

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts



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 T³, Koehoorn M¹, Li J², Ye M², Sereda, P², Salway T⁴, 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

1
2
3 healthcare setting, using a computable phenotype, and capacity to estimate the prevalence
4 of transgender status among the population living with HIV in the province.

- 5 • Development of a validated transgender computable phenotype algorithm lays the
6 foundation for future investigation of transgender-specific research questions related to
7 general and HIV-specific healthcare use and health outcomes for this key population.
8
9

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

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

1
2
3 data element). Development of CPs using standardized methods and definitions enables
4 identification and inclusion of transgender persons in research, as well as replication of analyses
5 across data sources, healthcare organizations/sites and studies. CPs have application in clinical
6 care, surveillance, and health research.
7

8
9 Recently, CP and other EHR-based algorithm methods have been applied in a number of settings
10 primarily in the US to identify transgender samples for health research.[14] Specifically, the
11 STRONG study identified a transgender cohort (n = 6,456) using EHR data from Kaiser
12 Permanente health plan members in California and Georgia, for investigation of general and
13 transgender-specific health outcomes.[18] Blosnich et al identified 3,177 people with a “gender
14 identity disorder” diagnosis among military veterans accessing care through the US Veterans
15 Health Administration healthcare system,[19] for examination of mental health and other
16 outcomes. Researchers with the US Centers for Medicare & Medicaid Services identified 4,098
17 transgender beneficiaries using national Medicare claims data,[20] and researchers at Vanderbilt
18 University identified 234 transgender patients in their university clinic EHR data.[16] While
19 these cohorts represent important opportunities for advancement of transgender health research,
20 these methods have yet to be applied widely outside the US context. This is particularly
21 important as different jurisdictions may vary in medical billing and coding practices, healthcare
22 system patient populations, and representativeness of the general population. Specifically, in
23 Canada, healthcare is delivered through a provincially administered universal healthcare system.
24 As such, research using EHR provides an opportunity to develop methods for population-based,
25 representative estimates of transgender populations within the Canadian context. Coupled with
26 the current absence of gender ascertainment measures in population-based routinely collected
27 data (e.g., census, national government health surveys, etc.) in Canada and many other
28 jurisdictions, this remains an evidence need.
29
30
31
32

33 **Summary of study rationale**

34 This study investigated the application of emerging transgender health research methods,
35 specifically CPs, in a Canadian context for the first time, testing the feasibility of identification
36 of a transgender sample using EHR data from a provincial healthcare administrative data-linked
37 cohort.
38

39 **METHODS**

40 **Data Sources and Participants**

41 The Comparative Outcomes and Service Utilization Trends Study (COAST)
42 COAST is a population-based cohort study focused on health services utilization research
43 questions among all people known to be living with HIV (PLWH) in the province of British
44 Columbia (BC) and a 10% random sample comparison group of the HIV-negative general
45 population.[21] The COAST cohort comprises individual-level, longitudinal data from PLWH
46 who have ever accessed HIV treatment in BC between 1996 and 2013, provided by Population
47 Data BC (PopDataBC)[21] via data linkage between two provincial data sources, by personal
48 health number: the Drug Treatment Program (DTP) [22] and the Ministry of Health. PopDataBC
49 provides infrastructure for access to, and linkage of, longitudinal and individual-level
50 administrative health data for all BC residents.[23]. The HIV-negative general population cohort
51 was drawn randomly from the Ministry of Health registry data by PopDataBC. The COAST study
52 has received approval from the University of British Columbia/Providence Health Care
53
54
55
56
57
58
59

1
2
3 Research Ethics Board (#H09-02905) and Simon Fraser University Office of Research Ethics
4 (#2013 s0566). The study complies with the BC Freedom of Information and Protection of
5 Privacy Act (FIPPA) and did not require informed consent as it is conducted using
6 retrospective administrative and anonymized data for research and statistical purposes
7 only.
8
9

10 Drug Treatment Program

11 In BC, antiretroviral therapy (ART) is provided to PLWH at no cost to the patient, and distributed
12 through the DTP.[22] The DTP contributed a provider-reported measure of transgender status for
13 COAST.
14
15

16 Ministry of Health

17 Ministry of Health data available via COAST included insured medical service billing records for
18 outpatient visits,[24,25] hospital (in-patient) visits,[26] prescription medications,[27,28] and vital
19 statistics.[29]
20
21

22 **Measures & Analyses**

23 Transgender computable phenotypes

24 Identification of transgender cases was tested in COAST using International Classification of
25 Disease (ICD) codes (9th and 10th editions) and exogenous sex hormone prescription use.
26 Transgender-specific diagnoses in medical and hospital billing records included the ICD-9 codes
27 302.5 Trans-sexualism with unspecified history, 302.51 Trans-sexualism with asexual history,
28 302.52 Trans-sexualism with homosexual history, 302.53 Trans-sexualism with heterosexual
29 history, 302.6 Gender Identity Disorder in children, 302.85 Gender Identity Disorder in
30 adolescents or adults; and ICD-10 codes F64.0 Gender Identity Disorder of childhood, F64.2
31 Gender Identity Disorder of childhood, F64.8 Other Gender Identity Disorder, and F64.9 Gender
32 Identity Disorder unspecified. The full list of androgen blockers and exogenous sex hormone
33 prescriptions included in analyses is available in the supplementary material.
34
35
36

37 Concordance

38 To assess face validity and utility of diagnosis and prescription data over time in CP
39 development, concordance analyses evaluated the presence of at least one included diagnosis and
40 prescription during the COAST study follow-up period with the presence of at least one included
41 diagnosis and prescription in the last study year. Concordance was assessed between
42 transgender-specific diagnoses, exogenous sex hormone and androgen blocker prescriptions, and
43 non-transgender specific diagnoses (ICD-9 259.9 Unspecified Endocrine Disorder and ICD-10
44 E34.9 Endocrine Disorder, Unspecified (see supplementary material)). Endocrine disorder
45 diagnosis codes are sometimes preferred by medical providers treating transgender people in
46 response to historic exclusions of transgender-specific care from insurance coverage and to
47 combat the stigma of transgender-specific diagnosis codes that have historically been classified
48 as psychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM).[30]
49 Exogenous sex hormone use, while common in transgender populations,[5,31] is not
50 transgender-specific. Cisgender populations also use androgen blocker and sex hormone
51 prescriptions (e.g. estrogen to treat menopausal symptoms in cisgender women, spironolactone is
52 used for hypertension), thus exogenous sex hormone and androgen blocker prescription use
53
54
55
56
57
58
59

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 geographically 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

1
2
3 disease, chronic kidney disease, osteoarthritis, schizophrenia and personality disorders, and
4 chronic liver disease, but a lower prevalence for hypertension.
5

6
7 *[Figure 4 here]*
8

9 **DISCUSSION**

10 This study demonstrates the feasibility of identification of a sample of transgender people in a
11 large linked provincial healthcare administrative database, using a CP based on prescriptions and
12 diagnoses. Among a growing number of studies using EHR and CP methods to identify
13 transgender samples for health research purposes, this is the first to do so in Canada., to
14 independently validate the CP using a ‘gold standard’ of provider-reported transgender status,
15 and the only to use population-based data.
16

17 **Concordance**

18 There was high concordance between transgender-specific diagnoses and exogenous sex
19 hormone or androgen blocker prescription use in this study. That nearly half of those with at
20 least one transgender-specific diagnosis had been dispensed hormones or blockers in the past
21 year is consistent with findings from US and Canadian studies (48.9% and 43.0%,
22 respectively)[20,32,33] - suggesting face validity for the current CP.
23
24
25

26 **CP development and validation**

27 The best-performing CP overall successfully identified cisgender people who were truly
28 cisgender (specificity) and correctly identified transgender people who were truly transgender
29 (0.2% false positive rate, results not shown). However, the selected CP had relatively low
30 sensitivity, missing approximately 72.5% of ‘true’ transgender people in COAST. Though a
31 relatively small proportion of the ‘true’ transgender sample was identified in this study, the
32 impact on future analyses comparing health outcomes for transgender and cisgender groups is
33 likely negligible, as even the large proportion of ‘true’ transgender people misclassified as
34 cisgender (approximate n=496) is a very small proportion of the total COAST sample. At worst,
35 this misclassification would bias results related to disparities between transgender and cisgender
36 health toward the null, producing a conservative attenuated effect. Further, as discussed below,
37 gender identity classification will likely greatly improve as transgender care shifts further into
38 the fee-for-service system in BC. As in other Canadian administrative data studies, low
39 sensitivity may be explained in part by provider and system billing preferences using 3-digit ICD
40 diagnosis coding instead of the more specific 4-digit coding, and inconsistencies in the BC
41 billing management system.[37]
42
43
44

45 The limited agreement between the CP and provider-report transgender status may be due to the
46 widely varying transgender status prevalence depending on study design and ascertainment
47 measures used.[14] In the BC context, the CP and the DTP measures are assessing transgender
48 status in different ways and for different purposes. In the DTP, transgender status is ascertained
49 in the context of HIV diagnosis and ART prescribing, during which demographics and HIV
50 transmission risk factors are recorded. This differs from recording diagnoses in EHR for those
51 accessing transgender-specific care as utilized in the CP. Ultimately, a single CP may not be
52 sufficient for all intended purposes and the best applicable CP (using different types of
53
54
55
56
57
58
59
60

1
2
3 diagnoses, prescriptions or procedures) may differ depending on the intended healthcare, health
4 research, or health policy application.[17]
5

6
7 There is limited literature on EHR-based studies with the ability to validate an administrative
8 transgender measure using a ‘gold standard’ comparison measure.[16] The two previous studies
9 that have developed and validated algorithms to identify transgender individuals have both been
10 conducted in non-representative samples in the US, one using Medicare data[38] and one in a
11 university medical center.[16] Similar to the current study, the Medicare study found high
12 specificity when comparing an EHR-based and a two-step survey-based transgender measure.
13 However, the Medicare study found that the EHR measure performed consistently well with high
14 sensitivity and a high Kappa statistic, unlike in the current study. Using chart review as the ‘gold
15 standard’ for comparison of transgender status, Ehrenfeld et al. found a low false positive rate for
16 their best-performing algorithm (3%), though not as low as the false positive rate in the current
17 study. The overall high levels of agreement for transgender measures in the two previous studies
18 is likely a function of the lack of independence between the ‘gold standard’ and the CP or
19 algorithm measures. Specifically, only those classified as transgender in the Medicare EHR data
20 were offered survey participation to complete the two-step ‘gold standard’ survey measure, and
21 only those cases identified as transgender in the university clinic EHR were included in chart
22 review. Thus, previous studies could assess agreement between the two measures, but not
23 robustly validate either. In the current study, the DTP provider-based transgender status measure
24 is independent and thus could be used for robust CP validation.
25
26
27

28 **Transgender status prevalence & ascertainment**

29
30 Based on a recent meta-analysis of transgender status prevalence in population-based probability
31 samples,[10] it was expected that an effective CP would identify 0.4% of the general population
32 as transgender, or approximately 54 of the HIV-positive COAST cohort (n=13 907) and 3,098 of
33 the HIV-negative cohort (n=516 340). Consistent with expectations, the best-performing CP
34 identified 51 PLWH as transgender, equivalent to a transgender status prevalence of 0.4% among
35 PLWH. Contrary to expectations, the best-performing CP identified less than 5% of the number
36 of transgender persons expected in the HIV-negative cohort. This is likely a result of a number of
37 factors including the limitation of CPs to the subset of a population accessing care as noted, and
38 the result of most transgender people in BC receiving care currently outside the main fee-for-
39 service healthcare delivery system. However, it is also consistent with the undercount of
40 transgender populations using diagnostic criteria compared to other methods of ascertainment
41 demonstrated in other studies.[14]
42
43

44
45 Using the broadest CP algorithm (any transgender-specific diagnosis ever, n=56) and those
46 identified by provider-report together (total n=106), the total transgender PLWH sample would
47 represent as high as 0.88% (range: 0.73-1.1%) of the prevalent HIV infections in BC in
48 2014.[39] This overrepresentation of transgender people among PLWH is consistent with
49 evidence of a disproportionate HIV burden for transgender populations globally,[5,40,41] as well
50 as in line with the only other available data on the proportion of PLWH who are transgender,
51 from US national surveillance data (2012 data: 1.1%, 95%CI: 0.8-1.4).[42]
52
53

54 **Demographics and chronic conditions**

55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 collection and legal changes in insurance coverage of transgender healthcare in the US potentiate
4 decline in accurate claims coding for gender-affirming care.[30] This highlights the utility of
5 work in this area from other jurisdictions, particularly those with transgender-inclusive universal
6 healthcare systems such as Canada.
7

8
9 Future research should build upon the methods developed in this study and explore
10 complimentary approaches for gender identity ascertainment in administrative and EHR data,
11 such as machine learning approaches, as have been used to develop algorithms based on
12 healthcare utilization data in other research areas. Finally, the current study lays the foundation
13 for future work with the ability to study transgender health and healthcare use patterns over time,
14 with linkage to laboratory data, as well as inclusion of appropriate comparison groups.[15,49]
15

16 17 **ACKNOWLEDGEMENTS & DISCLAIMER**

18 The authors thank the COAST study participants, the BC Centre for Excellence in HIV/AIDS,
19 the BC Ministry of Health, BC Vital Statistics Agency, PharmaNet and the institutional data
20 stewards for granting access to the data, and Population Data BC for facilitating the data linkage
21 process. In addition, we would like to thank the COAST core team members and other support
22 staff at these institutions for their administrative assistance with the data access and preparation.
23 All inferences, opinions, and conclusions drawn in this paper are those of the authors, and do not
24 reflect the opinions or policies of the Data Steward(s). There are no conflicts of interest to
25 declare.
26
27

28 29 **COMPETING INTERESTS**

30 None declared.
31

32 33 **FUNDING**

34 This work was supported by the Canadian Institutes of Health Research, through an Operation
35 Grant [grant number 130419], a Foundation Award to RSH [grant number 143342], and a
36 Doctoral Research Award to AJR [grant number 152382] and support from the British Columbia
37 Centre for Excellence in HIV/AIDS. The DTP receives funding from the provincial government
38 of British Columbia (PharmaCare). The funders had no role in the study design, analysis,
39 interpretation of the data, drafting of the manuscript or in the decision to submit for
40 publication.
41
42

43 44 **CONTRIBUTORS**

45 AJR and RSH conceived of and designed the study. RSH acquired study data and funding. JL, MY
46 and PS conducted data analysis, in consultation with AJR. AJR drafted the manuscript, with
47 commentary on drafts by TP, MK, and TS, and approval of final version by all co-authors.
48

49 50 **PATIENT CONSENT FOR PUBLICATION**

51 Not required.
52

53 54 **DATA SHARING STATEMENT** 55 56 57 58 59 60

1
2
3 The data used for this study are held by the BC Centre for Excellence in HIV/AIDS under
4 the authority of the BC Ministry of Health; as they contain confidential patient health
5 records including HIV serostatus, data are cannot be made available to other parties.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3 **FIGURES LEGENDS**
4

5 **Figure 1. Total transgender sample identified using a computable phenotype with**
6 **electronic health records**
7

8
9 **Figure 2. Geographic distribution of transgender people across province, by health**
10 **authority***
11

12 **% of transgender individuals with known health authority (n=182)*
13
14

15 **Figure 3. Age distribution of transgender sample, by HIV serostatus**
16
17

18 **Figure 4. Co-morbidities among transgender sample, by HIV serostatus**
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

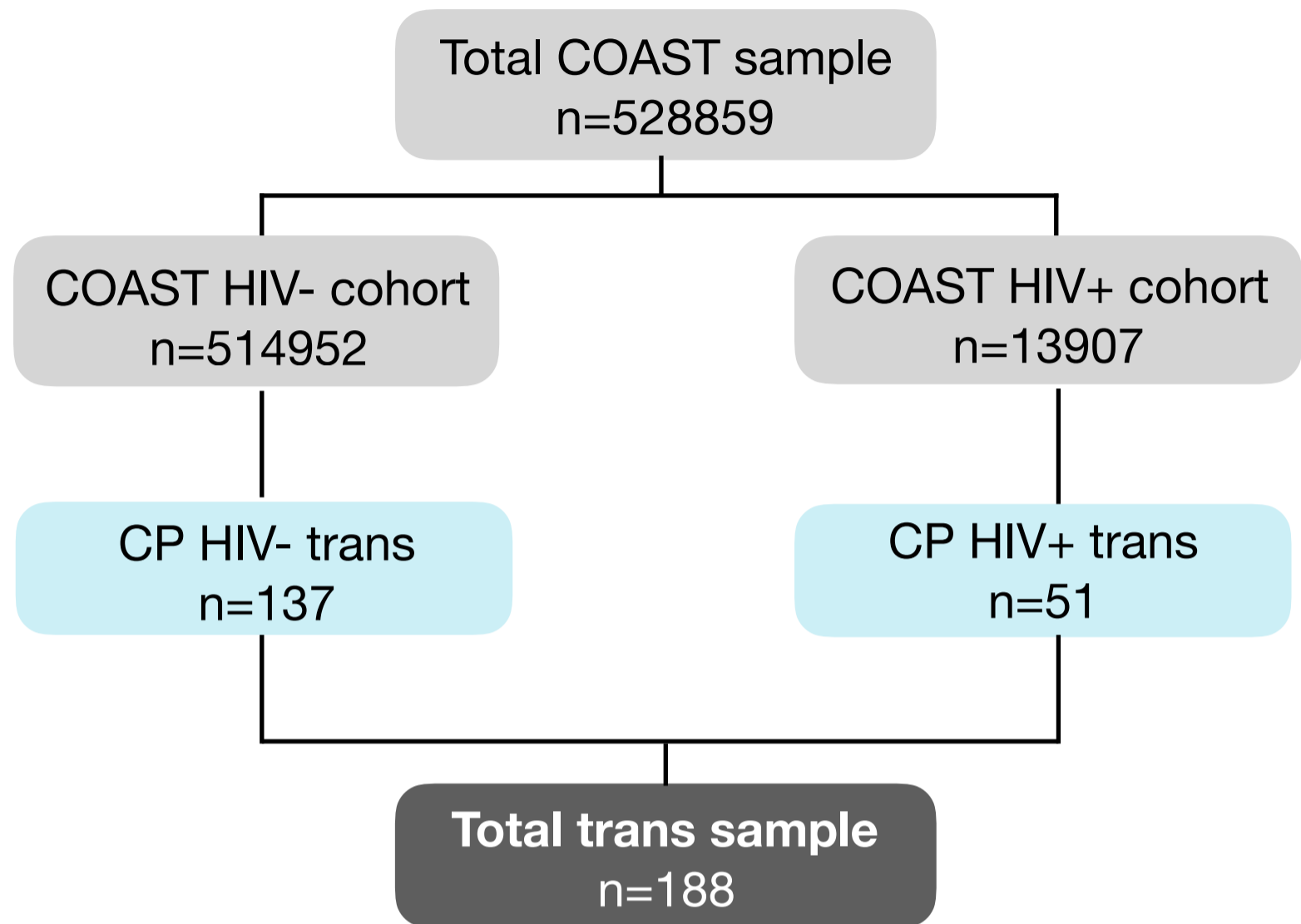
REFERENCES

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.

