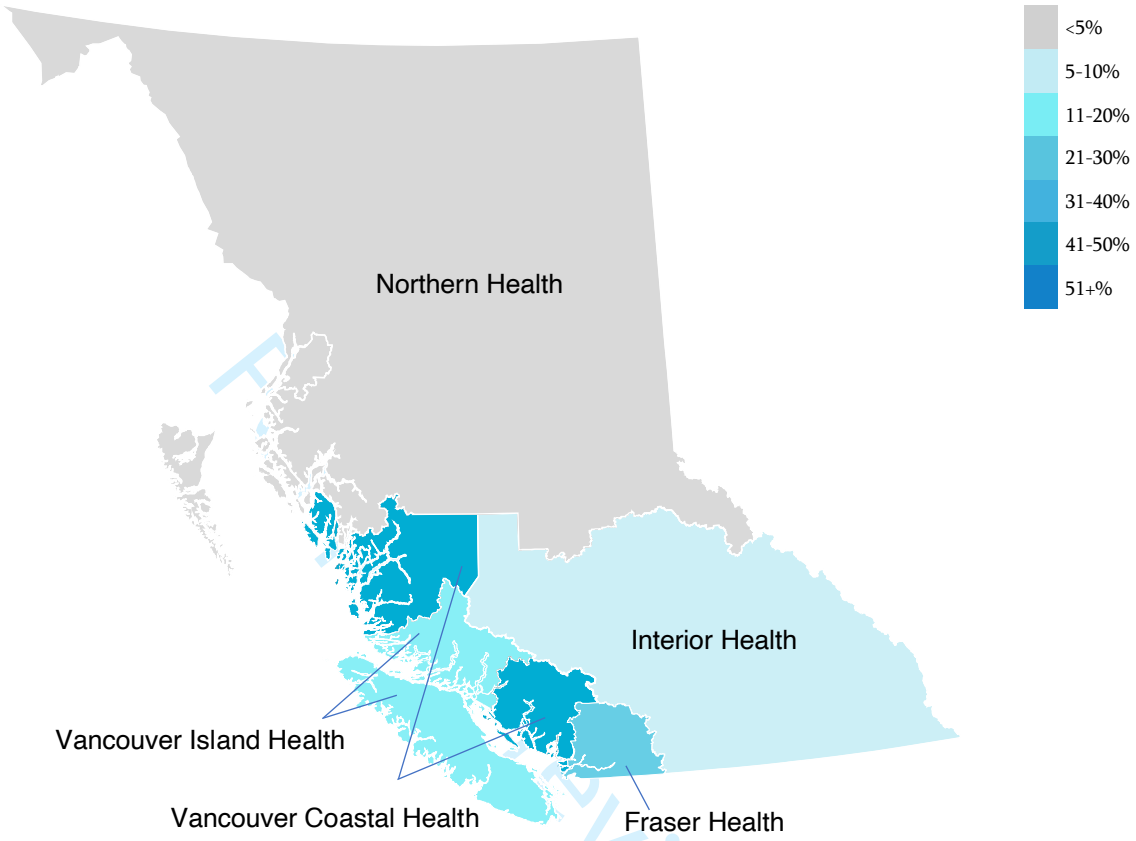
47. Cheung AS, Ooi O, Leemaqz S, Cundill P, Silberstein N, Bretherton I, et al. Sociodemographic and clinical characteristics of transgender adults in Australia. Transgender Health. 2018;3:229–38.

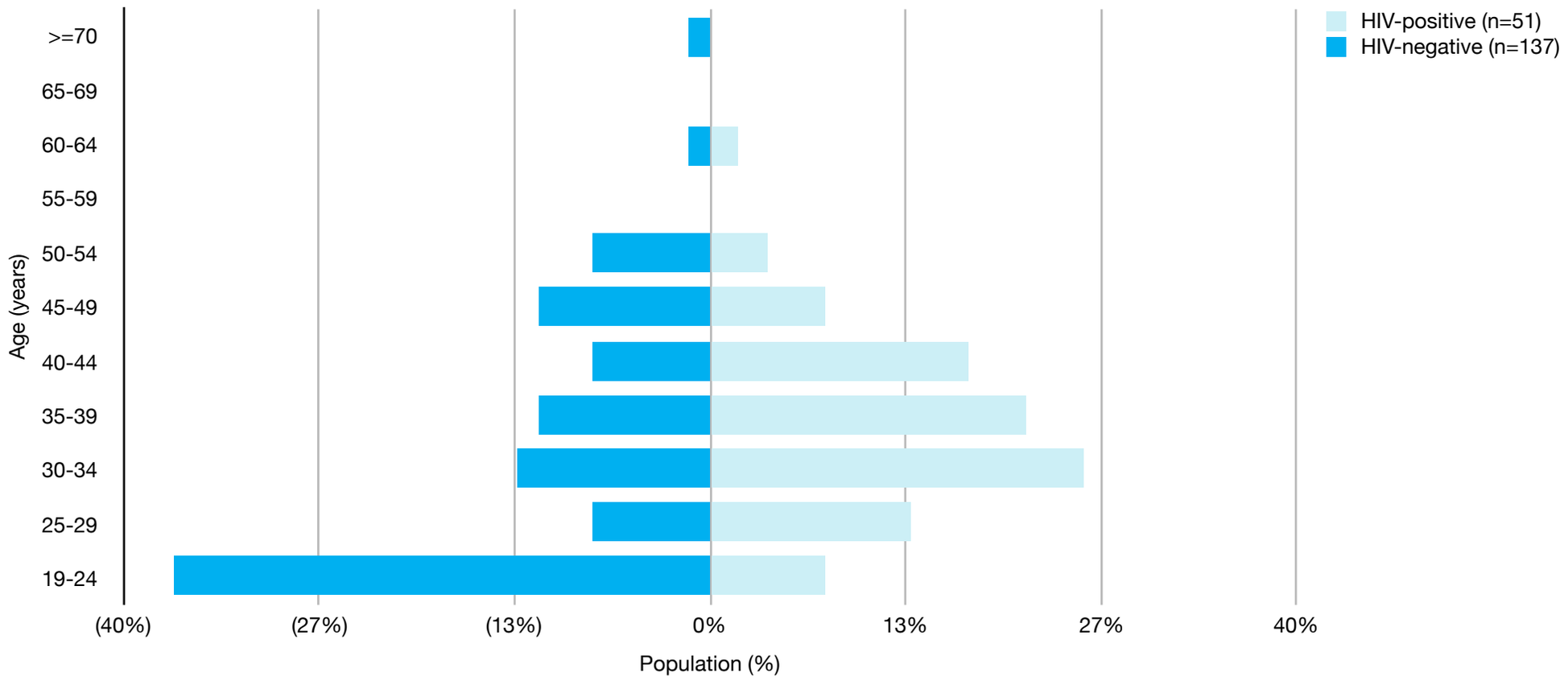
48. Swartz JA, Ducheny K, Holloway T, Stokes L, Willis S, Kuhns LM. A latent class analysis of chronic health conditions among HIV-Positive transgender women of color. AIDS Behav [Internet]. 2019 [cited 2019 Oct 2]; Available from: http://link.springer.com/10.1007/s10461-019-02543-3

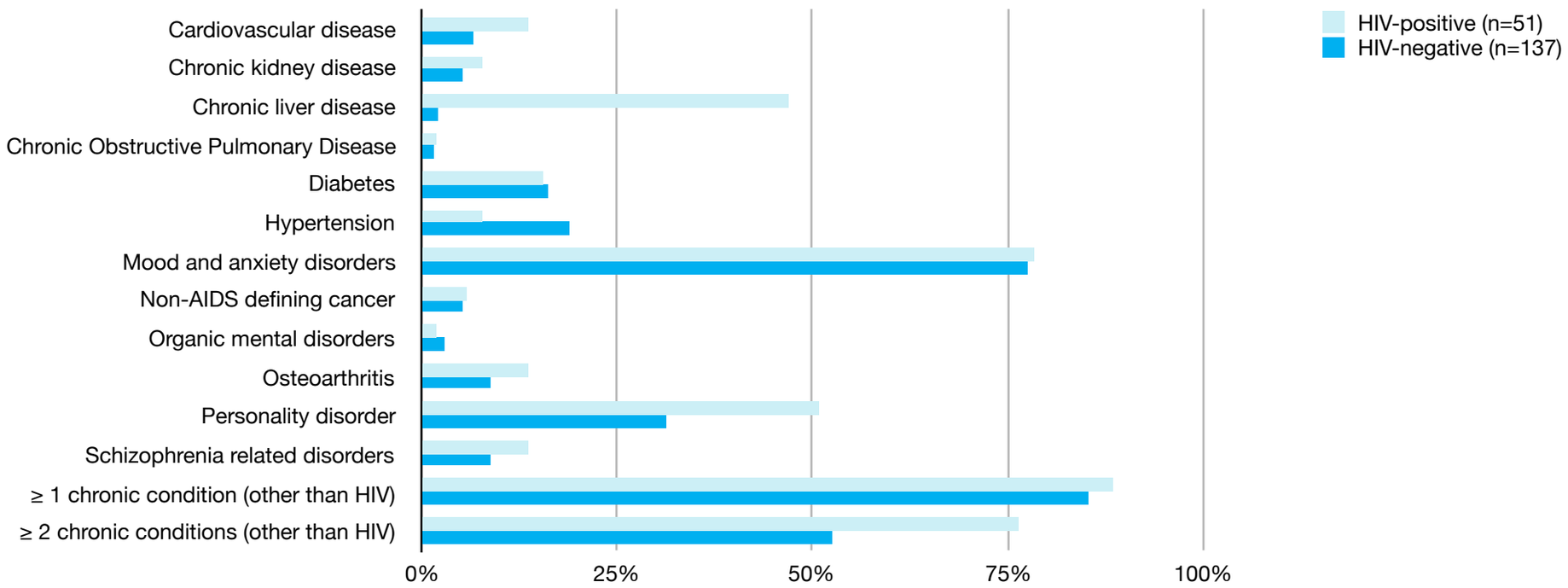
49. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy of Medicare's hospital claims data: progress has been made, but problems remain. Am J Public Health. 2011;