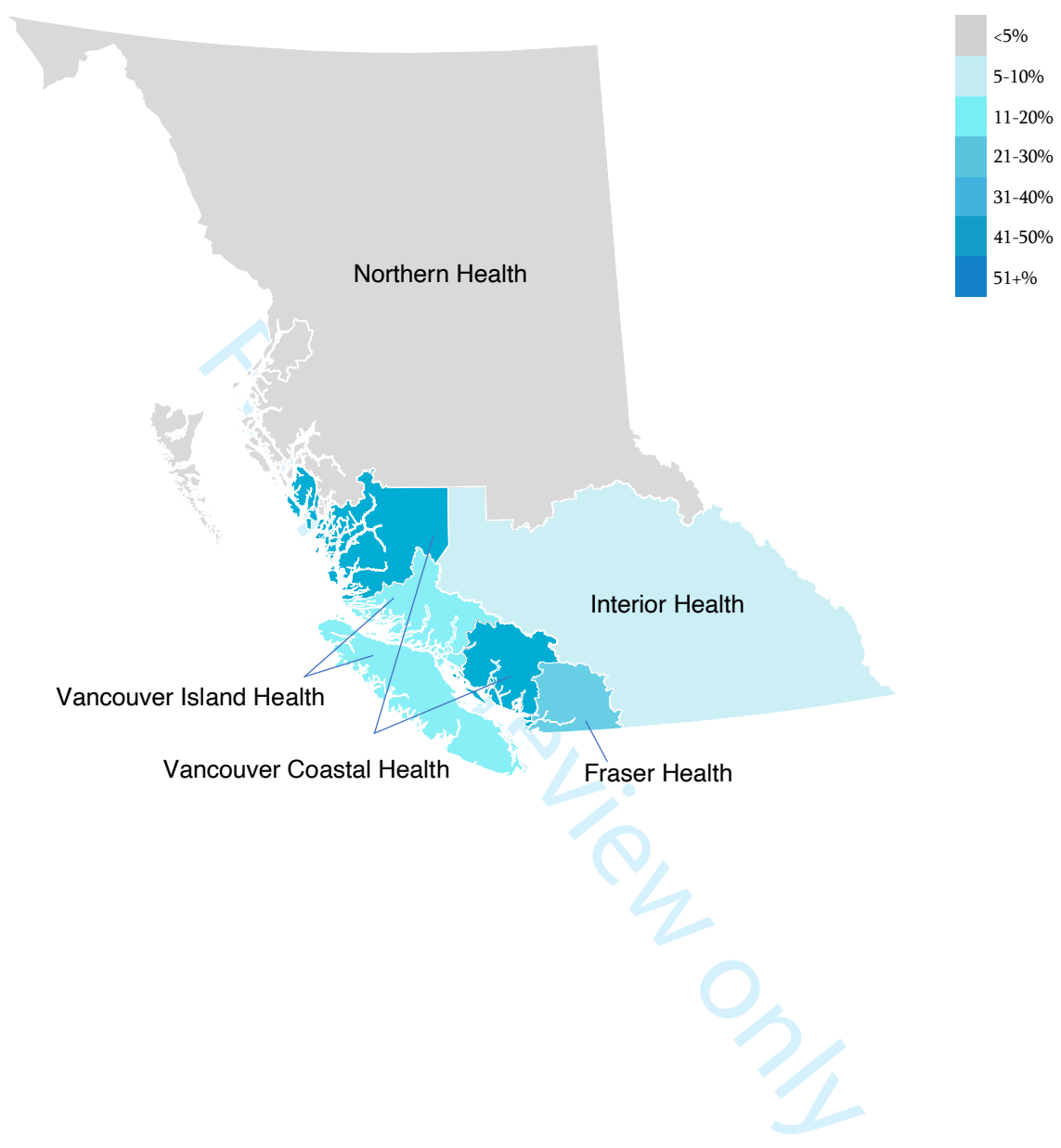
13. Bauer GR, Braimoh J, Scheim AI, Dharma C. Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. *PLOS ONE*. 2017;12:e0178043.
14. Collin L, Reisner SL, Tangpricha V, Goodman M. Prevalence of transgender depends on the “case” definition: 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 Textbook 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>

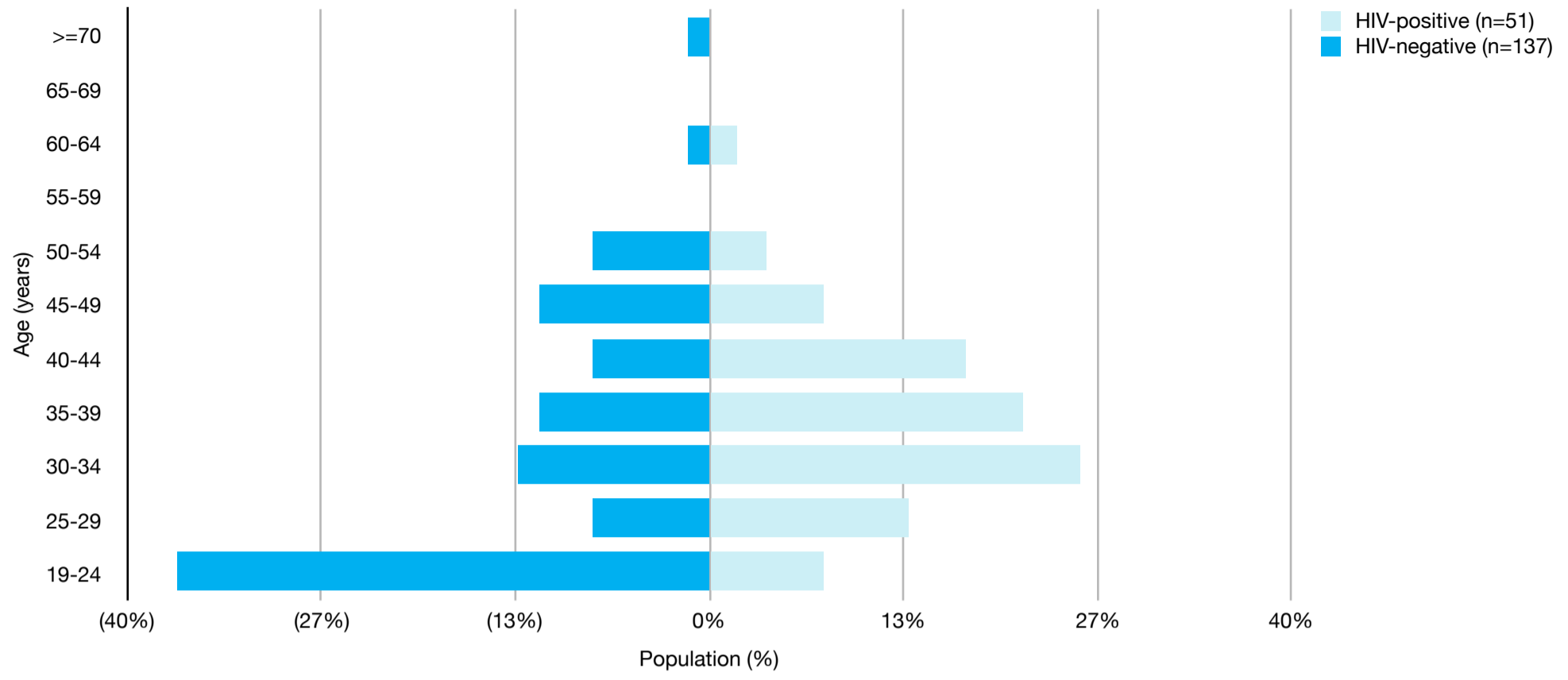
- 1
2
3 26. Canadian Institute for Health Information (2014). Discharge Abstract Database (Hospital
4 Separations). Population Data BC. Data Extract. MOH (2014). <http://www.popdata.bc.ca/data>
5
- 6
7 27. British Columbia Ministry of Health (2014). PharmaCare. Population Data BC. Data Extract.
8 MOH (2014). <http://www.popdata.bc.ca/data>
9
- 10 28. British Columbia Ministry of Health (2014). PharmaNet. Population Data BC. Data Extract.
11 Data Stewardship Committee (2014). <http://www.popdata.bc.ca/data>
12
- 13 29. British Columbia Vital Statistics Agency (2014). Vital Statistics Deaths. Population Data BC.
14 Data Extract. Vital Statistics Agency (2014). <http://www.popdata.bc.ca>
15
- 16 30. Arnold JD, Nelson AE, Loubier EM. Trends in Insurance Coverage for Gender-Affirming
17 Surgeries. *JAMA Surg.* 2018;153:972.
18
- 19 31. Radix A, Sevelius J, Deutsch MB. Transgender women, hormonal therapy and HIV
20 treatment: a comprehensive review of the literature and recommendations for best practices. *J Int*
21 *AIDS Soc.* 2016;19:20810.
22
- 23 32. Rotondi NK, Bauer GR, Scanlon K, Kaay M, Travers R, Travers A. Nonprescribed hormone
24 use and self-performed surgeries: “do-it-yourself” transitions in transgender communities in
25 Ontario, Canada. *Am J Public Health.* 2013;103:1830–1836.
26
- 27 33. Scheim AI, Bauer GR. Sex and gender diversity among transgender persons in Ontario,
28 Canada: results from a respondent-driven sampling survey. *J Sex Res.* 2015;52:1–14.
29
- 30 34. Chronic Disease Information Working Group. BC Chronic Disease and Selected Procedure
31 Case Definitions [Internet]. British Columbia, Canada: BC Ministry of Health; 2019 Apr p. 1–
32 132. Available from: [https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-
33 access/chronic-disease-registries-case-definitions.pdf](https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf)
34
35
- 36 35. British Columbia Cancer Agency (2014). BC Cancer Registry Data. Population Data BC. BC
37 Cancer (2014). <http://www.popdata.bc.ca/data>
38
- 39 36. BC Stats. Sub-provincial Population Estimates [Internet]. 2018 [cited 2019 Feb 1]. Available
40 from: <file://localhost/Users/AJRich/Zotero/storage/5VM2LBUN/PopulationEstimates.html>
41
- 42 37. Hu W. Diagnostic Codes in MSP Claim Data. Program Monitoring and Information
43 Management Branch, Resource Management Division, Medical Services Plan; 1996 p. 39.
44
- 45 38. Gerth J, Becerra-Culqui T, Bradlyn A, Getahun D, Hunkeler EM, Lash TL, et al. Agreement
46 between medical records and self-reports: Implications for transgender health research. *Rev*
47 *Endocr Metab Disord.* 2018;19:263–269.
48
- 49 39. BC Centre for Disease Control. HIV in British Columbia: Annual Surveillance Report 2015.
50 Vancouver, Canada: BC Centre for Disease Control; 2017.
51
52
53
54
55
56
57
58
59

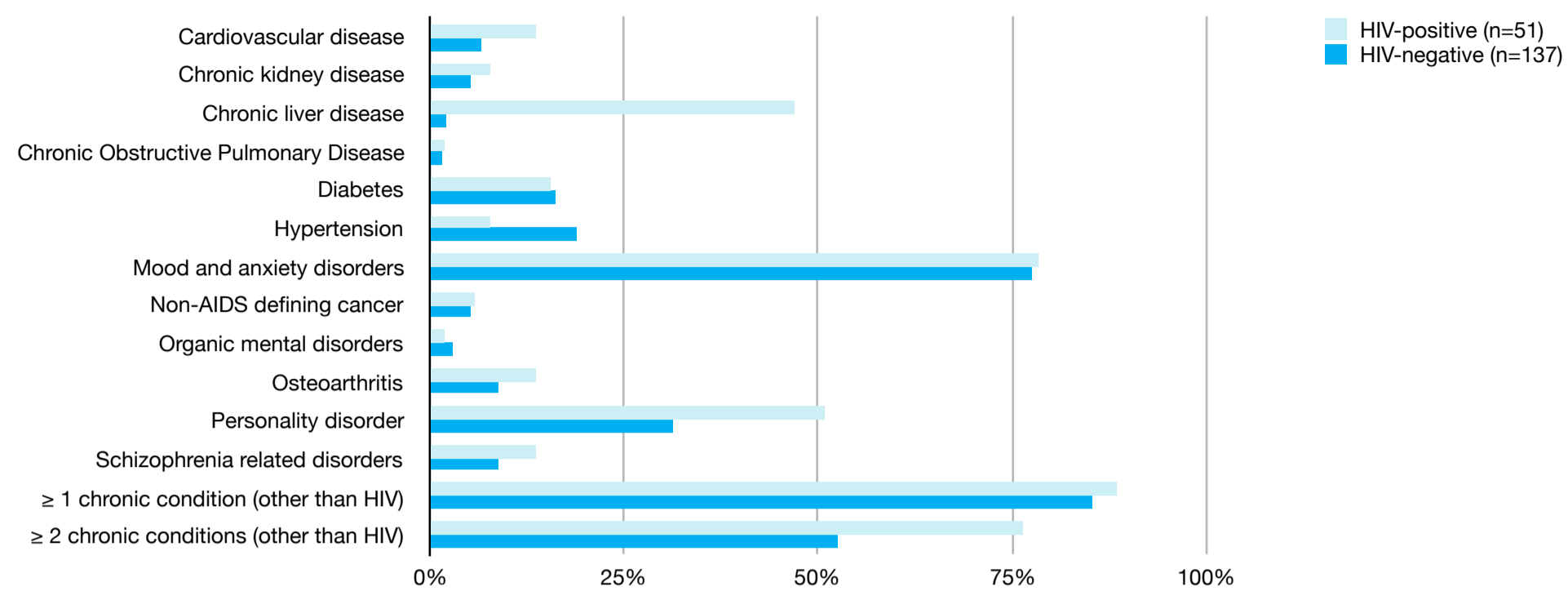
- 1
2
3 40. Poteat T, Scheim A, Xavier J, Reisner S, Baral S. Global epidemiology of HIV infection and
4 related syndemics affecting transgender people. *J Acquir Immune Defic Syndr*. 2016;72 Suppl
5 3:S210–9.
6
7
8 41. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of
9 HIV in transgender women: a systematic review and meta-analysis. *Lancet Infect Dis*.
10 2013;13:214–222.
11
12 42. Centers for Disease Control and Prevention. Behavioral and Clinical Characteristics of
13 Persons Receiving Medical Care for HIV Infection—Medical Monitoring Project, United States,
14 2012 [Internet]. Washington, D.C.: Centers for Disease Control; 2015 Aug p. 42. Report No.: 12.
15 Available from: [https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-](https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf)
16 [HSSR_MMP_2012.pdf](https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf)
17
18
19 43. Grabovac I, Veronese N, Stefanac S, Haider S, Jackson SE, Koyanagi A, et al. Human
20 Immunodeficiency Virus Infection and Diverse Physical Health Outcomes: An Umbrella Review
21 of Meta-analyses of Observational Studies. *Clin Infect Dis*. 2019;ciz539.
22
23 44. Nasi M, De Biasi S, Gibellini L, Bianchini E, Pecorini S, Bacca V, et al. Ageing and
24 inflammation in patients with HIV infection: Ageing, inflammation and HIV infection. *Clin Exp*
25 *Immunol*. 2017;187:44–52.
26
27
28 45. Judge C., O’Donovan C., Callaghan G., Gaoatswe G., O’Shea D. Gender dysphoria -
29 prevalence and co-morbidities in an Irish adult population. *Front Endocrinol*. 2014;5:87.
30
31 46. Cheung AS, Ooi O, Leemaqz S, Cundill P, Silberstein N, Bretherton I, et al.
32 Sociodemographic and clinical characteristics of transgender adults in Australia. *Transgender*
33 *Health*. 2018;3:229–38.
34
35 47. Swartz JA, Ducheny K, Holloway T, Stokes L, Willis S, Kuhns LM. A Latent Class Analysis
36 of Chronic Health Conditions Among HIV-Positive Transgender Women of Color. *AIDS Behav*
37 [Internet]. 2019 [cited 2019 Oct 2]; Available from: [http://link.springer.com/10.1007/s10461-](http://link.springer.com/10.1007/s10461-019-02543-3)
38 [019-02543-3](http://link.springer.com/10.1007/s10461-019-02543-3)
39
40
41 48. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy
42 of Medicare’s hospital claims data: progress has been made, but problems remain. *Am J Public*
43 *Health*. 2011;
44
45 49. Reisner SL, Radix A, Deutsch MB. Integrated and Gender-Affirming Transgender Clinical
46 Care and Research. *J Acquir Immune Defic Syndr* 1999. 2016;72 Suppl 3:S235–42.
47
48
49
50
51
52
53
54
55
56
57
58
59
60



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60







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		
Spirolactone		
	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

1		
2		
3	ETHYNODIOL D-ETHINYL ESTRADIOL	28630
4	ETHYNODIOL D-ETHINYL ESTRADIOL	469327
5	NORETHINDRONE-MESTRANOL	22608
6	NORETHINDRONE-MESTRANOL	22659
7	NORETHINDRONE A-E ESTRADIOL	297143
8	NORETHINDRONE A-E ESTRADIOL	315966
9	NORETHINDRONE-ETHINYL ESTRAD	317047
10	NORETHINDRONE-ETHINYL ESTRAD	372846
11	NORETHINDRONE-ETHINYL ESTRAD	373265
12	NORETHINDRONE-ETHINYL ESTRAD	531006
13	NORETHINDRONE-ETHINYL ESTRAD	538590
14	NORETHINDRONE-ETHINYL ESTRAD	602957
15	NORETHINDRONE-ETHINYL ESTRAD	620947
16	NORETHINDRONE-ETHINYL ESTRAD	2187086
17	NORETHINDRONE-ETHINYL ESTRAD	2187108
18	NORETHINDRONE-ETHINYL ESTRAD	2189054
19	NORGESTREL-ETHINYL ESTRADIOL	34207
20	NORGESTREL-ETHINYL ESTRADIOL	300640
21	LEVONORGESTREL-ETH ESTRA	579386
22	LEVONORGESTREL-ETH ESTRA	707600
23	LEVONORGESTREL-ETH ESTRA	782416
24	LEVONORGESTREL-ETH ESTRA	782432
25	LEVONORGESTREL-ETH ESTRA	2042320
26	NORGESTREL-ETHINYL ESTRADIOL	2043033
27	LEVONORGESTREL-ETH ESTRA	2043726
28	NORGESTIMATE-ETHINYL ESTRADIOL	2258560
29	NORETHINDRONE-MESTRANOL	30333
30	NORETHINDRONE-MESTRANOL	30341
31	LEVONORGESTREL-ETH ESTRA	2236974
32	ETHYNODIOL D-ETHINYL ESTRADIOL	471526
33	NORETHINDRONE-ETHINYL ESTRAD	340731
34	NORETHINDRONE-MESTRANOL	340758
35	NORETHINDRONE A-E ESTRADIOL	343838
36	NORETHINDRONE A-E ESTRADIOL	353027
37	NORETHINDRONE-ETHINYL ESTRAD	372838
38	NORETHINDRONE-ETHINYL ESTRAD	373273
39	NORETHINDRONE-ETHINYL ESTRAD	531014
40	NORETHINDRONE-ETHINYL ESTRAD	602965
41	NORETHINDRONE-ETHINYL ESTRAD	695734
42	NORETHINDRONE-ETHINYL ESTRAD	2187094
43		
44		
45		
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		

Supplementary material

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1		
2		
3	MEDROXYPROGESTERONE ACET	2246627
4	MEDROXYPROGESTERONE ACET	30937
5	MEDROXYPROGESTERONE ACET	2010739
6	MEDROXYPROGESTERONE ACET	2148560
7	MEDROXYPROGESTERONE ACET	2221292
8	MEDROXYPROGESTERONE ACET	2229839
9	MEDROXYPROGESTERONE ACET	2244727
10	MEDROXYPROGESTERONE ACET	2246628
11	MEDROXYPROGESTERONE ACET	729973
12	MEDROXYPROGESTERONE ACET	2010933
13	MEDROXYPROGESTERONE ACET	2148579
14	MEDROXYPROGESTERONE ACET	2221306
15	MEDROXYPROGESTERONE ACET	2229840
16	MEDROXYPROGESTERONE ACET	2246629
17	MEDROXYPROGESTERONE ACET	30945
18	MEDROXYPROGESTERONE ACET	2267640
19	MEDROXYPROGESTERONE ACET	37605
20	MEDROXYPROGESTERONE ACET	2166704
21	MEDROXYPROGESTERONE ACET	739952
22	MEDROXYPROGESTERONE ACET	1977652
23	MEDROXYPROGESTERONE ACET	2128470
24	MEDROXYPROGESTERONE ACET	2243005
25	NORETHINDRONE	
26	PROGESTERONE,MICRONIZED	
27	PROGESTERONE	
28	PROGESTERONE	
29	PROGESTERONE	
30	PROGESTERONE	
31	LEVONORGESTREL	
32		
33	Transmasculine	
34	Testosterone	
35	TESTOSTERONE	2249499
36	TESTOSTERONE CYPIONATE	30783
37	TESTOSTERONE PROPIONATE	1977571
38	TESTOSTERONE CYPIONATE	1977601
39	TESTOSTERONE CYPIONATE	2220318
40	TESTOSTERONE CYPIONATE	2246063
41	TESTOSTERONE ENANTHATE	29246
42	TESTOSTERONE ENANTHATE	716936
43	TESTOSTERONE ENANTHATE	739944
44	TESTOSTERONE UNDECANOATE	782327
45	TESTOSTERONE ENANTHATE/ESTRAD	108278
46	TESTOSTERONE ENANTHATE/ESTRAD	2061031
47	TESTOSTERONE	2239653
48	TESTOSTERONE	2245346
49	TESTOSTERONE	2245345
50	TESTOSTERONE	2245972
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		

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
	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-9: 410 Acute myocardial infarction 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
	Chronic heart failure	1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	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 ICD-10: I50 Heart failure
	Stroke- hospital	1 or more hospitalizations with relevant ICD codes	ICD-9: 428 Heart failure 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

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

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

Chronic Obstructive
 Pulmonary Disease*

1 or more hospitalization or 2 or more medical visits within 365 days

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.

Supplementary material

Hypertension*

Excludes gestational diabetes.

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

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

F68 Other disorders of adult personality & behavior

ICD-9:

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

For peer review only

Supplementary material

296 Affective psychoses 300
Neurotic disorders 311 Depressive
disorder, not elsewhere classified

MSP DX Code:
50B Anxiety/Depression
*Cancer case definition details
available from the British
Columbia Cancer Agency:
[http://www.bccancer.bc.ca/health-
info/types-of-cancer](http://www.bccancer.bc.ca/health-info/types-of-cancer)*

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)

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
ICD-10:
M15 Polyarthrosis
M16 Coxarthrosis [arthrosis of
hip]
M17 Gonarthrosis [arthrosis of
knee]
M18 Arthrosis of first
carpometacarpal joint
M19 Other arthrosis

Osteoarthritis*

1 or more
hospitalization or 2 or
more medical visits in
365 days with a
relevant ICD code

Supplementary material

1			
2			
3			
4			
5			ICD-9:
6			715 Osteoarthritis and allied
7			disorders
8	Personality disorder	1 or more	ICD-9:
9		hospitalizations or	301.x Personality disorders
10		medical visits with a	
11		relevant diagnosis	ICD-10:
12		within 365 days	F60.x Specified personality
13			disorders
14			F62 Enduring personality
15			changes, not attributable to brain
16			damage and disease
17			F68.1 Intentional production or
18			feigning of symptoms or
19			disabilities, either physical or
20			psychological
21			F68.8 Other specified disorders or
22			adult personality and behaviour
23			F69 Unspecified disorder or adult
24			personality and behaviour
25			ICD-9:
26			295.x Schizophrenic disorders
27	Schizophrenia related	1 or more medical visit	297.0 Paranoid state, simple
28	disorder	or hospitalizations with	297.1 Delusional disorder
29		relevant diagnoses	297.2 Paraphrenia
30		within 365 days	297.3 Shared psychotic disorder
31			
32			ICD-10:
33			F20.x Paranoid schizophrenia
34			F21.x Schizotypal disorder
35			F23.2 Acute schizophrenia-like
36			psychotic disorder
37			F25.x Schizoaffective disorders
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			

* 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	<p>(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up</p> <p><i>Case-control study</i> - 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</p> <p><i>Cross-sectional study</i> - Give the eligibility criteria, and the sources and methods of selection of participants</p> <p>(b) <i>Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed</p> <p><i>Case-control study</i> - For matched studies, give matching criteria and the number of controls per case</p>		<p>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.</p> <p>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.</p> <p>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.</p>	<p>6.1: Methods (pp 6-7); Supplementary Material</p> <p>6.2: Methods (pp 6-7)</p> <p>6.3: Figure 1</p>
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)

1 2 3 4 5 6 7 8 9 10	Bias	9	Describe any efforts to address potential sources of bias			Discussion (pp 9, 11)
11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34	Study size	10	Explain how the study size was arrived at			Methods (page 6); Figure 1
35 36 37 38 39 40 41 42 43 44 45 46 47	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) <i>Cohort study</i> - If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> - If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> - 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 (<i>e.g.</i> , 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>i.e.</i> , 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 (<i>e.g.</i> , 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) <i>Cohort study</i> - summarise follow-up time (<i>e.g.</i> , average and total amount)			Results (page 8)
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time <i>Case-control study</i> - Report numbers in each exposure			Results

		category, or summary measures of exposure <i>Cross-sectional study</i> - 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, Guttman 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.

*Checklist is protected under Creative Commons Attribution ([CC BY](https://creativecommons.org/licenses/by/4.0/)) license.

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts



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.

1
2
3 **1 TITLE PAGE**
4 2

5 3 **Title:** Development of a computable phenotype to identify a transgender sample for health
6 4 research purposes: A feasibility study in a large linked provincial healthcare
7 5 administrative cohort in British Columbia, Canada
8 6

9 7 **Authors:** Rich AJ^{1,2}, Poteat T³, Koehoorn M¹, Li J², Ye M², Sereda, P², Salway T⁴, Hogg RS^{2,4}
10 8
11 9

12 **Affiliations:**

- 13 10 1. School of Population and Public Health, University of British Columbia, Vancouver, Canada
14 11 2. British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada
15 12 3. Department of Social Medicine, University of North Carolina- Chapel Hill, Chapel Hill, USA
16 13 4. Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada
17 14

18 15 **Corresponding author:**

19 16 Ashleigh J Rich
20 17 ajrich@mail.ubc.ca
21 18

22 19 2206 East Mall
23 20 Vancouver, BC
24 21 V6T 1Z3
25 22 Canada
26 23

27 24 **Word count:** 4000/4000
28 25
29 26
30 27
31 28
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 ABSTRACT

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 1 healthcare setting, using a computable phenotype, and capacity to estimate the prevalence
4 2 of transgender status among the population living with HIV in the province.

- 5
6 3 • Development of a validated transgender computable phenotype algorithm lays the
7 4 foundation for future investigation of transgender-specific research questions related to
8 5 general and HIV-specific healthcare use and health outcomes for this key population.
9 6

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1 INTRODUCTION

2 Limited data on transgender people

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

15
16 The erasures or exclusions of transgender persons in health studies may be explained, in part, by
17 methodological challenges. Specific to electronic health record (EHR) data, a 2017 report
18 identified only one transgender person among 38,5820 cancer cases in a Minnesota cancer
19 registry,[9] clearly an undercount given that 0.4% of the US general population and 0.6% (95%
20 credibility intervals: 0.5%-0.7%) of the Minnesota population is estimated to be
21 transgender.[10,11] This highlights the need for improved gender ascertainment and transgender
22 inclusion in research relying on patient records and administrative data. The establishment of
23 best practices for measuring transgender status in survey research, such as the two-step method
24 (measuring sex assigned at birth and current gender identity), points to a way forward for
25 transgender-inclusive population health research.[12,13] However, innovative research methods
26 are needed to identify transgender people in studies that rely on existing data sources (in
27 particular EHR) and that optimize the use of transgender respondents' data in non-transgender
28 specific research.

29 Computable phenotypes for transgender health research

30 Previous research in transgender health largely comprises cross-sectional studies, case reports,
31 and qualitative or observational research.[7] Much consists of clinic- or venue-based
32 convenience samples or lack comparison groups.[7,8] The literature is further characterized by
33 inconsistent transgender status measurement,[14] small sample sizes, and focus on the United
34 States (US).[8] In response, researchers have called for advancing transgender health research
35 methods - namely ascertainment of high-quality samples via systematic approaches - including
36 for general population-based and health systems-based studies.[15] One opportunity for the
37 advancement of transgender health research methods is the emerging use of computable
38 phenotypes (CPs)[16] or case ascertainment algorithms, to identify transgender samples in
39 healthcare utilization data. A computable phenotype is an algorithm for identifying a clinical
40 feature, condition, or set of characteristics that can be determined directly from EHR and other
41 ancillary health care data systems (e.g., disease registries, insurance claims data) data.[17] CPs
42 are developed using a combination of data elements (e.g., sociodemographic variables, clinical
43 diagnoses) and value sets (i.e., the selection of a set of relevant values for each data element).
44 Development of CPs using standardized methods and definitions enables identification and
45 inclusion of transgender persons in research, as well as replication of analyses across data
46

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

3
4
5
6
7 4 Recently, CP and other EHR-based algorithm methods have been applied in a number of settings
8 primarily in the US to identify transgender samples for health research.[14] Specifically, the
9 STRONG study identified a transgender cohort (n = 6,456) using EHR data from Kaiser
10 Permanente health plan members in California and Georgia, for investigation of general and
11 transgender-specific health outcomes.[18] Blosnich et al identified 3,177 people with a “gender
12 identity disorder” diagnosis among military veterans accessing care through the US Veterans
13 Health Administration healthcare system,[19] for examination of mental health and other
14 outcomes. Researchers with the US Centers for Medicare & Medicaid Services identified 4,098
15 transgender beneficiaries using national Medicare claims data,[20] and researchers at Vanderbilt
16 University identified 234 transgender patients in their university clinic EHR data.[16] While
17 these cohorts represent important opportunities for advancement of transgender health research,
18 these methods have yet to be applied widely outside the US context. This is particularly
19 important as different jurisdictions may vary in medical billing and coding practices, healthcare
20 system patient populations, and representativeness of the general population. Specifically, in
21 Canada, healthcare is delivered through a provincially administered universal healthcare system.
22 As such, research using EHR provides an opportunity to develop methods for population-based,
23 representative estimates of transgender populations within the Canadian context. Coupled with
24 the current absence of gender ascertainment measures in population-based routinely collected
25 data (e.g., census, national government health surveys, etc.) in Canada and many other
26 jurisdictions, this remains an evidence need.

27 **Summary of study rationale**

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

32 **METHODS**

33 **Data Sources and Participants**

34 The Comparative Outcomes and Service Utilization Trends Study (COAST)
35 COAST is a population-based cohort study focused on health services utilization research
36 questions among all people known to be living with HIV (PLWH) in the province of British
37 Columbia (BC) and a 10% random sample comparison group of the HIV-negative general
38 population.[21] The COAST cohort comprises individual-level, longitudinal data from PLWH
39 who have ever accessed HIV treatment in BC between 1996 and 2013, provided by Population
40 Data BC (PopDataBC)[21] via data linkage between two provincial data sources, by personal
41 health number: the Drug Treatment Program (DTP) [22] and the Ministry of Health. PopDataBC
42 provides infrastructure for access to, and linkage of, longitudinal and individual-level
43 administrative health data for all BC residents.[23]. The HIV-negative general population cohort
44 was drawn randomly from the Ministry of Health registry data by PopDataBC. The COAST study
45 has received approval from the University of British Columbia/Providence Health Care
46 Research Ethics Board (#H09-02905) and Simon Fraser University Office of Research Ethics
47 (#2013 s0566). The study complies with the BC Freedom of Information and Protection of

1
2
3 1 Privacy Act (FIPPA) and did not require informed consent as it is conducted using
4 2 retrospective administrative and anonymized data for research and statistical purposes
5 3 only. No patients or public were involved in this study.
6 4

5 Drug Treatment Program

6 In BC, antiretroviral therapy (ART) is provided to PLWH at no cost to the patient, and distributed
7 through the DTP.[22] The DTP contributed a provider-reported measure of transgender status for
8 COAST.
9

10 Ministry of Health

11 Ministry of Health data available via COAST included insured medical service billing records for
12 outpatient visits,[24,25] hospital (in-patient) visits,[26] prescription medications,[27,28] and vital
13 statistics.[29]
14

15 **Measures & Analyses**

16 Transgender computable phenotypes

17 Identification of transgender cases was tested in COAST using International Classification of
18 Disease (ICD) codes (9th and 10th editions) and exogenous sex hormone prescription use.

19 Transgender-specific diagnoses in medical and hospital billing records included the ICD-9 codes
20 302.5 Trans-sexualism with unspecified history, 302.51 Trans-sexualism with asexual history,
21 302.52 Trans-sexualism with homosexual history, 302.53 Trans-sexualism with heterosexual
22 history, 302.6 Gender Identity Disorder in children, 302.85 Gender Identity Disorder in
23 adolescents or adults; and ICD-10 codes F64.0 Gender Identity Disorder of childhood, F64.2
24 Gender Identity Disorder of childhood, F64.8 Other Gender Identity Disorder, and F64.9 Gender
25 Identity Disorder unspecified. The full list of androgen blockers and exogenous sex hormone
26 prescriptions included in analyses is available in the supplementary material.
27

28 Concordance

29 To assess face validity and utility of diagnosis and prescription data over time in CP
30 development (i.e. whether the identified transgender sample had exogenous sex hormone
31 prescription use and other diagnoses patterns consistent with that of transgender populations in
32 other studies), concordance analyses evaluated the presence of at least one included diagnosis
33 and prescription during the COAST study follow-up period with the presence of at least one
34 included diagnosis and prescription in the last study year. Concordance was assessed between
35 transgender-specific diagnoses, exogenous sex hormone and androgen blocker prescriptions, and
36 non-transgender specific diagnoses (ICD-9 259.9 Unspecified Endocrine Disorder and ICD-10
37 E34.9 Endocrine Disorder, Unspecified [see supplementary material]). Endocrine disorder
38 diagnosis codes are sometimes preferred by medical providers treating transgender people in
39 response to historic exclusions of transgender-specific care from insurance coverage and to
40 combat the stigma of transgender-specific diagnosis codes that have historically been classified
41 as psychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM).[30]
42 Exogenous sex hormone use, while common in transgender populations,[5,31] is not
43 transgender-specific. Cisgender populations also use androgen blocker and sex hormone
44 prescriptions (e.g. estrogen to treat menopausal symptoms in cisgender women, spironolactone is
45 used for hypertension), thus exogenous sex hormone and androgen blocker prescription use
56
57
58
59
60

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

1
2
3 1 disease, chronic kidney disease, osteoarthritis, schizophrenia and personality disorders, and
4 2 chronic liver disease, but a lower prevalence for hypertension.
5 3

6 4 *[Figure 4 here]*
7 5

6 6 **DISCUSSION**

7 7 This study demonstrates the feasibility of identification of a sample of transgender people in a
8 8 large linked provincial healthcare administrative database, using a CP based on prescriptions and
9 9 diagnoses. Among a growing number of studies using EHR and CP methods to identify
10 10 transgender samples for health research purposes, this is the first to do so in Canada., to
11 11 independently validate the CP using a ‘gold standard’ of provider-reported transgender status,
12 12 and the only to use population-based data.
13 13

14 14 **Concordance**

15 15 There was high concordance between transgender-specific diagnoses and exogenous sex
16 16 hormone or androgen blocker prescription use in this study. That nearly half of those with at
17 17 least one transgender-specific diagnosis had been dispensed hormones or blockers in the past
18 18 year is consistent with findings from US and Canadian studies (48.9% and 43.0%,
19 19 respectively)[20,32,33] - suggesting face validity for the current CP.
20 20

21 21 **CP development and validation**

22 22 The best-performing CP overall successfully identified cisgender people who were truly
23 23 cisgender (specificity) and correctly identified transgender people who were truly transgender
24 24 (0.2% false positive rate, results not shown). However, the selected CP had relatively low
25 25 sensitivity, missing approximately 72.5% of ‘true’ transgender people in COAST, as identified
26 26 by the gold standard provider-based measure. Though a relatively small proportion of the ‘true’
27 27 transgender sample was identified in this study, the impact on future analyses comparing health
28 28 outcomes for transgender and cisgender groups is likely negligible, as even the large proportion
29 29 of ‘true’ transgender people misclassified as cisgender (approximate n=496) is a very small
30 30 proportion of the total COAST sample. At worst, this misclassification would bias results related
31 31 to disparities between transgender and cisgender health toward the null, producing a conservative
32 32 attenuated effect. Further, as discussed below, gender identity classification will likely greatly
33 33 improve as transgender care shifts further into the fee-for-service system in BC. As in other
34 34 Canadian administrative data studies, low sensitivity may be explained in part by provider and
35 35 system billing preferences using 3-digit ICD diagnosis coding instead of the more specific 4-
36 36 digit coding, and inconsistencies in the BC billing management system.[37] Despite the low
37 37 sensitivity, CP development in this study with high specificity offers an advancement for
38 38 transgender health research. A measure that correctly identifies cases for transgender samples in
39 39 research with good success translates to better opportunities to include transgender people in
40 40 health studies and to investigate their health relative to other groups. While future research may
41 41 lead to improvements in CP development, the CP identified in the current study with good
42 42 specificity, albeit relatively poor sensitivity, has important utility in advancing opportunities in
43 43 transgender health research.
44 44

45 45 The limited agreement between the CP and provider-report transgender status may be due to the
46 46 widely varying transgender status prevalence depending on study design and ascertainment
47 47
48 48
49 49
50 50
51 51
52 52
53 53
54 54
55 55
56 56
57 57
58 58
59 59
60 60

1
2
3 1 measures used.[14] In the BC context, the CP and the DTP measures are assessing transgender
4 2 status in different ways and for different purposes. In the DTP, transgender status is ascertained
5 3 in the context of HIV diagnosis and ART prescribing, during which demographics and HIV
6 4 transmission risk factors are recorded. This differs from recording diagnoses in EHR for those
7 5 accessing transgender-specific care as utilized in the CP. This may explain the lower PPV for the
8 6 best-performing CP compared to the CP based on recent transgender diagnoses, suggesting the
9 7 DTP provider-reported transgender status measure has better coverage for recent cases and the
10 8 potential for use of recent diagnosis over ever to be beneficial in future CP development.
11 9 Ultimately, a single CP may not be sufficient for all intended purposes and the best applicable
12 10 CP (using different types of diagnoses, prescriptions or procedures) may differ depending on the
13 11 intended healthcare, health research, or health policy application.[17]

14 12
15 13 There is limited literature on EHR-based studies with the ability to validate an administrative
16 14 transgender measure using a ‘gold standard’ comparison measure.[16] The two previous studies
17 15 that have developed and validated algorithms to identify transgender individuals have both been
18 16 conducted in non-representative samples in the US, one using Medicare data[38] and one in a
19 17 university medical center.[16] Similar to the current study, the Medicare study found high
20 18 specificity when comparing an EHR-based and a two-step survey-based transgender measure.
21 19 However, the Medicare study found that the EHR measure performed consistently well with high
22 20 sensitivity and a high Kappa statistic, unlike in the current study. Using chart review as the ‘gold
23 21 standard’ for comparison of transgender status, Ehrenfeld et al. found a low false positive rate for
24 22 their best-performing algorithm (3%), though not as low as the false positive rate in the current
25 23 study. The overall high levels of agreement for transgender measures in the two previous studies
26 24 is likely a function of the lack of independence between the ‘gold standard’ and the CP or
27 25 algorithm measures. Specifically, only those classified as transgender in the Medicare EHR data
28 26 were offered survey participation to complete the two-step ‘gold standard’ survey measure, and
29 27 only those cases identified as transgender in the university clinic EHR were included in chart
30 28 review. Thus, previous studies could assess agreement between the two measures, but not
31 29 robustly validate either. In the current study, the DTP provider-based transgender status measure
32 30 is independent and thus could be used for robust CP validation.