50. Reisner SL, Radix A, Deutsch MB. Integrated and gender-affirming transgender clinical care and research. J Acquir Immune Defic Syndr. 2016;72 Suppl 3:S235–42.








Cardiovascular disease
Chronic kidney disease
Chronic liver disease
Chronic Obstructive Pulmonary Disease
Diabetes
Hypertension
Mood and anxiety disorders
Non-AIDS defining cancer
Organic mental disorders
Osteoarthritis
Personality disorder
Schizophrenia related disorders
≥ 1 chronic condition (other than HIV)
≥ 2 chronic conditions (other than HIV)
0%
25%
50%
75%
100%
HIV-positive (n=51)
HIV-negative (n=137)

Supplementary material

Data sources and description of data elements

Data source	Data steward	Description	Data elements provided for this study
Medical Services Plan	British Columbia Ministry of Health	For individuals covered by the provincial universal health insurance plan- Medically necessary services provided by fee-for-service physicians and other healthcare providers, laboratory services, diagnostic procedures, dental/oral surgery	ICD-9 and ICD-10 codes
Consolidation file	British Columbia Ministry of Health	Demographic data on individuals receiving or registered to receive care in BC, pooled from multiple PopData sources	Sociodemographics
Discharge Abstract Database	British Columbia Ministry of Health	Demographic, administrative and clinical data for inpatient hospital discharges and day surgeries	ICD-9 and ICD-10 codes
Vital statistics deaths	British Columbia Vital Statistics Agency	Records of all registered deaths in BC	Death data
Pharmacare	British Columbia Ministry of Health	Data related to prescription drugs dispensed under the BC public drug insurance program.	Prescription drug name/identifier
Pharmanet	British Columbia Ministry of Health Data Stewardship Committee	Data related to prescription drugs dispensed by community and outpatient pharmacies	Prescription drug name/identifier

Supplementary material

Drug Treatment Program and laboratory	British Columbia Centre for Excellence in HIV/AIDS	Antiretroviral therapy use history, laboratory testing, immunological and virologic testing, and demographic data on PLWH who have accessed antiretrovirals in BC	Providers-reported transgender status, laboratory confirmed HIV serostatus

Prescription drugs with drug identification numbers (DIN)s

	Generic Name	DIN
Transfeminine		
Androgen Blockers		
Spironolactone		
	SPIRONOLACTONE	28606
	SPIRONOLACTONE	613215
	SPIRONOLACTONE	285455
	SPIRONOLACTONE	613223
	SPIRONOLACT/HYDROCHLOROTHIAZID	180408
	SPIRONOLACT/HYDROCHLOROTHIAZID	613231
	SPIRONOLACT/HYDROCHLOROTHIAZID	594377
	SPIRONOLACT/HYDROCHLOROTHIAZID	657182
Cyproterone		
	ETHINYL ESTRADIOL/CYPROTERONE	2233542
	NO GENERIC FORMULARY	634514
	CYPROTERONE ACETATE	704431
	CYPROTERONE ACETATE	2229449
	CYPROTERONE ACETATE	2229723
	CYPROTERONE ACETATE	2232872
	CYPROTERONE ACETATE	2245898
	CYPROTERONE ACETATE	704423
Finasteride		
	FINASTERIDE	2010909
	FINASTERIDE	2238213
Dutasteride		
	DUTASTERIDE	2247813
Estrogens		
Estrogen		
	ESTROGENS,CONJUGATED	830240
	ESTROGENS,CONJUGATED	831395

Supplementary material

ETHYNODIOL D-ETHINYL ESTRADIOL	28630
ETHYNODIOL D-ETHINYL ESTRADIOL	469327
NORETHINDRONE-MESTRANOL	22608
NORETHINDRONE-MESTRANOL	22659
NORETHINDRONE A-E ESTRADIOL	297143
NORETHINDRONE A-E ESTRADIOL	315966
NORETHINDRONE-ETHINYL ESTRAD	317047
NORETHINDRONE-ETHINYL ESTRAD	372846
NORETHINDRONE-ETHINYL ESTRAD	373265
NORETHINDRONE-ETHINYL ESTRAD	531006
NORETHINDRONE-ETHINYL ESTRAD	538590
NORETHINDRONE-ETHINYL ESTRAD	602957
NORETHINDRONE-ETHINYL ESTRAD	620947
NORETHINDRONE-ETHINYL ESTRAD	2187086
NORETHINDRONE-ETHINYL ESTRAD	2187108
NORETHINDRONE-ETHINYL ESTRAD	2189054
NORGESTREL-ETHINYL ESTRADIOL	34207
NORGESTREL-ETHINYL ESTRADIOL	300640
LEVONORGESTREL-ETH ESTRA	579386
LEVONORGESTREL-ETH ESTRA	707600
LEVONORGESTREL-ETH ESTRA	782416
LEVONORGESTREL-ETH ESTRA	782432
LEVONORGESTREL-ETH ESTRA	2042320
NORGESTREL-ETHINYL ESTRADIOL	2043033
LEVONORGESTREL-ETH ESTRA	2043726
NORGESTIMATE-ETHINYL ESTRADIOL	2258560
NORETHINDRONE-MESTRANOL	30333
NORETHINDRONE-MESTRANOL	30341
LEVONORGESTREL-ETH ESTRA	2236974
ETHYNODIOL D-ETHINYL ESTRADIOL	471526
NORETHINDRONE-ETHINYL ESTRAD	340731
NORETHINDRONE-MESTRANOL	340758
NORETHINDRONE A-E ESTRADIOL	343838
NORETHINDRONE A-E ESTRADIOL	353027
NORETHINDRONE-ETHINYL ESTRAD	372838
NORETHINDRONE-ETHINYL ESTRAD	373273
NORETHINDRONE-ETHINYL ESTRAD	531014
NORETHINDRONE-ETHINYL ESTRAD	602965
NORETHINDRONE-ETHINYL ESTRAD	695734
NORETHINDRONE-ETHINYL ESTRAD	2187094

Supplementary material

NORETHINDRONE-ETHINYL ESTRAD	2187116
NORETHINDRONE-ETHINYL ESTRAD	2189062
ETHINYL ESTRADIOL/NORETH AC	2242531
NORGESTREL-ETHINYL ESTRADIOL	340766
NORGESTREL-ETHINYL ESTRADIOL	342815
LEVONORGESTREL-ETH ESTRA	586609
LEVONORGESTREL-ETH ESTRA	707503
LEVONORGESTREL-ETH ESTRA	782424
LEVONORGESTREL-ETH ESTRA	782440
LEVONORGESTREL-ETH ESTRA	2042339
NORGESTREL-ETHINYL ESTRADIOL	2043041
LEVONORGESTREL-ETH ESTRA	2043734
NORGESTIMATE-ETHINYL ESTRADIOL	2258587
LEVONORGESTREL-ETH ESTRA	2236975
NORGESTIMATE-ETHINYL ESTRADIOL	1968440
NORGESTIMATE-ETHINYL ESTRADIOL	2028700
NORGESTIMATE-ETHINYL ESTRADIOL	1992872
NORGESTIMATE-ETHINYL ESTRADIOL	2029421
DESOGESTREL-ETHINYL ESTRADIOL	2042487
DESOGESTREL-ETHINYL ESTRADIOL	2042541
DESOGESTREL-ETHINYL ESTRADIOL	2042479
DESOGESTREL-ETHINYL ESTRADIOL	2042533
ESTRADIOL/NORETH AC	2241835
ESTRADIOL/NORETH AC	2241837
LEVONORGESTREL	2241674
ESTROGEN,CON/M-PROGEST ACET	2242878
ESTROGEN,CON/M-PROGEST ACET	2242879
ESTRADIOL/NORETH AC	2243529
ESTRADIOL/NORETH AC	2243530
ETHINYL ESTRADIOL/DROSPIRENONE	2261723
ETHINYL ESTRADIOL/DROSPIRENONE	2261731
ETONOGESTREL/ETHINYL ESTRADIOL	2253186
ETHINYL ESTRADIOL/NORELGEST	2248297
DIENESTROL	441295
DIETHYLSTILBESTROL	3360
DIETHYLSTILBESTROL	2091461
DIETHYLSTILBESTROL	2091488
ESTRADIOL	464791
ESTRADIOL	2148587
ESTRADIOL	464805