33 31
34 32 While not possible to incorporate free-text records in case-finding algorithms in the current study
35 33 as only structured EHR data is linked through COAST, it is worth noting the opportunities
36 34 potentiated by use of NLP and machine learning approaches as methods for identifying
37 35 transgender samples in EHR data as this research area continues to grow. Outside of transgender
38 36 health, the use of NLP and machine learning to mine unstructured free-text EHR data has
39 37 demonstrated efficiency in improving case ascertainment algorithm accuracy .[39] As ‘gold
40 38 standard’ two-step sex assigned at birth and current gender identity measures of transgender
41 39 status[12] are slowly being implemented in routinely collected healthcare data sources, in the
42 40 meantime NLPs to extract free-text data can be used to produce better gold standards against
43 41 which to measure algorithm performance, as demonstrated by the Medicare study.[38]

44 42 **Transgender status prevalence & ascertainment**

45 43 Based on a recent meta-analysis of transgender status prevalence in population-based probability
46 44 samples,[10] it was expected that an effective CP would identify 0.4% of the general population
47 45 as transgender, or approximately 54 of the HIV-positive COAST cohort (n=13 907) and 3,098 of
48 46
49 47
50 48
51 49
52 50
53 51
54 52
55 53
56 54
57 55
58 56
59 57
60 58

1
2
3 1 the HIV-negative cohort (n=516 340). Consistent with expectations, the best-performing CP
4 2 identified 51 PLWH as transgender, equivalent to a transgender status prevalence of 0.4% among
5 3 PLWH. Contrary to expectations, the best-performing CP identified less than 5% of the number
6 4 of transgender persons expected in the HIV-negative cohort. This is likely a result of a number of
7 5 factors including the limitation of CPs to the subset of a population accessing care as noted, and
8 6 the result of most transgender people in BC receiving care currently outside the main fee-for-
9 7 service healthcare delivery system. However, it is also consistent with the undercount of
10 8 transgender populations using diagnostic criteria compared to other methods of ascertainment
11 9 demonstrated in other studies.[14]
12
13

14 10
15 11 Using the broadest CP algorithm (any transgender-specific diagnosis ever, n=56) and those
16 12 identified by provider-report together (total n=106), the total transgender PLWH sample would
17 13 represent as high as 0.88% (range: 0.73-1.1%) of the prevalent HIV infections in BC in
18 14 2014.[40] This overrepresentation of transgender people among PLWH is consistent with
19 15 evidence of a disproportionate HIV burden for transgender populations globally,[5,41,42] as well
20 16 as in line with the only other available data on the proportion of PLWH who are transgender,
21 17 from US national surveillance data (2012 data: 1.1%, 95%CI: 0.8-1.4).[43]
22
23

24 19 **Demographics and chronic conditions**

25 20 Despite moderate to low performance by some validation metrics, particularly low sensitivity,
26 21 the CP was able to detect meaningful results in the characterization of demographics and chronic
27 22 condition burden for the transgender sample - supporting CP face validity. The population
28 23 density and age distribution by HIV-status of transgender people in this study is largely
29 24 consistent with general population patterns, as well as the larger COAST cohort.[21,36] The
30 25 overall higher burden of chronic illness for transgender people living with HIV versus without
31 26 HIV in this study is consistent with elevated chronic illness risk and morbidity among non-
32 27 transgender PLWH.[44] This higher chronic disease burden is linked to HIV disease processes
33 28 and related inflammatory immune response.[45] While a small but growing number of studies
34 29 have begun to investigate the chronic illness burden for transgender populations in other
35 30 industrialized settings,[16,19,46–48] including using EHR data, findings vary widely due to
36 31 differences in sampling, study design, setting and measurement.
37
38
39

40 33 **Limitations**

41 34 Findings from this study should be interpreted in the context of a few key limitations. CPs are by
42 35 design only applicable to people accessing healthcare services, often motivated by illness and
43 36 aided by the ability to access care. As such, this study is limited to those transgender people
44 37 accessing medical transition care in BC and may only represent 24% to 47% of the total
45 38 transgender population.[33] This study was also limited by the inability to validate the
46 39 transgender CP among the HIV-negative COAST cohort, as a 'gold standard' provider-based
47 40 transgender measure was only available for the HIV-positive cohort. It is possible that the
48 41 transgender CPs would perform differently in populations living without HIV, particularly as
49 42 healthcare contact is higher among populations living with HIV. Additionally, this study should
50 43 be considered in light of the context in which it was conducted, an environment in which
51 44 transgender healthcare delivery in BC is currently shifting from specialized care settings to the
52 45 main primary care fee-for-service settings. Given that COAST only includes fee-for-service data,
53 46 this study was limited by the inability to capture transgender people who access transgender care
54
55
56
57
58
59
60

1
2
3 1 outside the fee-for-service system. However, fortunately, as the shift to the fee-for-service
4 2 system occurs, transgender ascertainment via CPs in BC will likely improve. The administrative
5 3 data used in this study may also be susceptible to coding error (and coding biases/practices)
6 4 across conditions and settings,[49] potentially introducing misclassification bias in terms of
7 5 transgender ascertainment. Finally, chronic condition prevalence data reported in this study
8 6 should be interpreted with caution, given potential selection bias by serostatus in the COAST
9 7 cohort; though any such bias likely resulted in conservative estimates of difference by serostatus
10 8 in this analysis.
11 9

10 CONCLUSION

11 This study makes a number of important contributions to the literature on innovative methods in
12 12 transgender health. Major contributions include development and validation of a transgender CP,
13 13 using a population-based representative source population, in the Canadian context. Another
14 14 strength is the approximately complete ascertainment of the population of transgender PLWH in
15 15 BC, and capacity to estimate transgender status prevalence among PLWH. In a current funding
16 16 environment of limited support for longitudinal transgender health studies in the US and none to
17 17 date in Canada, this study and the methods employed offer an efficient, replicable and cost-
18 18 effective way forward in creating electronic cohorts for advancing transgender health
19 19 research.[15] Moreover, the recent rollback of sexual orientation and gender identity data
20 20 collection and legal changes in insurance coverage of transgender healthcare in the US potentiate
21 21 decline in accurate claims coding for gender-affirming care.[30] This highlights the utility of
22 22 work in this area from other jurisdictions, particularly those with transgender-inclusive universal
23 23 healthcare systems such as Canada.

24 24
25 25 Future research should build upon the methods developed in this study and explore
26 26 complimentary approaches for gender identity ascertainment in administrative and EHR data,
27 27 such as machine learning approaches, as have been used to develop algorithms based on
28 28 healthcare utilization data in other research areas. Finally, the current study lays the foundation
29 29 for future work with the ability to study transgender health and healthcare use patterns over time,
30 30 with linkage to laboratory data, as well as inclusion of appropriate comparison groups.[15,50]

31 ACKNOWLEDGEMENTS & DISCLAIMER

32 32 The authors thank the COAST study participants, the BC Centre for Excellence in HIV/AIDS,
33 33 the BC Ministry of Health, BC Vital Statistics Agency, PharmaNet and the institutional data
34 34 stewards for granting access to the data, and Population Data BC for facilitating the data linkage
35 35 process. In addition, we would like to thank the COAST core team members and other support
36 36 staff at these institutions for their administrative assistance with the data access and preparation.
37 37 All inferences, opinions, and conclusions drawn in this paper are those of the authors, and do not
38 38 reflect the opinions or policies of the Data Steward(s). There are no conflicts of interest to
39 39 declare.
40 40

41 COMPETING INTERESTS

42 42 None declared.
43 43

44 FUNDING

45 45

1
2
3 1 This work was supported by the Canadian Institutes of Health Research, through an Operation
4 2 Grant [grant number 130419], a Foundation Award to RSH [grant number 143342], and a
5 3 Doctoral Research Award to AJR [grant number 152382] and support from the British Columbia
6 4 Centre for Excellence in HIV/AIDS. The DTP receives funding from the provincial government
7 5 of British Columbia (PharmaCare). The funders had no role in the study design, analysis,
8 6 interpretation of the data, drafting of the manuscript or in the decision to submit for
9 7 publication.
10 8

11 9 **CONTRIBUTORS**

12 10 AJR led the study from conceptualization to analysis plan to interpretation, drafting of the first
13 11 manuscript version, revisions and final version. RSH acquired study data and funding. TP, MK,
14 12 PS, TS, and RSH all contributed to study design, interpretation of results, and reviewed manuscript
15 13 versions. JL and MY contributed to study analysis and reviewed manuscript versions. All authors
16 14 provided critical review of first and subsequent manuscript drafts, approved the final version, and
17 15 agree to be accountable for the work presented.
18 16

19 17 **PATIENT CONSENT FOR PUBLICATION**

20 18 Not required.
21 19

22 20 **DATA SHARING STATEMENT**

23 21 The data used for this study are held by the BC Centre for Excellence in HIV/AIDS under the
24 22 authority of the BC Ministry of Health; as they contain confidential patient health records
25 23 including HIV serostatus, data are cannot be made available to other parties.
26 24
27 25
28 26
29 27
30 28
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **FIGURES LEGENDS**

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

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

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

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

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

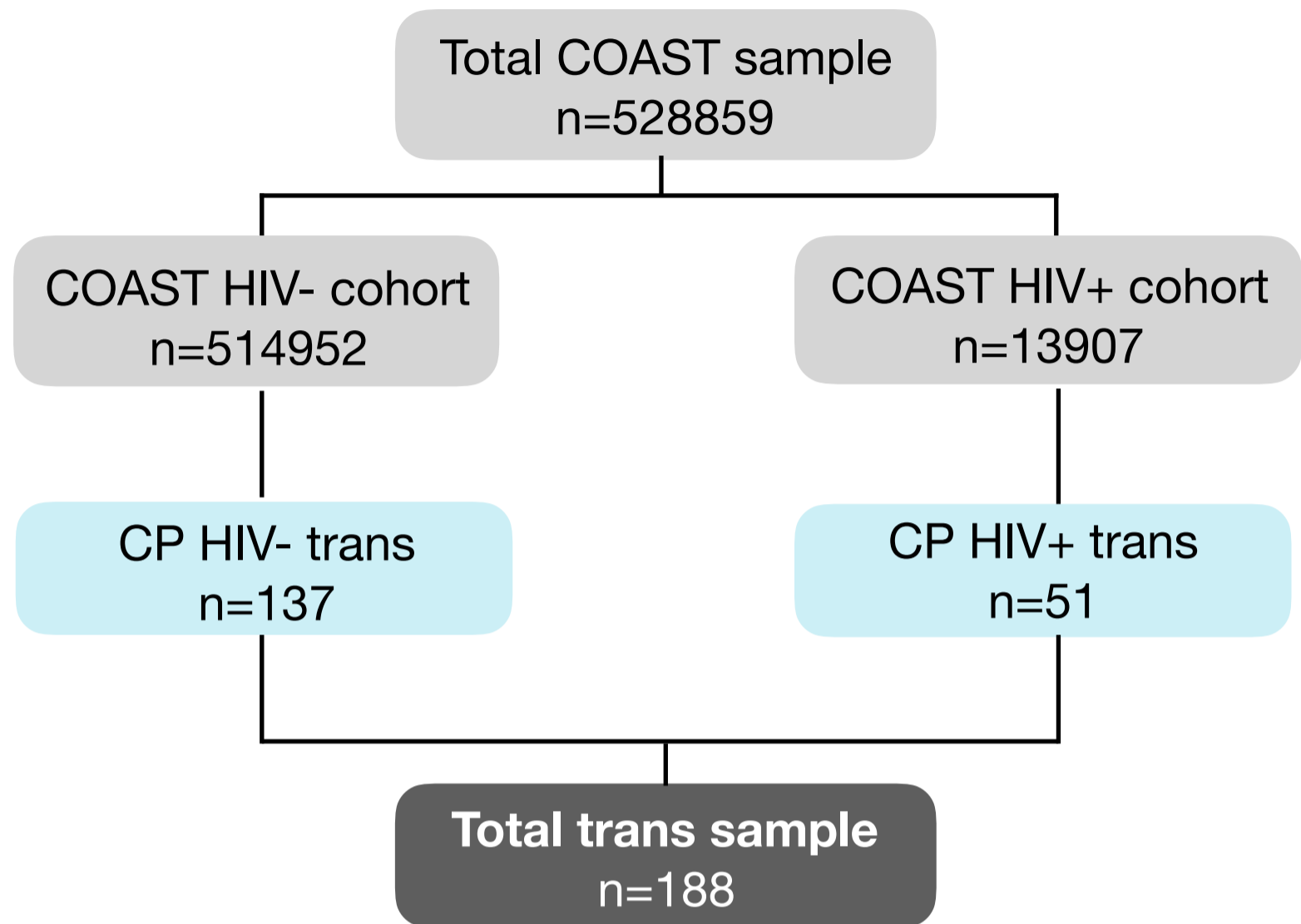
1 REFERENCES

- 2 1. Bauer GR. Making sure everyone counts: Considerations for inclusion, identification, and
3 analysis of transgender and transsexual participants in health surveys. *What Differ Sex Gen*
4 *Make Gen Sex Health Res Casebook*. Vancouver, Canada: Canadian Institutes of Health
5 Research (CIHR), Institute for Gender and Health (IGH); 2012. p. 59–67.
- 6 2. Bauer GR, Hammond R, Travers R, Kaay M, Hohenadel KM, Boyce M. “I don’t think this is
7 theoretical; this is our lives”: How erasure impacts health care for transgender people. *JANAC*.
8 2009;20:348–361.
- 9 3. Poteat T, German D, Flynn C. The conflation of gender and sex: Gaps and opportunities in
10 HIV data among transgender women and MSM. *Glob Public Health*. 2016;11:835–848.
- 11 4. Wanta JW, Unger CA. Review of the transgender literature: Where do we go from here?
12 *Transgender Health*. 2017;2:119–128.
- 13 5. Herbst JH, Jacobs ED, Finlayson TJ, McKleroy VS, Neumann MS, Crepaz N, et al.
14 Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a
15 systematic review. *AIDS Behav*. 2008;12:1–17.
- 16 6. Becasen JS, Denard CL, Mullins MM, Higa DH, Sipe TA. Estimating the prevalence of HIV
17 and sexual behaviors among the US transgender population: A systematic review and meta-
18 analysis, 2006–2017. *Am J Public Health*. 2019;109:e1–8.
- 19 7. Marshall Z, Welch V, Thomas J, Brunger F, Swab M, Shemilt I, et al. Documenting research
20 with transgender and gender diverse people: protocol for an evidence map and thematic analysis.
21 *Syst Rev*. 2017;6:35.
- 22 8. Reisner SL, Poteat T, Keatley J, Cabral M, Mothopeng T, Dunham E, et al. Global health
23 burden and needs of transgender populations: a review. *The Lancet*. 2016;388:412–436.
- 24 9. Cathcart-Rake EJ, Lightner DJ, Quevedo FJ, Ertz DC, Jatoi A. Cancer in transgender patients:
25 One case in 385,820 is indicative of a paucity of data. *J Oncol Pract*. 2017;JOP2017027714.
- 26 10. Meerwijk EL, Sevelius JM. Transgender population size in the United States: A meta-
27 regression of population-based probability samples. *Am J Public Health*. 2017;107:e1–e8.
- 28 11. Flores AR, Herman JL, Gates GJ, Brown TNT. How Many Adults Identify as Transgender in
29 the United States? [Internet]. Los Angeles, CA: Williams Institute; 2016 Jun p. 13. Available
30 from: [https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-](https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf)
31 [Transgender-in-the-United-States.pdf](https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf)
- 32 12. The GenIUSS Group. Best Practices for Asking Questions to Identify Transgender and Other
33 Gender Minority Respondents on Population-Based Surveys. Los Angeles, CA: The Williams
34 Institute; 2014.