Supplementary material

ESTRADIOL	2148595
ESTRADIOL VALERATE	29238
ESTRADIOL	756849
ESTRADIOL	2237807
ESTRADIOL	2243722
ESTRADIOL	2245676
ESTRADIOL	756857
ESTRADIOL	2204428
ESTRADIOL	2231509
ESTRADIOL	2237808
ESTRADIOL	2243724
ESTRADIOL	2244000
ESTRADIOL	2246967
ESTRADIOL	756792
ESTRADIOL	2204444
ESTRADIOL	2231510
ESTRADIOL	2244002
ESTRADIOL	2246969
ESTRADIOL	2168898
ESTRADIOL	2204436
ESTRADIOL	2244001
ESTRADIOL	2246968
ESTRADIOL	2225190
ESTRADIOL	2204401
ESTRADIOL	2238704
ESTRADIOL	2243999
ESTRADIOL	2241332
ESTRADIOL	2247499
ESTRADIOL	2247500
ESTROGENS,CONJUGATED	2569
ESTROGENS,CONJUGATED	2043394
ESTROGENS,CONJUGATED	2230891
ESTROGENS,CONJUGATED	2239654
ESTROGENS,CONJUGATED	2577
ESTROGENS,CONJUGATED	265470
ESTROGENS,CONJUGATED	587281
ESTROGENS,CONJUGATED	2043408
ESTROGENS,CONJUGATED	2089
ESTROGENS,CONJUGATED	2043440
ESTROGENS,CONJUGATED	403466

Supplementary material

	ESTROGENS,CONJUGATED	2043416
	ESTROGENS,CONJUGATED	2230892
	ESTROGENS,CONJUGATED	2239655
	ESTROGENS,CONJUGATED	2585
	ESTROGENS,CONJUGATED	265489
	ESTROGENS,CONJUGATED	587303
	ESTROGENS,CONJUGATED	2043424
	ESTROGENS,CONJUGATED	2043432
	ESTROGENS,CONJUGATED	2043386
	ME-TESTOSTERONE/ESTROGEN,CON	53538
	ESTROPIPATE	282685
	ESTROPIPATE	2089769
	ESTROPIPATE	282677
	ESTROPIPATE	2089777
	ESTROPIPATE	2089793
	ESTRADIOL/NORETH AC	2108186
	NORGESTIMATE-ETHINYL ESTRADIOL	2229064
	NORGESTIMATE-ETHINYL ESTRADIOL	2229218
	NORGESTIMATE-ETHINYL ESTRADIOL	2229226
	ETHINYL ESTRADIOL/CYPROTERONE	2233542
	ETHINYL ESTRADIOL/NORELGEST	2246340
	NO GENERIC FORMULARY	66124057
	NO GENERIC FORMULARY	66124058
	NO GENERIC FORMULARY	66124060
	NO GENERIC FORMULARY	66124061
	NO GENERIC FORMULARY	66124062
	NO GENERIC FORMULARY	66124063
	NO GENERIC FORMULARY	66124064
Progestogens		
Progesterone		
	PROGESTERONE,MICRONIZED	2241013
	MEDROXYPROGESTERONE ACET	30848
	MEDROXYPROGESTERONE ACET	30856
	MEDROXYPROGESTERONE ACET	585092
	NO GENERIC FORMULARY	66123240
	MEDROXYPROGESTERONE ACET	708917
	MEDROXYPROGESTERONE ACET	2148552
	MEDROXYPROGESTERONE ACET	2221284
	MEDROXYPROGESTERONE ACET	2229838
	MEDROXYPROGESTERONE ACET	2244726

Supplementary material

	MEDROXYPROGESTERONE ACET	2246627
	MEDROXYPROGESTERONE ACET	30937
	MEDROXYPROGESTERONE ACET	2010739
	MEDROXYPROGESTERONE ACET	2148560
	MEDROXYPROGESTERONE ACET	2221292
	MEDROXYPROGESTERONE ACET	2229839
	MEDROXYPROGESTERONE ACET	2244727
	MEDROXYPROGESTERONE ACET	2246628
	MEDROXYPROGESTERONE ACET	729973
	MEDROXYPROGESTERONE ACET	2010933
	MEDROXYPROGESTERONE ACET	2148579
	MEDROXYPROGESTERONE ACET	2221306
	MEDROXYPROGESTERONE ACET	2229840
	MEDROXYPROGESTERONE ACET	2246629
	MEDROXYPROGESTERONE ACET	30945
	MEDROXYPROGESTERONE ACET	2267640
	NORETHINDRONE	37605
	PROGESTERONE,MICRONIZED	2166704
	PROGESTERONE	739952
	PROGESTERONE	1977652
	PROGESTERONE	2128470
	LEVONORGESTREL	2243005
Transmasculine		
Testosterone		
	TESTOSTERONE	2249499
	TESTOSTERONE CYPIONATE	30783
	TESTOSTERONE PROPIONATE	1977571
	TESTOSTERONE CYPIONATE	1977601
	TESTOSTERONE CYPIONATE	2220318
	TESTOSTERONE CYPIONATE	2246063
	TESTOSTERONE ENANTHATE	29246
	TESTOSTERONE ENANTHATE	716936
	TESTOSTERONE ENANTHATE	739944
	TESTOSTERONE UNDECANOATE	782327
	TESTOSTERONE ENANTHATE/ESTRAD	108278
	TESTOSTERONE ENANTHATE/ESTRAD	2061031
	TESTOSTERONE	2239653
	TESTOSTERONE	2245346
	TESTOSTERONE	2245345
	TESTOSTERONE	2245972

Supplementary material

Chronic condition case definitions

Chronic condition		Case definition	Codes
Cardiovascular disease*	Acute myocardial infarction	1 or more hospitalizations with relevant ICD codes	ICD-10: I21 Acute myocardial infarction I22 Subsequent myocardial infarction ICD-9: 410 Acute myocardial infarction
	Ischemic heart disease	At least one of the following: 2 medical visits with Angina ICD-9 code 413 plus 1 heart disease prescription in 365 days; or 1 specialist visit with Angina ICD-9 code 413 plus 1 prescription in 365 days; or 2 medical visits with two ICD9 codes 410, 411, 412, 413, 414 in 365 days; or 1 CCI/CCP CABG,PCI/PCTA procedure code; or 1 hospitalization with relevant ICD code.	ICD-10: I20 Angina pectoris I21 Acute myocardial infarction I 22 Subsequent myocardial infarction I23 Certain current complications following acute myocardial infarction I24 Other acute ischaemic heart diseases I25 Chronic ischaemic heart disease ICD-9: 410 Acute myocardial infarction 411 Other acute and subacute forms of ischaemic heart disease 412 Old myocardial infarction 413 Angina pectoris 414 Other forms of chronic ischaemic heart disease
	Chronic heart failure	1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	ICD-10: I50 Heart failure ICD-9: 428 Heart failure
	Stroke- hospital	1 or more hospitalizations with relevant ICD codes	ICD-10: H34.1 Central retinal artery occlusion I60 Subarachnoid hemorrhage I61 Intracerebral haemorrhage I63 Cerebral infarction (exclude I63.6 Cerebral infarction due to

Supplementary material

			cerebral venous thrombosis, nonpyogenic) I64 Stroke, not specified as haemorrhage or infarction 362.3 Retinal vascular occlusion 430 Subarachnoid hemorrhage 431 Intracerebral hemorrhage 433.x1 Occlusion and stenosis of precerebral arteries 434.x Occlusion cerebral arteries 436 Acute but ill-defined cerebrovascular disease *Excludes any traumatic brain injury*
	Transient ischemic attack	1 or more hospitalizations with relevant ICD codes	ICD-10: H34.0 Transient retinal artery occlusion G45.0 Vertebro-basilar artery syndrome G45.1 Carotid artery syndrome (hemispheric) G45.2 Multiple and bilateral precerebral artery syndromes G45.3 Amaurosis fugax G45.8 Other transient cerebral ischemic attacks and related syndromes G45.9 Transient cerebral ischemic attack, unspecified ICD-9: 435 Transient cerebral ischemia *Excludes any traumatic brain injury*
Chronic kidney disease*		1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	ICD-10: N01 Rapidly progressive nephritic syndrome N03 Chronic nephritic syndrome N04 Nephrotic syndrome N05 Unspecified nephritic syndrome N06 Isolated proteinuria with specified morphological lesion N07 Hereditary nephropathy, not elsewhere classified