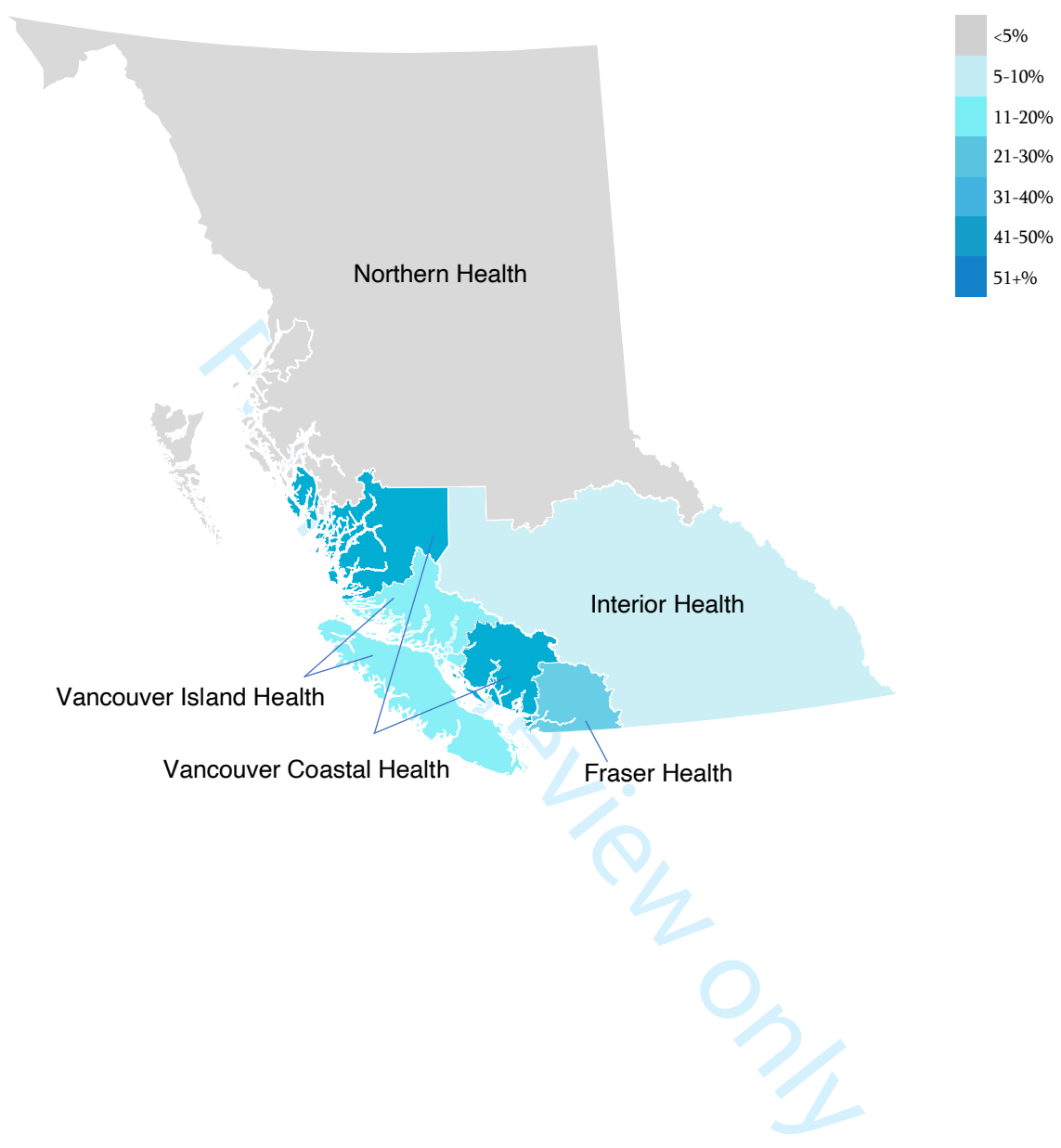
13. Bauer GR, Braimoh J, Scheim AI, Dharma C. Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. *PLOS ONE*. 2017;12:e0178043.
14. Collin L, Reisner SL, Tangpricha V, Goodman M. Prevalence of transgender depends on the “case” definition: 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 Textbook 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. HIV/AIDS. 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>

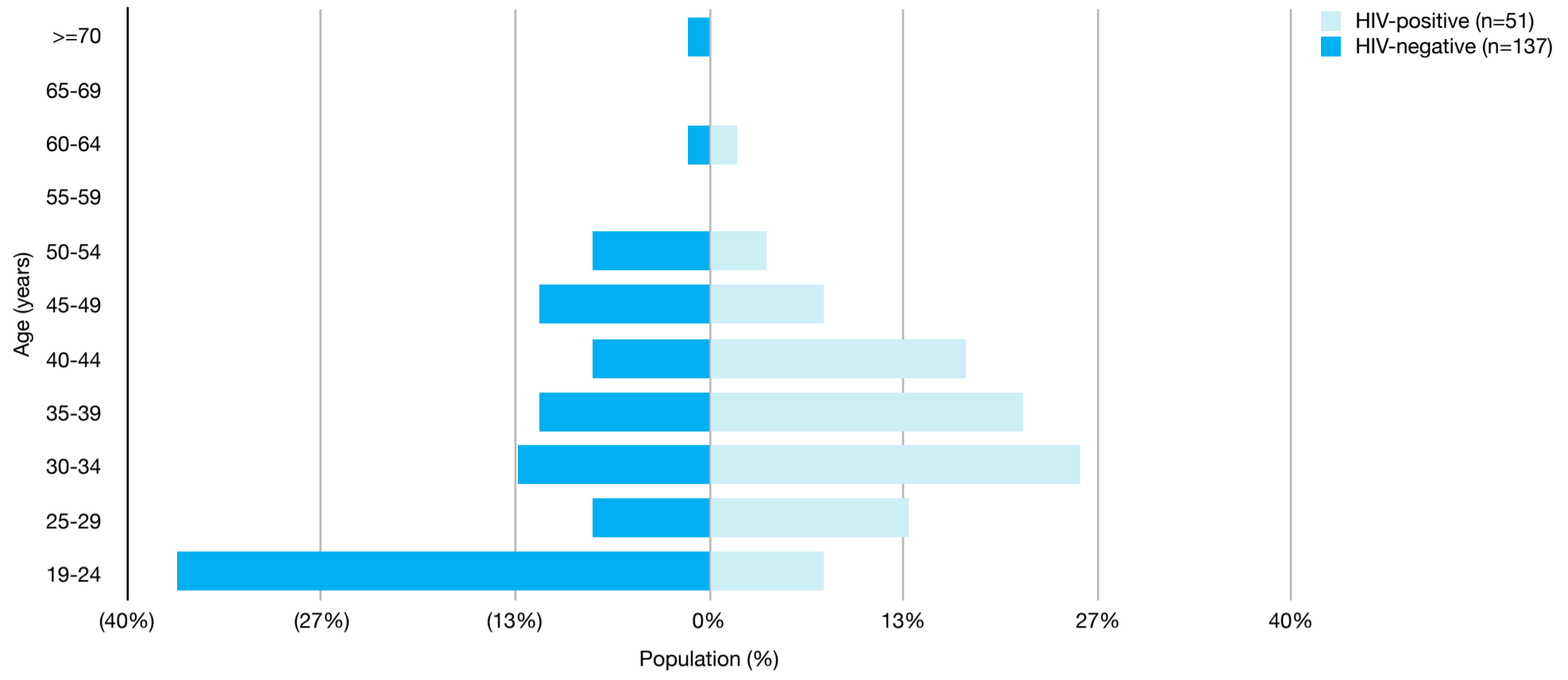
- 1
2
3 1 26. Canadian Institute for Health Information (2014). Discharge Abstract Database (Hospital
4 2 Separations) [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from:
5 3 <http://www.popdata.bc.ca/data>
6
7
8 4 27. British Columbia Ministry of Health (2014). PharmaCare [Internet]. Popul. Data BC. 2014
9 5 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca/data>
10
11 6 28. British Columbia Ministry of Health (2014). PharmaNet [Internet]. Popul. Data BC. 2014
12 7 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca/data>
13
14 8 29. British Columbia Vital Statistics Agency (2014). Vital Statistics Deaths [Internet]. Popul.
15 9 Data BC. 2014 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca>
16
17
18 10 30. Arnold JD, Nelson AE, Loubier EM. Trends in Insurance Coverage for Gender-Affirming
19 11 Surgeries. *JAMA Surg.* 2018;153:972.
20
21 12 31. Radix A, Sevelius J, Deutsch MB. Transgender women, hormonal therapy and HIV
22 13 treatment: a comprehensive review of the literature and recommendations for best practices. *J Int*
23 14 *AIDS Soc.* 2016;19:20810.
24
25 15 32. Rotondi NK, Bauer GR, Scanlon K, Kaay M, Travers R, Travers A. Nonprescribed hormone
26 16 use and self-performed surgeries: “do-it-yourself” transitions in transgender communities in
27 17 Ontario, Canada. *Am J Public Health.* 2013;103:1830–1836.
28
29
30 18 33. Scheim AI, Bauer GR. Sex and gender diversity among transgender persons in Ontario,
31 19 Canada: results from a respondent-driven sampling survey. *J Sex Res.* 2015;52:1–14.
32
33 20 34. Chronic Disease Information Working Group. BC Chronic Disease and Selected Procedure
34 21 Case Definitions [Internet]. British Columbia, Canada: BC Ministry of Health; 2019 Apr p. 1–
35 22 132. Available from: [https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-](https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf)
36 23 [access/chronic-disease-registries-case-definitions.pdf](https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf)
37
38
39 24 35. British Columbia Cancer Agency. BC Cancer Registry Data [Internet]. Popul. Data BC. 2014
40 25 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca/data>
41
42 26 36. BC Stats. Sub-provincial Population Estimates [Internet]. 2018 [cited 2019 Feb 1]. Available
43 27 from: <file://localhost/Users/AJRich/Zotero/storage/5VM2LBUN/PopulationEstimates.html>
44
45 28 37. Hu W. Diagnostic Codes in MSP Claim Data. Program Monitoring and Information
46 29 Management Branch, Resource Management Division, Medical Services Plan; 1996 p. 39.
47
48 30 38. Gerth J, Becerra-Culqui T, Bradlyn A, Getahun D, Hunkeler EM, Lash TL, et al. Agreement
49 31 between medical records and self-reports: Implications for transgender health research. *Rev*
50 32 *Endocr Metab Disord.* 2018;19:263–269.
51
52
53 33 39. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al.
54 34 Development of phenotype algorithms using electronic medical records and incorporating natural
55
56
57
58
59
60

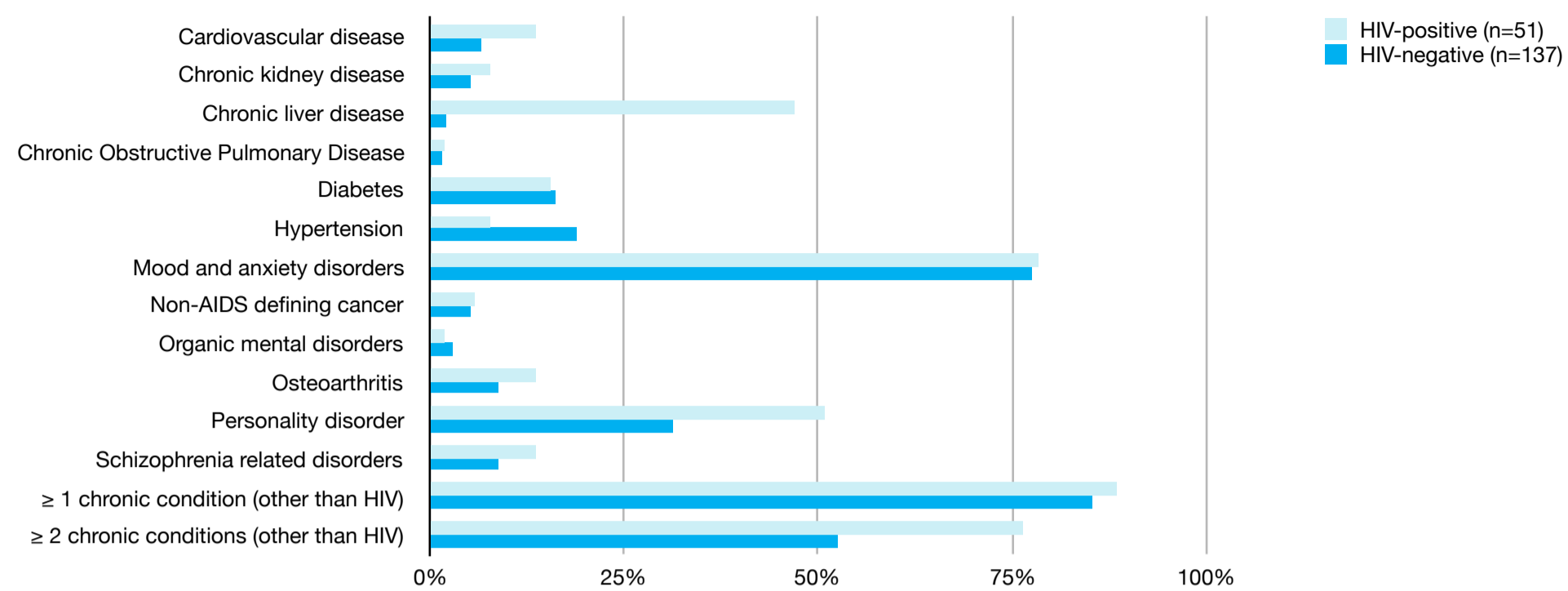
- 1
2
3 1 language processing. The BMJ [Internet]. 2015 [cited 2020 Oct 22];350. Available from:
4 2 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4707569/>
5
6 3 40. BC Centre for Disease Control. HIV in British Columbia: Annual Surveillance Report 2015.
7 4 Vancouver, Canada: BC Centre for Disease Control; 2017.
8
9 5 41. Poteat T, Scheim A, Xavier J, Reisner S, Baral S. Global epidemiology of HIV infection and
10 6 related syndemics affecting transgender people. *J Acquir Immune Defic Syndr*. 2016;72 Suppl
11 7 3:S210–9.
12
13 8 42. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of
14 9 HIV in transgender women: a systematic review and meta-analysis. *Lancet Infect Dis*.
15 10 2013;13:214–222.
16
17 11 43. Centers for Disease Control and Prevention. Behavioral and Clinical Characteristics of
18 12 Persons Receiving Medical Care for HIV Infection—Medical Monitoring Project, United States,
19 13 2012 [Internet]. Washington, D.C.: Centers for Disease Control; 2015 Aug p. 42. Report No.: 12.
20 14 Available from: [https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-](https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf)
21 15 [HSSR_MMP_2012.pdf](https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf)
22
23 16 44. Grabovac I, Veronese N, Stefanac S, Haider S, Jackson SE, Koyanagi A, et al. Human
24 17 Immunodeficiency Virus Infection and Diverse Physical Health Outcomes: An Umbrella Review
25 18 of Meta-analyses of Observational Studies. *Clin Infect Dis*. 2019;ciz539.
26
27 19 45. Nasi M, De Biasi S, Gibellini L, Bianchini E, Pecorini S, Bacca V, et al. Ageing and
28 20 inflammation in patients with HIV infection: Ageing, inflammation and HIV infection. *Clin Exp*
29 21 *Immunol*. 2017;187:44–52.
30
31 22 46. Judge C., O’Donovan C., Callaghan G., Gaoatswe G., O’Shea D. Gender dysphoria -
32 23 prevalence and co-morbidities in an Irish adult population. *Front Endocrinol*. 2014;5:87.
33
34 24 47. Cheung AS, Ooi O, Leemaqz S, Cundill P, Silberstein N, Bretherton I, et al.
35 25 Sociodemographic and clinical characteristics of transgender adults in Australia. *Transgender*
36 26 *Health*. 2018;3:229–38.
37
38 27 48. Swartz JA, Ducheny K, Holloway T, Stokes L, Willis S, Kuhns LM. A latent class analysis
39 28 of chronic health conditions among HIV-Positive transgender women of color. *AIDS Behav*
40 29 [Internet]. 2019 [cited 2019 Oct 2]; Available from: [http://link.springer.com/10.1007/s10461-](http://link.springer.com/10.1007/s10461-019-02543-3)
41 30 [019-02543-3](http://link.springer.com/10.1007/s10461-019-02543-3)
42
43 31 49. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy
44 32 of Medicare’s hospital claims data: progress has been made, but problems remain. *Am J Public*
45 33 *Health*. 2011;
46
47 34 50. Reisner SL, Radix A, Deutsch MB. Integrated and gender-affirming transgender clinical care
48 35 and research. *J Acquir Immune Defic Syndr*. 2016;72 Suppl 3:S235–42.
49
50
51
52
53
54
55
56
57
58
59
60



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60







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		
Spirolactone		
	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

1		
2		
3		
4	ETHYNODIOL D-ETHINYL ESTRADIOL	28630
5	ETHYNODIOL D-ETHINYL ESTRADIOL	469327
6	NORETHINDRONE-MESTRANOL	22608
7	NORETHINDRONE-MESTRANOL	22659
8		
9	NORETHINDRONE A-E ESTRADIOL	297143
10	NORETHINDRONE A-E ESTRADIOL	315966
11	NORETHINDRONE-ETHINYL ESTRAD	317047
12	NORETHINDRONE-ETHINYL ESTRAD	372846
13	NORETHINDRONE-ETHINYL ESTRAD	373265
14	NORETHINDRONE-ETHINYL ESTRAD	531006
15	NORETHINDRONE-ETHINYL ESTRAD	538590
16	NORETHINDRONE-ETHINYL ESTRAD	602957
17	NORETHINDRONE-ETHINYL ESTRAD	620947
18	NORETHINDRONE-ETHINYL ESTRAD	2187086
19	NORETHINDRONE-ETHINYL ESTRAD	2187108
20	NORETHINDRONE-ETHINYL ESTRAD	2189054
21	NORETHINDRONE-ETHINYL ESTRAD	34207
22	NORGESTREL-ETHINYL ESTRADIOL	300640
23	NORGESTREL-ETHINYL ESTRADIOL	579386
24	LEVONORGESTREL-ETH ESTRA	707600
25	LEVONORGESTREL-ETH ESTRA	782416
26	LEVONORGESTREL-ETH ESTRA	782432
27	LEVONORGESTREL-ETH ESTRA	2042320
28	LEVONORGESTREL-ETH ESTRA	2043033
29	NORGESTREL-ETHINYL ESTRADIOL	2043033
30	LEVONORGESTREL-ETH ESTRA	2043726
31	NORGESTREL-ETHINYL ESTRADIOL	2258560
32	NORGESTIMATE-ETHINYL ESTRADIOL	2258560
33	NORETHINDRONE-MESTRANOL	30333
34	NORETHINDRONE-MESTRANOL	30341
35	NORETHINDRONE-MESTRANOL	30341
36	LEVONORGESTREL-ETH ESTRA	2236974
37	LEVONORGESTREL-ETH ESTRA	2236974
38	ETHYNODIOL D-ETHINYL ESTRADIOL	471526
39	ETHYNODIOL D-ETHINYL ESTRADIOL	471526
40	NORETHINDRONE-ETHINYL ESTRAD	340731
41	NORETHINDRONE-ETHINYL ESTRAD	340731
42	NORETHINDRONE-MESTRANOL	340758
43	NORETHINDRONE-MESTRANOL	340758
44	NORETHINDRONE A-E ESTRADIOL	343838
45	NORETHINDRONE A-E ESTRADIOL	343838
46	NORETHINDRONE A-E ESTRADIOL	353027
47	NORETHINDRONE A-E ESTRADIOL	353027
48	NORETHINDRONE-ETHINYL ESTRAD	372838
49	NORETHINDRONE-ETHINYL ESTRAD	372838
50	NORETHINDRONE-ETHINYL ESTRAD	373273
51	NORETHINDRONE-ETHINYL ESTRAD	373273
52	NORETHINDRONE-ETHINYL ESTRAD	531014
53	NORETHINDRONE-ETHINYL ESTRAD	531014
54	NORETHINDRONE-ETHINYL ESTRAD	602965
55	NORETHINDRONE-ETHINYL ESTRAD	602965
56	NORETHINDRONE-ETHINYL ESTRAD	695734
57	NORETHINDRONE-ETHINYL ESTRAD	695734
58	NORETHINDRONE-ETHINYL ESTRAD	2187094
59	NORETHINDRONE-ETHINYL ESTRAD	2187094
60		

Supplementary material

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1		
2		
3	MEDROXYPROGESTERONE ACET	2246627
4	MEDROXYPROGESTERONE ACET	30937
5	MEDROXYPROGESTERONE ACET	2010739
6	MEDROXYPROGESTERONE ACET	2148560
7	MEDROXYPROGESTERONE ACET	2221292
8	MEDROXYPROGESTERONE ACET	2229839
9	MEDROXYPROGESTERONE ACET	2244727
10	MEDROXYPROGESTERONE ACET	2246628
11	MEDROXYPROGESTERONE ACET	729973
12	MEDROXYPROGESTERONE ACET	2010933
13	MEDROXYPROGESTERONE ACET	2148579
14	MEDROXYPROGESTERONE ACET	2221306
15	MEDROXYPROGESTERONE ACET	2229840
16	MEDROXYPROGESTERONE ACET	2246629
17	MEDROXYPROGESTERONE ACET	30945
18	MEDROXYPROGESTERONE ACET	2267640
19	MEDROXYPROGESTERONE ACET	2267640
20	MEDROXYPROGESTERONE ACET	2267640
21	MEDROXYPROGESTERONE ACET	2267640
22	MEDROXYPROGESTERONE ACET	2267640
23	MEDROXYPROGESTERONE ACET	2267640
24	MEDROXYPROGESTERONE ACET	2267640
25	MEDROXYPROGESTERONE ACET	2267640
26	MEDROXYPROGESTERONE ACET	2267640
27	MEDROXYPROGESTERONE ACET	2267640
28	MEDROXYPROGESTERONE ACET	2267640
29	MEDROXYPROGESTERONE ACET	2267640
30	MEDROXYPROGESTERONE ACET	2267640
31	MEDROXYPROGESTERONE ACET	2267640
32	MEDROXYPROGESTERONE ACET	2267640
33	Transmasculine	
34	Testosterone	
35	TESTOSTERONE	2249499
36	TESTOSTERONE CYPIONATE	30783
37	TESTOSTERONE PROPIONATE	1977571
38	TESTOSTERONE CYPIONATE	1977601
39	TESTOSTERONE CYPIONATE	2220318
40	TESTOSTERONE CYPIONATE	2246063
41	TESTOSTERONE ENANTHATE	29246
42	TESTOSTERONE ENANTHATE	716936
43	TESTOSTERONE ENANTHATE	739944
44	TESTOSTERONE UNDECANOATE	782327
45	TESTOSTERONE ENANTHATE/ESTRAD	108278
46	TESTOSTERONE ENANTHATE/ESTRAD	2061031
47	TESTOSTERONE	2239653
48	TESTOSTERONE	2245346
49	TESTOSTERONE	2245345
50	TESTOSTERONE	2245972
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		

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
	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-9: 410 Acute myocardial infarction 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
	Chronic heart failure	1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	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 ICD-10: I50 Heart failure
	Stroke- hospital	1 or more hospitalizations with relevant ICD codes	ICD-9: 428 Heart failure 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

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

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

Chronic Obstructive
 Pulmonary Disease*

1 or more hospitalization or 2 or more medical visits within 365 days

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.

Supplementary material

Hypertension*

Excludes gestational diabetes.

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

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

F68 Other disorders of adult personality & behavior

ICD-9:

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

Supplementary material

296 Affective psychoses 300
Neurotic disorders 311 Depressive
disorder, not elsewhere classified

MSP DX Code:
50B Anxiety/Depression
*Cancer case definition details
available from the British
Columbia Cancer Agency:
[http://www.bccancer.bc.ca/health-
info/types-of-cancer](http://www.bccancer.bc.ca/health-info/types-of-cancer)*

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)

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 Amnesic 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

1			
2			
3			
4			
5			ICD-9:
6			715 Osteoarthritis and allied
7			disorders
8	Personality disorder	1 or more	ICD-9:
9		hospitalizations or	301.x Personality disorders
10		medical visits with a	
11		relevant diagnosis	ICD-10:
12		within 365 days	F60.x Specified personality
13			disorders
14			F62 Enduring personality
15			changes, not attributable to brain
16			damage and disease
17			F68.1 Intentional production or
18			feigning of symptoms or
19			disabilities, either physical or
20			psychological
21			F68.8 Other specified disorders or
22			adult personality and behaviour
23			F69 Unspecified disorder or adult
24			personality and behaviour
25			ICD-9:
26			295.x Schizophrenic disorders
27	Schizophrenia related	1 or more medical visit	297.0 Paranoid state, simple
28	disorder	or hospitalizations with	297.1 Delusional disorder
29		relevant diagnoses	297.2 Paraphrenia
30		within 365 days	297.3 Shared psychotic disorder
31			
32			ICD-10:
33			F20.x Paranoid schizophrenia
34			F21.x Schizotypal disorder
35			F23.2 Acute schizophrenia-like
36			psychotic disorder
37			F25.x Schizoaffective disorders
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			

* 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)

<p>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27</p> <p>Participants</p>	<p>6</p>	<p>(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> - 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 <i>Cross-sectional study</i> - Give the eligibility criteria, and the sources and methods of selection of participants</p> <p>(b) <i>Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> - For matched studies, give matching criteria and the number of controls per case</p>		<p>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.</p> <p>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.</p> <p>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.</p>	<p>6.1: Methods (pp 6-7); Supplementary Material</p> <p>6.2: Methods (pp 6-7)</p> <p>6.3: Figure 1</p>
<p>28 29 30 31 32 33 34</p> <p>Variables</p>	<p>7</p>	<p>Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.</p>		<p>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.</p>	<p>7.1: Methods (pp 6-7); Supplementary Material</p>
<p>35 36 37 38 39 40 41 42 43 44 45 46 47</p> <p>Data sources/ measurement</p>	<p>8</p>	<p>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</p>			<p>Methods (pp 6-7)</p>

1 2 3 4 5 6 7 8 9 10	Bias	9	Describe any efforts to address potential sources of bias			Discussion (pp 9, 11)
11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34	Study size	10	Explain how the study size was arrived at			Methods (page 6); Figure 1
35 36 37 38 39 40 41 42 43 44 45 46 47	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) <i>Cohort study</i> - If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> - If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> - 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 (<i>e.g.</i> , 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>i.e.</i> , 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 (<i>e.g.</i> , 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) <i>Cohort study</i> - summarise follow-up time (<i>e.g.</i> , average and total amount)			Results (page 8)
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time <i>Case-control study</i> - Report numbers in each exposure			Results

		category, or summary measures of exposure <i>Cross-sectional study</i> - 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, Guttman 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.

*Checklist is protected under Creative Commons Attribution ([CC BY](https://creativecommons.org/licenses/by/4.0/)) license.

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts



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.

1
2
3 **1 TITLE PAGE**
4 2

5 3 **Title:** Development of a computable phenotype to identify a transgender sample for health
6 4 research purposes: A feasibility study in a large linked provincial healthcare
7 5 administrative cohort in British Columbia, Canada
8 6

9 7 **Authors:** Rich AJ^{1,2}, Poteat T³, Koehoorn M¹, Li J², Ye M², Sereda, P², Salway T⁴, Hogg RS^{2,4}
10 8
11 9

12 **Affiliations:**

- 13 10 1. School of Population and Public Health, University of British Columbia, Vancouver, Canada
14 11 2. British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada
15 12 3. Department of Social Medicine, University of North Carolina- Chapel Hill, Chapel Hill, USA
16 13 4. Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada
17 14

18 15 **Corresponding author:**

19 16 Ashleigh J Rich
20 17 ajrich@mail.ubc.ca
21 18

22 19 2206 East Mall
23 20 Vancouver, BC
24 21 V6T 1Z3
25 22 Canada
26 23
27 24
28 25
29 26
30 27
31 28
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 ABSTRACT

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 1 healthcare setting, using a computable phenotype, and capacity to estimate the prevalence
4 2 of transgender status among the population living with HIV in the province.

- 5
6 3 • Development of a validated transgender computable phenotype algorithm lays the
7 4 foundation for future investigation of transgender-specific research questions related to
8 5 general and HIV-specific healthcare use and health outcomes for this key population.
9 6

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1 INTRODUCTION

2 Limited data on transgender people

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

15
16 The erasures or exclusions of transgender persons in health studies may be explained, in part, by
17 methodological challenges. Specific to electronic health record (EHR) data, a 2017 report
18 identified only one transgender person among 38,5820 cancer cases in a Minnesota cancer
19 registry,[9] clearly an undercount given that 0.4% of the US general population and 0.6% (95%
20 credibility intervals: 0.5%-0.7%) of the Minnesota population is estimated to be
21 transgender.[10,11] This highlights the need for improved gender ascertainment and transgender
22 inclusion in research relying on patient records and administrative data. The establishment of
23 best practices for measuring transgender status in survey research, such as the two-step method
24 (measuring sex assigned at birth and current gender identity), points to a way forward for
25 transgender-inclusive population health research.[12,13] However, innovative research methods
26 are needed to identify transgender people in studies that rely on existing data sources (in
27 particular EHR) and that optimize the use of transgender respondents' data in non-transgender
28 specific research.

29 Computable phenotypes for transgender health research

30 Previous research in transgender health largely comprises cross-sectional studies, case reports,
31 and qualitative or observational research.[7] Much consists of clinic- or venue-based
32 convenience samples or lack comparison groups.[7,8] The literature is further characterized by
33 inconsistent transgender status measurement,[14] small sample sizes, and focus on the United
34 States (US).[8] In response, researchers have called for advancing transgender health research
35 methods - namely ascertainment of high-quality samples via systematic approaches - including
36 for general population-based and health systems-based studies.[15] One opportunity for the
37 advancement of transgender health research methods is the emerging use of computable
38 phenotypes (CPs)[16] or case ascertainment algorithms, to identify transgender samples in
39 healthcare utilization data. A computable phenotype is an algorithm for identifying a clinical
40 feature, condition, or set of characteristics that can be determined directly from EHR and other
41 ancillary health care data systems (e.g., disease registries, insurance claims data) data.[17] CPs
42 are developed using a combination of data elements (e.g., sociodemographic variables, clinical
43 diagnoses) and value sets (i.e., the selection of a set of relevant values for each data element).
44 Development of CPs using standardized methods and definitions enables identification and
45 inclusion of transgender persons in research, as well as replication of analyses across data
46

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

3
4
5
6
7 4 Recently, CP and other EHR-based algorithm methods have been applied in a number of settings
8 primarily in the US to identify transgender samples for health research.[14] Specifically, the
9 STRONG study identified a transgender cohort (n = 6,456) using EHR data from Kaiser
10 Permanente health plan members in California and Georgia, for investigation of general and
11 transgender-specific health outcomes.[18] Blosnich et al identified 3,177 people with a “gender
12 identity disorder” diagnosis among military veterans accessing care through the US Veterans
13 Health Administration healthcare system,[19] for examination of mental health and other
14 outcomes. Researchers with the US Centers for Medicare & Medicaid Services identified 4,098
15 transgender beneficiaries using national Medicare claims data,[20] and researchers at Vanderbilt
16 University identified 234 transgender patients in their university clinic EHR data.[16] While
17 these cohorts represent important opportunities for advancement of transgender health research,
18 these methods have yet to be applied widely outside the US context. This is particularly
19 important as different jurisdictions may vary in medical billing and coding practices, healthcare
20 system patient populations, and representativeness of the general population. Specifically, in
21 Canada, healthcare is delivered through a provincially administered universal healthcare system.
22 As such, research using EHR provides an opportunity to develop methods for population-based,
23 representative estimates of transgender populations within the Canadian context. Coupled with
24 the current absence of gender ascertainment measures in population-based routinely collected
25 data (e.g., census, national government health surveys, etc.) in Canada and many other
26 jurisdictions, this remains an evidence need.

27 **Summary of study rationale**

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

32 **METHODS**

33 **Data Sources and Participants**

34 The Comparative Outcomes and Service Utilization Trends Study (COAST)
35 COAST is a population-based cohort study focused on health services utilization research
36 questions among all people known to be living with HIV (PLWH) in the province of British
37 Columbia (BC) and a 10% random sample comparison group of the HIV-negative general
38 population.[21] The COAST cohort comprises individual-level, longitudinal data from PLWH
39 who have ever accessed HIV treatment in BC between 1996 and 2013, provided by Population
40 Data BC (PopDataBC)[21] via data linkage between two provincial data sources, by personal
41 health number: the Drug Treatment Program (DTP) [22] and the Ministry of Health. PopDataBC
42 provides infrastructure for access to, and linkage of, longitudinal and individual-level
43 administrative health data for all BC residents.[23]. The HIV-negative general population cohort
44 was drawn randomly from the Ministry of Health registry data by PopDataBC. The COAST study
45 has received approval from the University of British Columbia/Providence Health Care
46 Research Ethics Board (#H09-02905) and Simon Fraser University Office of Research Ethics
47 (#2013 s0566). The study complies with the BC Freedom of Information and Protection of

1
2
3 1 Privacy Act (FIPPA) and did not require informed consent as it is conducted using
4 2 retrospective administrative and anonymized data for research and statistical purposes
5 3 only. No patients or public were involved in this study.
6 4

5 Drug Treatment Program

6 In BC, antiretroviral therapy (ART) is provided to PLWH at no cost to the patient, and distributed
7 through the DTP.[22] The DTP contributed a provider-reported measure of transgender status for
8 COAST.
9

10 Ministry of Health

11 Ministry of Health data available via COAST included insured medical service billing records for
12 outpatient visits,[24,25] hospital (in-patient) visits,[26] prescription medications,[27,28] and vital
13 statistics.[29]
14

15 **Measures & Analyses**

16 Transgender computable phenotypes

17 Identification of transgender cases was tested in COAST using International Classification of
18 Disease (ICD) codes (9th and 10th editions) and exogenous sex hormone prescription use.

19 Transgender-specific diagnoses in medical and hospital billing records included the ICD-9 codes
20 302.5 Trans-sexualism with unspecified history, 302.51 Trans-sexualism with asexual history,
21 302.52 Trans-sexualism with homosexual history, 302.53 Trans-sexualism with heterosexual
22 history, 302.6 Gender Identity Disorder in children, 302.85 Gender Identity Disorder in
23 adolescents or adults; and ICD-10 codes F64.0 Gender Identity Disorder of childhood, F64.2
24 Gender Identity Disorder of childhood, F64.8 Other Gender Identity Disorder, and F64.9 Gender
25 Identity Disorder unspecified. The full list of androgen blockers and exogenous sex hormone
26 prescriptions included in analyses is available in the supplementary material.
27

28 Concordance

29 To assess face validity and utility of diagnosis and prescription data over time in CP
30 development (i.e. whether the identified transgender sample had exogenous sex hormone
31 prescription use and other diagnoses patterns consistent with that of transgender populations in
32 other studies), concordance analyses evaluated the presence of at least one included diagnosis
33 and prescription during the COAST study follow-up period with the presence of at least one
34 included diagnosis and prescription in the last study year. Concordance was assessed between
35 transgender-specific diagnoses, exogenous sex hormone and androgen blocker prescriptions, and
36 non-transgender specific diagnoses (ICD-9 259.9 Unspecified Endocrine Disorder and ICD-10
37 E34.9 Endocrine Disorder, Unspecified [see supplementary material]). Endocrine disorder
38 diagnosis codes are sometimes preferred by medical providers treating transgender people in
39 response to historic exclusions of transgender-specific care from insurance coverage and to
40 combat the stigma of transgender-specific diagnosis codes that have historically been classified
41 as psychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM).[30]
42 Exogenous sex hormone use, while common in transgender populations,[5,31] is not
43 transgender-specific. Cisgender populations also use androgen blocker and sex hormone
44 prescriptions (e.g. estrogen to treat menopausal symptoms in cisgender women, spironolactone is
45 used for hypertension), thus exogenous sex hormone and androgen blocker prescription use
56
57
58
59
60

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

39

1 Validation

2 While no one CP consistently performed well across all validation metrics, the CP with the best
 3 overall performance across test statistics was based on having received at least one transgender-
 4 specific diagnosis and at least one androgen blocker/exogenous sex hormone prescription over
 5 the study follow-up period (Table 2). This CP had high specificity (99.8%, 95% CI: 99.6-99.8),
 6 low sensitivity (27.5%, 95% CI: 17.8-39.8), low to moderate Kappa coefficients (0.3, 95% CI:
 7 0.2-0.5) and moderate positive predictive values (43.2%, 95% CI: 28.7-58.9). This CP also had
 8 the second longest mean follow-up time (mean: 136.3, range: 21.0-203.0), similar overall to the
 9 other CP groups (mean: 136.5, range: 21.0-203.0; mean: 117.1, range: 24.0-198.0; mean: 130.4,
 10 range: 69.0-198.0; respectively).

11
 12 **Table 2. Validation measures of transgender computable phenotype (CP) with provider-
 13 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)

14 Transgender phenotype

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

17 Demographics and chronic conditions

18 Demographic characteristics and chronic conditions for the 188 transgender people identified via
 19 the best-performing CP are presented in Figures 2 to 4. Transgender people were geographical
 20 located throughout BC health regions. The Vancouver Coastal Health Authority region, which
 21 includes the largest municipal area in BC, had the highest concentration of transgender people
 22 (44.2%) while the Northern Health Authority region - a predominantly rural and remote area of
 23 the province - had the lowest (1.6%).^[36] The HIV-positive group had a higher median age than
 24 the HIV-negative group (35 [Q1, Q3: 30,42] and 30 [Q1, Q3: 19,42], respectively). For the HIV-
 25 negative sample, the largest proportion of transgender people were aged 19 to 29 years (44.5%)
 26 and the smallest proportion aged 55 years and older (<3%). For the HIV-positive sample, the
 27 largest proportion were aged 30 to 34 years (25.5%) and the smallest proportion aged 55 years
 28 and older (<2%).