Supplementary material

		N18 Chronic kidney disease N19 Unspecified kidney failure N26 Unspecified contracted kidney N27 Small kidney of unknown cause ICD-9: 581 Nephrotic syndrome 582 Chronic glomerulonephritis 583 Nephritis and nephropathy, not specified as acute or chronic 585 Chronic renal failure 586 Renal failure, unspecified 587 Renal sclerosis, unspecified 589 Small kidney of unknown cause
Chronic liver disease	1 or more hospitalization or medical visit with relevant diagnosis within 365 days	ICD-9: 571.0 Alcoholic fatty liver 571.2 Alcoholic cirrhosis of liver 571.3 Alcoholic liver damage, unspecified 571.4 Chronic hepatitis 571.5 Cirrhosis of liver without mention of alcohol 571.6 Billiary cirrhosis 571.8 Other chronic nonalcoholic liver disease 571.9 Unspecified chronic liver disease without mention of alcohol 070.3 Viral hepatitis B without mention of hepatic coma 070.30 Viral hepatitis B without mention of hepatic coma, acute or unspecified, without mention of hepatitis delta 070.31 Viral hepatitis B without mention of hepatic coma, acute or unspecified, with hepatitis delta 070.32 Viral hepatitis B without mention of hepatic coma, chronic, without mention of hepatitis delta 070.33 Viral hepatitis B without mention of hepatic coma, chronic, with hepatitis delta

Supplementary material

		070.52 Hepatitis delta without mention of active Hepatitis B disease or hepatic coma V02.61 Hepatitis B carrier 070.42 Hepatitis delta without mention of active Hepatitis B disease with hepatic coma 070.54 Chronic hepatitis C without mention of hepatic coma V02.62 Hepatitis C carrier
Chronic Obstructive Pulmonary Disease*	1 or more hospitalization or 2 or more medical visits within 365 days	ICD-10: J41 Simple and mucopurulent chronic bronchitis J42 Unspecified chronic bronchitis J43 Emphysema J44 Other chronic obstructive pulmonary disease ICD-9: 491 Chronic bronchitis 492 Emphysema 496 Chronic airways obstruction, not elsewhere classified
Diabetes Mellitus*	At least 1 of the following: 1 hospitalization or 2 medical visits in 365 days with relevant ICD codes; or 2 or more insulin prescriptions in 365 days; or 2 or more oral antihyperglycemic (not including metformin) prescriptions in 365 days; or 1 insulin and 1 oral antihyperglycemic (including metformin) in 365 days; or 2 metformin prescriptions and 1 medical visit in one year with relevant ICD codes.	ICD-10: E10 Type 1 diabetes mellitus E11 Type 2 diabetes mellitus E13 Other specified diabetes mellitus E14 Unspecified diabetes mellitus ICD-9: 250 Diabetes mellitus

Supplementary material

	Excludes gestational diabetes.	
Hypertension*	1 or more hospitalizations or 2 or more medical visits within 2 years with relevant ICD codes. *Excludes gestational hypertension.*	ICD-10: I10 Essential (primary) hypertension I11 Hypertensive heart disease I12 Hypertensive renal disease I13 Hypertensive heart and renal disease I15 Secondary hypertension ICD-9: 401 Essential hypertension 402 Hypertensive heart disease 403 Hypertensive renal disease 404 Hypertensive heart and renal disease 405 Secondary hypertension
Mood and anxiety disorders*	1 or more hospitalizations with a relevant ICD code or 2 or more medical visits with a relevant code within 2 years	ICD-10: F30 Manic episode F31 Bipolar affective disorder F32 Depressive episode F33 Recurrent depressive disorder F34 Persistent mood [affective] disorders F38 Other mood [affective] disorders F39 Unspecified mood [affective] disorder F40 Phobic anxiety disorders F41 Other anxiety disorders F42 Obsessive-compulsive disorder F43 Reaction to severe stress, and adjustment disorders F44 Dissociative (conversion) disorders F45 Somatoform disorders F48 Other neurotic disorders F68 Other disorders of adult personality & behavior ICD-9:

Supplementary material

			296 Affective psychoses 300 Neurotic disorders 311 Depressive disorder, not elsewhere classified MSP DX Code: 50B Anxiety/Depression
Non-AIDS defining cancer†	All prevalent cancer cases were included, with the exception of AIDS defining malignancies (Kaposi's sarcoma, non-Hodgkin's lymphoma, invasive cervical cancer)		*Cancer case definition details available from the British Columbia Cancer Agency: http://www.bccancer.bc.ca/health-info/types-of-cancer*
Organic mental disorders		1 or more medical visits or hospitalizations with relevant diagnoses within 365 days	ICD-9: 290.x Dementias 294.x Other organic psychotic conditions 331.x Alzheimer's ICD-10: F00.x Dementia in Alzheimer's disease F01.x Vascular Dementia F02.x Dementia in other disease classified elsewhere F03.x Unspecified dementia F04 Amnestic disorder due to physiological condition F06 Other mental disorders due to known physiological condition F09 Unspecified mental disorder due to known physiological condition G30 Alzheimer's disease with early onset
Osteoarthritis*		1 or more hospitalization or 2 or more medical visits in 365 days with a relevant ICD code	ICD-10: M15 Polyarthrosis M16 Coxarthrosis [arthrosis of hip] M17 Gonarthrosis [arthrosis of knee] M18 Arthrosis of first carpometacarpal joint M19 Other arthrosis

Supplementary material

		ICD-9: 715 Osteoarthrosis and allied disorders
Personality disorder	1 or more hospitalizations or medical visits with a relevant diagnosis within 365 days	ICD-9: 301.x Personality disorders ICD-10: F60.x Specified personality disorders F62 Enduring personality changes, not attributable to brain damage and disease F68.1 Intentional production or feigning of symptoms or disabilities, either physical or psychological F68.8 Other specified disorders or adult personality and behaviour F69 Unspecified disorder or adult personality and behaviour
Schizophrenia related disorder	1 or more medical visit or hospitalizations with relevant diagnoses within 365 days	ICD-9: 295.x Schizophrenic disorders 297.0 Paranoid state, simple 297.1 Delusional disorder 297.2 Paraphrenia 297.3 Shared psychotic disorder ICD-10: F20.x Paranoid schizophrenia F21.x Schizotypal disorder F23.2 Acute schizophrenia-like psychotic disorder F25.x Schizoaffective disorders

* Case definition adapted from British Columbia Ministry of Health version 2017, April 4 2019 update

† Case-definition adapted from British Columbia Cancer Agency

The RECORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational studies using routinely collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location in manuscript where items are reported
Title and abstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found		RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	1.1-3: Title page
Introduction					
Background rationale	2	Explain the scientific background and rationale for the investigation being reported			Introduction (pp 4-5)
Objectives	3	State specific objectives, including any prespecified hypotheses			Introduction (page 5)
Methods					
Study Design	4	Present key elements of study design early in the paper			Methods (page 5); Supplementary Material
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection			Methods (page 5)

Participants	6	*(a) Cohort study* - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		RECORD 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	6.1: Methods (pp 6-7); Supplementary Material 6.2: Methods (pp 6-7) 6.3: Figure 1
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	7.1: Methods (pp 6-7); Supplementary Material
Data sources/ measurement	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group			Methods (pp 6-7)

Bias	9	Describe any efforts to address potential sources of bias			Discussion (pp 9, 11)
Study size	10	Explain how the study size was arrived at			Methods (page 6); Figure 1
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Methods (pp 6-7)
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses			Methods (pp 6-7)
Data access and cleaning methods		..		RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	12.1-2: Methods

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	
Linkage		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Methods (page 5)
Results					
Participants	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.*, study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	Results (page 8); Figure 1
Descriptive data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			Results (page 8)
Outcome data	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report numbers in each exposure			Results

		category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			NA
Other analyses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			NA
Discussion					
Key results	18	Summarise key results with reference to study objectives			Discussion: Page 9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	Discussion: Page 9, Page 11
Interpretation	20	Give a cautious overall interpretation of results considering objectives,			Conclusion (Page 11)

		limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
Generalisability	21	Discuss the generalisability (external validity) of the study results			Limitations (Page 9)
Other Information					
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			Funding (Page 12)
Accessibility of protocol, raw data, and programming code		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	Page 12

*Reference: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD Working Committee. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine* 2015; in press.