1
2
3 1 [Figures 2 and 3 here]
4 2

5 3 Overall, HIV-positive transgender people had a higher prevalence of at least one chronic
6 4 condition (other than HIV) compared to HIV-negative transgender people (88.2% versus 85.4%,
7 5 respectively), and of two or more chronic conditions (76.5% versus 52.6%, respectively).
8 6 Specific chronic disease differences between transgender people living with and without HIV
9 7 were most notable for a higher prevalence among the HIV-positive cohort of cardiovascular
10 8 disease, chronic kidney disease, osteoarthritis, schizophrenia and personality disorders, and
11 9 chronic liver disease, but a lower prevalence for hypertension.
12
13

14 10
15 11 [Figure 4 here]
16 12

17 13 **DISCUSSION**

18 14 This study demonstrates the feasibility of identification of a sample of transgender people in a
19 15 large linked provincial healthcare administrative database, using a CP based on prescriptions and
20 16 diagnoses. Among a growing number of studies using EHR and CP methods to identify
21 17 transgender samples for health research purposes, this is the first to do so in Canada., to
22 18 independently validate the CP using a ‘gold standard’ of provider-reported transgender status,
23 19 and the only to use population-based data.
24
25

26 21 **Concordance**

27 22 There was high concordance between transgender-specific diagnoses and exogenous sex
28 23 hormone or androgen blocker prescription use in this study. That nearly half of those with at
29 24 least one transgender-specific diagnosis had been dispensed hormones or blockers in the past
30 25 year is consistent with findings from US and Canadian studies (48.9% and 43.0%,
31 26 respectively)[20,32,33] - suggesting face validity for the current CP.
32
33

34 28 **CP development and validation**

35 29 The best-performing CP overall successfully identified cisgender people who were truly
36 30 cisgender (specificity) and correctly identified transgender people who were truly transgender
37 31 (0.2% false positive rate, results not shown). However, the selected CP had relatively low
38 32 sensitivity, missing approximately 72.5% of ‘true’ transgender people in COAST, as identified
39 33 by the gold standard provider-based measure. Though a relatively small proportion of the ‘true’
40 34 transgender sample was identified in this study, the impact on future analyses comparing health
41 35 outcomes for transgender and cisgender groups is likely negligible, as even the large proportion
42 36 of ‘true’ transgender people misclassified as cisgender (approximate n=496) is a very small
43 37 proportion of the total COAST sample. At worst, this misclassification would bias results related
44 38 to disparities between transgender and cisgender health toward the null, producing a conservative
45 39 attenuated effect in COAST, and other such administrative datasets. Further, as discussed below,
46 40 gender identity classification will likely greatly improve as transgender care shifts further into
47 41 the fee-for-service system in BC. As in other Canadian administrative data studies, low
48 42 sensitivity may be explained in part by provider and system billing preferences using 3-digit ICD
49 43 diagnosis coding instead of the more specific 4-digit coding, and inconsistencies in the BC
50 44 billing management system.[37] Despite the low sensitivity, CP development in this study with
51 45 high specificity offers an advancement for transgender health research. A measure that correctly
52 46 identifies cases for transgender samples in research with good success translates to better
53
54
55
56
57
58
59
60

1
2
3 1 opportunities to include transgender people in health studies and to investigate their health
4 2 relative to other groups. While future research may lead to improvements in CP development, the
5 3 CP identified in the current study with good specificity, albeit relatively poor sensitivity, has
6 4 important utility in advancing opportunities in transgender health research. Additionally, while
7 5 differential follow-up time can affect algorithm performance, the similar mean and range follow-
8 6 up time for all CPs in this study suggests that differential follow-up time was not an important
9 7 source of bias in this study.
10 8

11 9
12 10 The limited agreement between the CP and provider-report transgender status may be due to the
13 11 widely varying transgender status prevalence depending on study design and ascertainment
14 12 measures used.[14] In the BC context, the CP and the DTP measures are assessing transgender
15 13 status in different ways and for different purposes. In the DTP, transgender status is ascertained
16 14 in the context of HIV diagnosis and ART prescribing, during which demographics and HIV
17 15 transmission risk factors are recorded. This differs from recording diagnoses in EHR for those
18 16 accessing transgender-specific care as utilized in the CP. This may explain the lower PPV for the
19 17 best-performing CP compared to the CP based on recent transgender diagnoses, suggesting the
20 18 DTP provider-reported transgender status measure has better coverage for recent cases and the
21 19 potential for use of recent diagnosis over ever to be beneficial in future CP development.
22 20 Ultimately, a single CP may not be sufficient for all intended purposes and the best applicable
23 21 CP (using different types of diagnoses, prescriptions or procedures) may differ depending on the
24 22 intended healthcare, health research, or health policy application.[17]
25 23
26 24

27 25
28 26 There is limited literature on EHR-based studies with the ability to validate an administrative
29 27 transgender measure using a ‘gold standard’ comparison measure.[16] The two previous studies
30 28 that have developed and validated algorithms to identify transgender individuals have both been
31 29 conducted in non-representative samples in the US, one using Medicare data[38] and one in a
32 30 university medical center.[16] Similar to the current study, the Medicare study found high
33 31 specificity when comparing an EHR-based and a two-step survey-based transgender measure.
34 32 However, the Medicare study found that the EHR measure performed consistently well with high
35 33 sensitivity and a high Kappa statistic, unlike in the current study. Using chart review as the ‘gold
36 34 standard’ for comparison of transgender status, Ehrenfeld et al. found a low false positive rate for
37 35 their best-performing algorithm (3%), though not as low as the false positive rate in the current
38 36 study. The overall high levels of agreement for transgender measures in the two previous studies
39 37 is likely a function of the lack of independence between the ‘gold standard’ and the CP or
40 38 algorithm measures. Specifically, only those classified as transgender in the Medicare EHR data
41 39 were offered survey participation to complete the two-step ‘gold standard’ survey measure, and
42 40 only those cases identified as transgender in the university clinic EHR were included in chart
43 41 review. Thus, previous studies could assess agreement between the two measures, but not
44 42 robustly validate either. In the current study, the DTP provider-based transgender status measure
45 43 is independent and thus could be used for robust CP validation.
46 44
47 45
48 46
49 47
50 48
51 49
52 50
53 51
54 52
55 53
56 54
57 55
58 56
59 57
60 58

59 59 While not possible to incorporate free-text records in case-finding algorithms in the current study
60 60 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

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

7 **Transgender status prevalence & ascertainment**

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

20
21 Using the broadest CP algorithm (any transgender-specific diagnosis ever, n=56) and those
22 identified by provider-report together (total n=106), the total transgender PLWH sample would
23 represent as high as 0.88% (range: 0.73-1.1%) of the prevalent HIV infections in BC in
24 2014.[40] This overrepresentation of transgender people among PLWH is consistent with
25 evidence of a disproportionate HIV burden for transgender populations globally,[5,41,42] as well
26 as in line with the only other available data on the proportion of PLWH who are transgender,
27 from US national surveillance data (2012 data: 1.1%, 95%CI: 0.8-1.4).[43]

29 **Demographics and chronic conditions**

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

43 **Limitations**

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

1
2
3 1 accessing medical transition care in BC and may only represent 24% to 47% of the total
4 2 transgender population.[33] This study was also limited by the inability to validate the
5 3 transgender CP among the HIV-negative COAST cohort, as a 'gold standard' provider-based
6 4 transgender measure was only available for the HIV-positive cohort. It is possible that the
7 5 transgender CPs would perform differently in populations living without HIV, particularly as
8 6 healthcare contact is higher among populations living with HIV. Additionally, this study should
9 7 be considered in light of the context in which it was conducted, an environment in which
10 8 transgender healthcare delivery in BC is currently shifting from specialized care settings to the
11 9 main primary care fee-for-service settings. Given that COAST only includes fee-for-service data,
12 10 this study was limited by the inability to capture transgender people who access transgender care
13 11 outside the fee-for-service system. However, fortunately, as the shift to the fee-for-service
14 12 system occurs, transgender ascertainment via CPs in BC will likely improve. The administrative
15 13 data used in this study may also be susceptible to coding error (and coding biases/practices)
16 14 across conditions and settings,[49] potentially introducing misclassification bias in terms of
17 15 transgender ascertainment. Finally, chronic condition prevalence data reported in this study
18 16 should be interpreted with caution, given potential selection bias by serostatus in the COAST
19 17 cohort; though any such bias likely resulted in conservative estimates of difference by serostatus
20 18 in this analysis.
21
22
23
24

25 20 CONCLUSION

26 21 This study makes a number of important contributions to the literature on innovative methods in
27 22 transgender health. Major contributions include development and validation of a transgender CP,
28 23 using a population-based representative source population, in the Canadian context. Another
29 24 strength is the approximately complete ascertainment of the population of transgender PLWH in
30 25 BC, and capacity to estimate transgender status prevalence among PLWH. In a current funding
31 26 environment of limited support for longitudinal transgender health studies in the US and none to
32 27 date in Canada, this study and the methods employed offer an efficient, replicable and cost-
33 28 effective way forward in creating electronic cohorts for advancing transgender health
34 29 research.[15] Moreover, the recent rollback of sexual orientation and gender identity data
35 30 collection and legal changes in insurance coverage of transgender healthcare in the US potentiate
36 31 decline in accurate claims coding for gender-affirming care.[30] This highlights the utility of
37 32 work in this area from other jurisdictions, particularly those with transgender-inclusive universal
38 33 healthcare systems such as Canada.
39
40
41

42 35 Future research should build upon the methods developed in this study and explore
43 36 complimentary approaches for gender identity ascertainment in administrative and EHR data,
44 37 such as machine learning approaches, as have been used to develop algorithms based on
45 38 healthcare utilization data in other research areas. Finally, the current study lays the foundation
46 39 for future work with the ability to study transgender health and healthcare use patterns over time,
47 40 with linkage to laboratory data, as well as inclusion of appropriate comparison groups.[15,50]
48
49

50 42 ACKNOWLEDGEMENTS & DISCLAIMER

51 43 The authors thank the COAST study participants, the BC Centre for Excellence in HIV/AIDS,
52 44 the BC Ministry of Health, BC Vital Statistics Agency, PharmaNet and the institutional data
53 45 stewards for granting access to the data, and Population Data BC for facilitating the data linkage
54 46 process. In addition, we would like to thank the COAST core team members and other support
55
56
57
58
59
60

1 staff at these institutions for their administrative assistance with the data access and preparation.
2 All inferences, opinions, and conclusions drawn in this paper are those of the authors, and do not
3 reflect the opinions or policies of the Data Steward(s). There are no conflicts of interest to
4 declare.
5

6 **COMPETING INTERESTS**

7 None declared.
8

9 **FUNDING**

10 This work was supported by the Canadian Institutes of Health Research, through an Operation
11 Grant [grant number 130419], a Foundation Award to RSH [grant number 143342], and a
12 Doctoral Research Award to AJR [grant number 152382] and support from the British Columbia
13 Centre for Excellence in HIV/AIDS. The DTP receives funding from the provincial government
14 of British Columbia (PharmaCare). The funders had no role in the study design, analysis,
15 interpretation of the data, drafting of the manuscript or in the decision to submit for
16 publication.
17

18 **CONTRIBUTORS**

19 AJR led the study from conceptualization to analysis plan to interpretation, drafting of the first
20 manuscript version, revisions and final version. RSH acquired study data and funding. TP, MK,
21 PS, TS, and RSH all contributed to study design, interpretation of results, and reviewed manuscript
22 versions. JL and MY contributed to study analysis and reviewed manuscript versions. All authors
23 provided critical review of first and subsequent manuscript drafts, approved the final version, and
24 agree to be accountable for the work presented.
25

26 **PATIENT AND PUBLIC INVOLVEMENT**

27 No patients involved.
28

29 **DATA SHARING STATEMENT**

30 The data used for this study are held by the BC Centre for Excellence in HIV/AIDS under the
31 authority of the BC Ministry of Health; as they contain confidential patient health records
32 including HIV serostatus, data cannot be made available to other parties.
33
34
35
36
37

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **FIGURES LEGENDS**

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

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

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

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

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

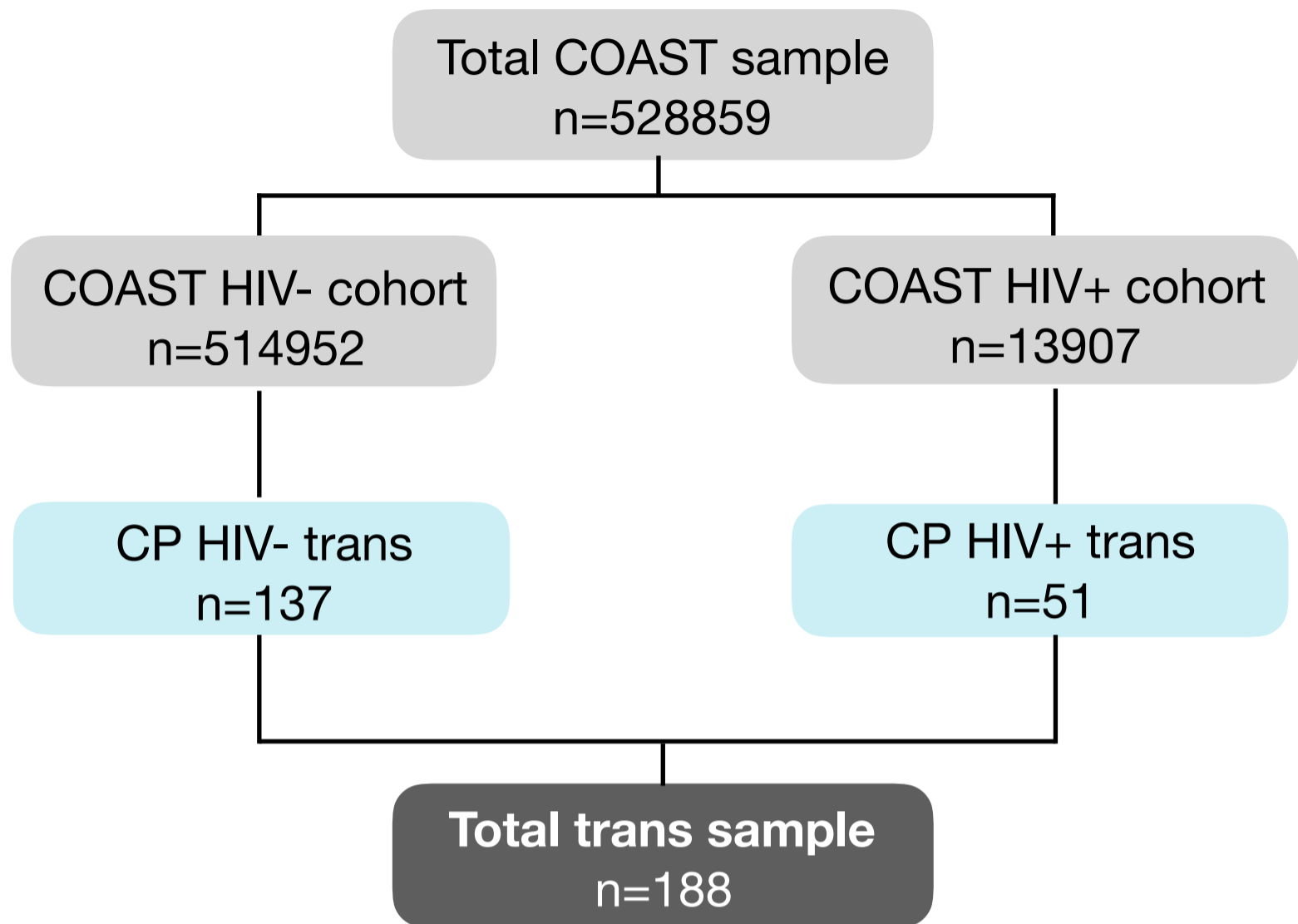
1 REFERENCES

- 2 1. Bauer GR. Making sure everyone counts: Considerations for inclusion, identification, and
3 analysis of transgender and transsexual participants in health surveys. *What Differ Sex Gen*
4 *Make Gen Sex Health Res Casebook*. Vancouver, Canada: Canadian Institutes of Health
5 Research (CIHR), Institute for Gender and Health (IGH); 2012. p. 59–67.
- 6 2. Bauer GR, Hammond R, Travers R, Kaay M, Hohenadel KM, Boyce M. “I don’t think this is
7 theoretical; this is our lives”: How erasure impacts health care for transgender people. *JANAC*.
8 2009;20:348–361.
- 9 3. Poteat T, German D, Flynn C. The conflation of gender and sex: Gaps and opportunities in
10 HIV data among transgender women and MSM. *Glob Public Health*. 2016;11:835–848.
- 11 4. Wanta JW, Unger CA. Review of the transgender literature: Where do we go from here?
12 *Transgender Health*. 2017;2:119–128.
- 13 5. Herbst JH, Jacobs ED, Finlayson TJ, McKleroy VS, Neumann MS, Crepaz N, et al.
14 Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a
15 systematic review. *AIDS Behav*. 2008;12:1–17.
- 16 6. Becasen JS, Denard CL, Mullins MM, Higa DH, Sipe TA. Estimating the prevalence of HIV
17 and sexual behaviors among the US transgender population: A systematic review and meta-
18 analysis, 2006–2017. *Am J Public Health*. 2019;109:e1–8.
- 19 7. Marshall Z, Welch V, Thomas J, Brunger F, Swab M, Shemilt I, et al. Documenting research
20 with transgender and gender diverse people: protocol for an evidence map and thematic analysis.
21 *Syst Rev*. 2017;6:35.
- 22 8. Reisner SL, Poteat T, Keatley J, Cabral M, Mothopeng T, Dunham E, et al. Global health
23 burden and needs of transgender populations: a review. *The Lancet*. 2016;388:412–436.
- 24 9. Cathcart-Rake EJ, Lightner DJ, Quevedo FJ, Ertz DC, Jatoi A. Cancer in transgender patients:
25 One case in 385,820 is indicative of a paucity of data. *J Oncol Pract*. 2017;JOP2017027714.
- 26 10. Meerwijk EL, Sevelius JM. Transgender population size in the United States: A meta-
27 regression of population-based probability samples. *Am J Public Health*. 2017;107:e1–e8.
- 28 11. Flores AR, Herman JL, Gates GJ, Brown TNT. How Many Adults Identify as Transgender in
29 the United States? [Internet]. Los Angeles, CA: Williams Institute; 2016 Jun p. 13. Available
30 from: [https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-](https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf)
31 [Transgender-in-the-United-States.pdf](https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf)
- 32 12. The GenIUSS Group. Best Practices for Asking Questions to Identify Transgender and Other
33 Gender Minority Respondents on Population-Based Surveys. Los Angeles, CA: The Williams
34 Institute; 2014.

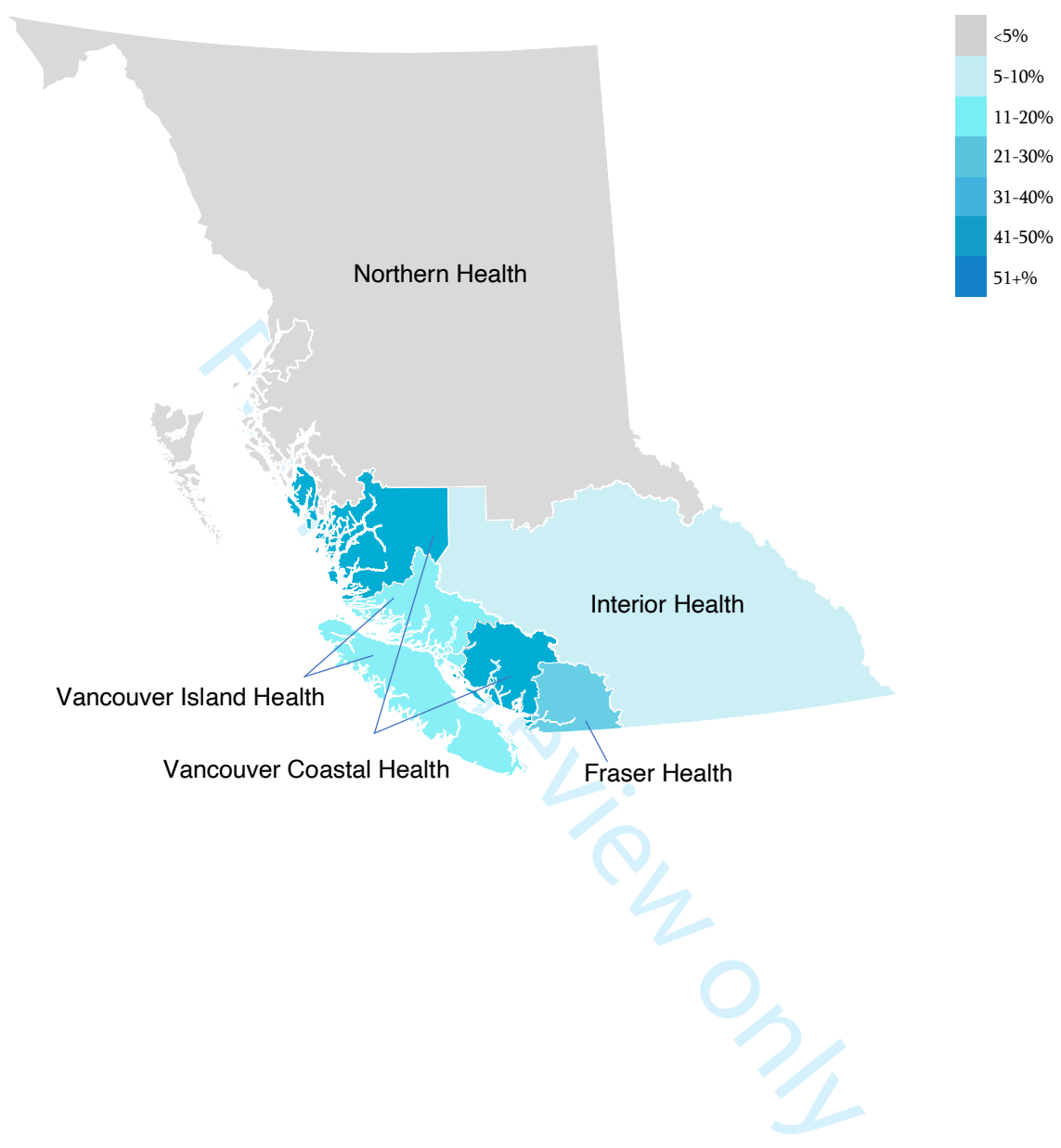
13. Bauer GR, Braimoh J, Scheim AI, Dharma C. Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. *PLOS ONE*. 2017;12:e0178043.
14. Collin L, Reisner SL, Tangpricha V, Goodman M. Prevalence of transgender depends on the “case” definition: 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 Textbook 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. HIV/AIDS. 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>

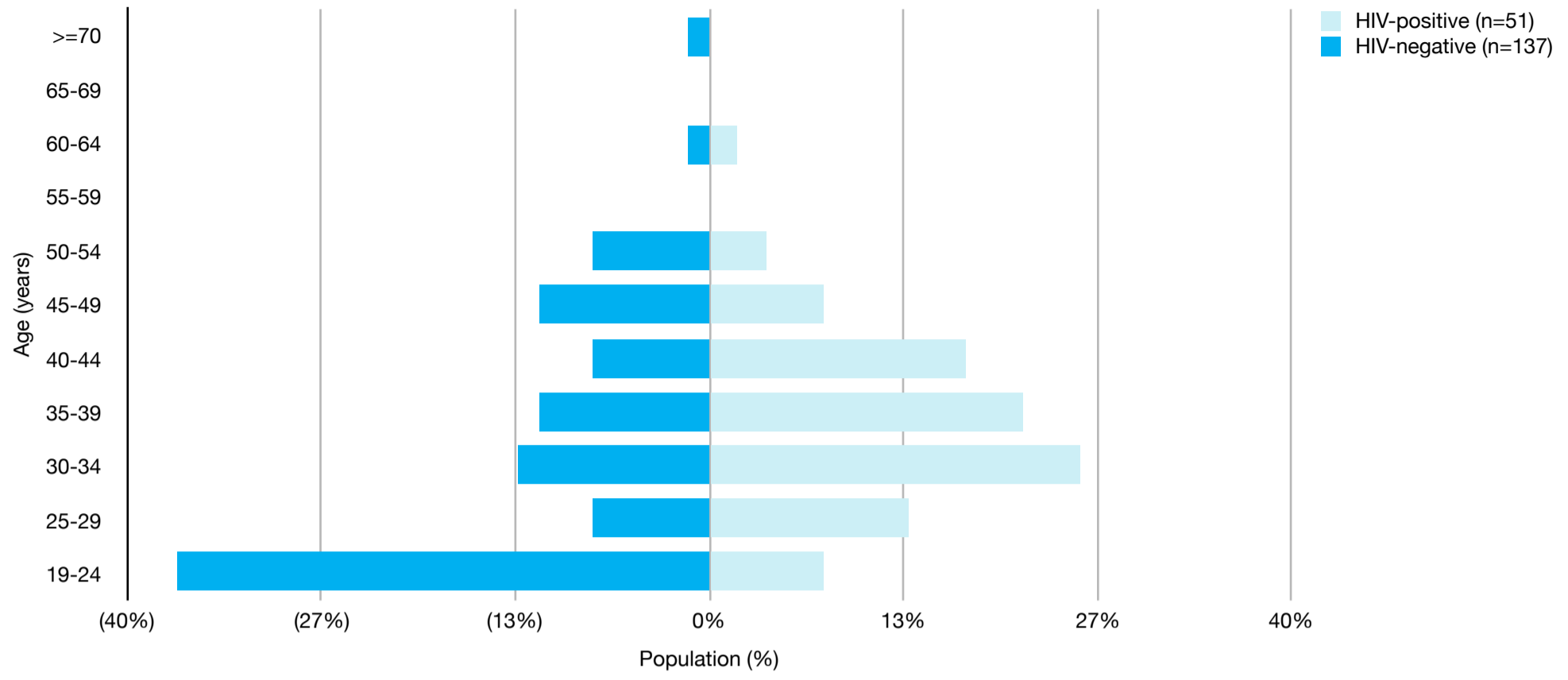
- 1
2
3 1 26. Canadian Institute for Health Information (2014). Discharge Abstract Database (Hospital
4 2 Separations) [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from:
5 3 <http://www.popdata.bc.ca/data>
6
7
8 4 27. British Columbia Ministry of Health (2014). PharmaCare [Internet]. Popul. Data BC. 2014
9 5 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca/data>
10
11 6 28. British Columbia Ministry of Health (2014). PharmaNet [Internet]. Popul. Data BC. 2014
12 7 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca/data>
13
14 8 29. British Columbia Vital Statistics Agency (2014). Vital Statistics Deaths [Internet]. Popul.
15 9 Data BC. 2014 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca>
16
17
18 10 30. Arnold JD, Nelson AE, Loubier EM. Trends in Insurance Coverage for Gender-Affirming
19 11 Surgeries. *JAMA Surg.* 2018;153:972.
20
21 12 31. Radix A, Sevelius J, Deutsch MB. Transgender women, hormonal therapy and HIV
22 13 treatment: a comprehensive review of the literature and recommendations for best practices. *J Int*
23 14 *AIDS Soc.* 2016;19:20810.
24
25 15 32. Rotondi NK, Bauer GR, Scanlon K, Kaay M, Travers R, Travers A. Nonprescribed hormone
26 16 use and self-performed surgeries: “do-it-yourself” transitions in transgender communities in
27 17 Ontario, Canada. *Am J Public Health.* 2013;103:1830–1836.
28
29
30 18 33. Scheim AI, Bauer GR. Sex and gender diversity among transgender persons in Ontario,
31 19 Canada: results from a respondent-driven sampling survey. *J Sex Res.* 2015;52:1–14.
32
33 20 34. Chronic Disease Information Working Group. BC Chronic Disease and Selected Procedure
34 21 Case Definitions [Internet]. British Columbia, Canada: BC Ministry of Health; 2019 Apr p. 1–
35 22 132. Available from: [https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-](https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf)
36 23 [access/chronic-disease-registries-case-definitions.pdf](https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf)
37
38
39 24 35. British Columbia Cancer Agency. BC Cancer Registry Data [Internet]. Popul. Data BC. 2014
40 25 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca/data>
41
42 26 36. BC Stats. Sub-provincial Population Estimates [Internet]. 2018 [cited 2019 Feb 1]. Available
43 27 from: <file://localhost/Users/AJRich/Zotero/storage/5VM2LBUN/PopulationEstimates.html>
44
45 28 37. Hu W. Diagnostic Codes in MSP Claim Data. Program Monitoring and Information
46 29 Management Branch, Resource Management Division, Medical Services Plan; 1996 p. 39.
47
48 30 38. Gerth J, Becerra-Culqui T, Bradlyn A, Getahun D, Hunkeler EM, Lash TL, et al. Agreement
49 31 between medical records and self-reports: Implications for transgender health research. *Rev*
50 32 *Endocr Metab Disord.* 2018;19:263–269.
51
52
53 33 39. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al.
54 34 Development of phenotype algorithms using electronic medical records and incorporating natural
55
56
57
58
59
60

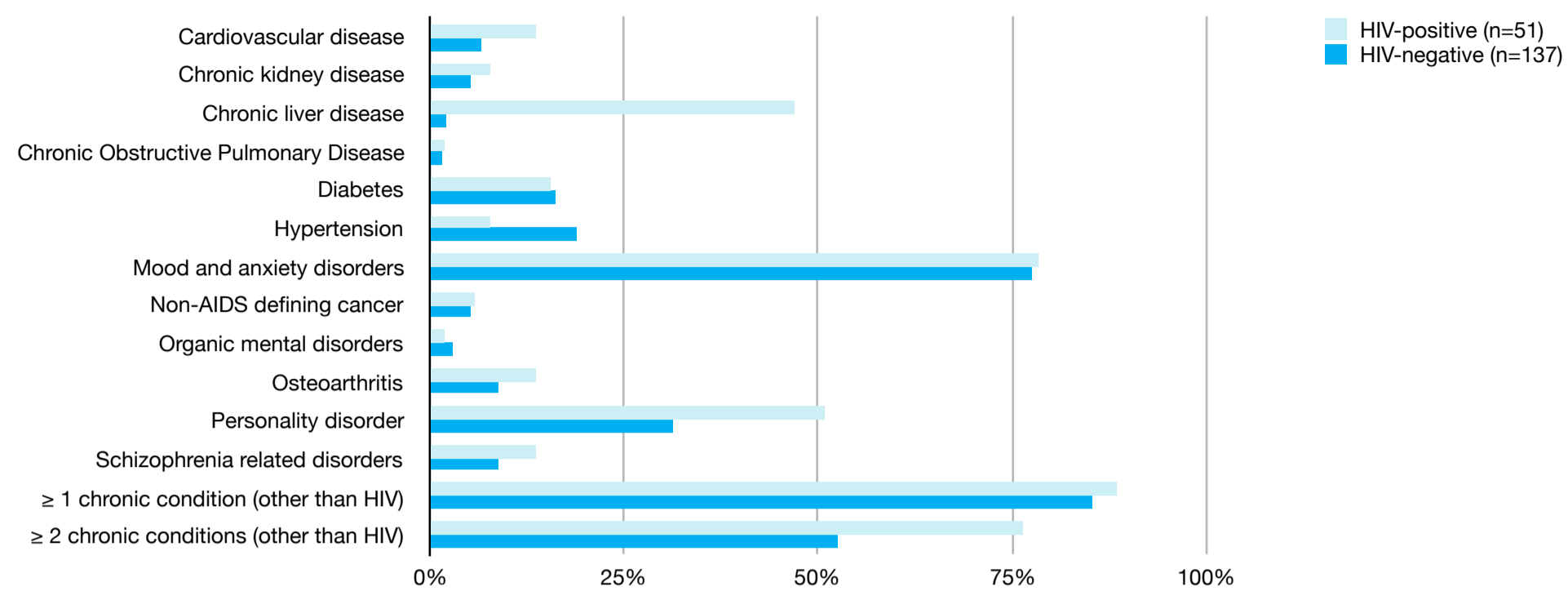
- 1
2
3 1 language processing. The BMJ [Internet]. 2015 [cited 2020 Oct 22];350. Available from:
4 2 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4707569/>
5
6 3 40. BC Centre for Disease Control. HIV in British Columbia: Annual Surveillance Report 2015.
7 4 Vancouver, Canada: BC Centre for Disease Control; 2017.
8
9 5 41. Poteat T, Scheim A, Xavier J, Reisner S, Baral S. Global epidemiology of HIV infection and
10 6 related syndemics affecting transgender people. *J Acquir Immune Defic Syndr*. 2016;72 Suppl
11 7 3:S210–9.
12
13 8 42. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of
14 9 HIV in transgender women: a systematic review and meta-analysis. *Lancet Infect Dis*.
15 10 2013;13:214–222.
16
17 11 43. Centers for Disease Control and Prevention. Behavioral and Clinical Characteristics of
18 12 Persons Receiving Medical Care for HIV Infection—Medical Monitoring Project, United States,
19 13 2012 [Internet]. Washington, D.C.: Centers for Disease Control; 2015 Aug p. 42. Report No.: 12.
20 14 Available from: [https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-](https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf)
21 15 [HSSR_MMP_2012.pdf](https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf)
22
23 16 44. Grabovac I, Veronese N, Stefanac S, Haider S, Jackson SE, Koyanagi A, et al. Human
24 17 Immunodeficiency Virus Infection and Diverse Physical Health Outcomes: An Umbrella Review
25 18 of Meta-analyses of Observational Studies. *Clin Infect Dis*. 2019;ciz539.
26
27 19 45. Nasi M, De Biasi S, Gibellini L, Bianchini E, Pecorini S, Bacca V, et al. Ageing and
28 20 inflammation in patients with HIV infection: Ageing, inflammation and HIV infection. *Clin Exp*
29 21 *Immunol*. 2017;187:44–52.
30
31 22 46. Judge C., O’Donovan C., Callaghan G., Gaoatswe G., O’Shea D. Gender dysphoria -
32 23 prevalence and co-morbidities in an Irish adult population. *Front Endocrinol*. 2014;5:87.
33
34 24 47. Cheung AS, Ooi O, Leemaqz S, Cundill P, Silberstein N, Bretherton I, et al.
35 25 Sociodemographic and clinical characteristics of transgender adults in Australia. *Transgender*
36 26 *Health*. 2018;3:229–38.
37
38 27 48. Swartz JA, Ducheny K, Holloway T, Stokes L, Willis S, Kuhns LM. A latent class analysis
39 28 of chronic health conditions among HIV-Positive transgender women of color. *AIDS Behav*
40 29 [Internet]. 2019 [cited 2019 Oct 2]; Available from: [http://link.springer.com/10.1007/s10461-](http://link.springer.com/10.1007/s10461-019-02543-3)
41 30 [019-02543-3](http://link.springer.com/10.1007/s10461-019-02543-3)
42
43 31 49. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy
44 32 of Medicare’s hospital claims data: progress has been made, but problems remain. *Am J Public*
45 33 *Health*. 2011;
46
47 34 50. Reisner SL, Radix A, Deutsch MB. Integrated and gender-affirming transgender clinical care
48 35 and research. *J Acquir Immune Defic Syndr*. 2016;72 Suppl 3:S235–42.
49
50
51
52
53
54
55
56
57
58
59
60



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60







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		
Spirolactone		
	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

1		
2		
3		
4	ETHYNODIOL D-ETHINYL ESTRADIOL	28630
5	ETHYNODIOL D-ETHINYL ESTRADIOL	469327
6	NORETHINDRONE-MESTRANOL	22608
7	NORETHINDRONE-MESTRANOL	22659
8		
9	NORETHINDRONE A-E ESTRADIOL	297143
10	NORETHINDRONE A-E ESTRADIOL	315966
11	NORETHINDRONE-ETHINYL ESTRAD	317047
12	NORETHINDRONE-ETHINYL ESTRAD	372846
13	NORETHINDRONE-ETHINYL ESTRAD	373265
14	NORETHINDRONE-ETHINYL ESTRAD	531006
15	NORETHINDRONE-ETHINYL ESTRAD	538590
16	NORETHINDRONE-ETHINYL ESTRAD	602957
17	NORETHINDRONE-ETHINYL ESTRAD	620947
18	NORETHINDRONE-ETHINYL ESTRAD	2187086
19	NORETHINDRONE-ETHINYL ESTRAD	2187108
20	NORETHINDRONE-ETHINYL ESTRAD	2189054
21	NORETHINDRONE-ETHINYL ESTRAD	34207
22	NORGESTREL-ETHINYL ESTRADIOL	300640
23	NORGESTREL-ETHINYL ESTRADIOL	579386
24	LEVONORGESTREL-ETH ESTRA	707600
25	LEVONORGESTREL-ETH ESTRA	782416
26	LEVONORGESTREL-ETH ESTRA	782432
27	LEVONORGESTREL-ETH ESTRA	2042320
28	LEVONORGESTREL-ETH ESTRA	2043033
29	NORGESTREL-ETHINYL ESTRADIOL	2043726
30	LEVONORGESTREL-ETH ESTRA	2258560
31	NORGESTIMATE-ETHINYL ESTRADIOL	30333
32	NORETHINDRONE-MESTRANOL	30341
33	NORETHINDRONE-MESTRANOL	2236974
34	LEVONORGESTREL-ETH ESTRA	471526
35	ETHYNODIOL D-ETHINYL ESTRADIOL	340731
36	NORETHINDRONE-ETHINYL ESTRAD	340758
37	NORETHINDRONE-MESTRANOL	343838
38	NORETHINDRONE A-E ESTRADIOL	343838
39	NORETHINDRONE A-E ESTRADIOL	353027
40	NORETHINDRONE-ETHINYL ESTRAD	372838
41	NORETHINDRONE-ETHINYL ESTRAD	373273
42	NORETHINDRONE-ETHINYL ESTRAD	531014
43	NORETHINDRONE-ETHINYL ESTRAD	602965
44	NORETHINDRONE-ETHINYL ESTRAD	695734
45	NORETHINDRONE-ETHINYL ESTRAD	2187094
46		
47		
48		
49		
50		
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		

Supplementary material

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1		
2		
3	MEDROXYPROGESTERONE ACET	2246627
4	MEDROXYPROGESTERONE ACET	30937
5	MEDROXYPROGESTERONE ACET	2010739
6	MEDROXYPROGESTERONE ACET	2148560
7	MEDROXYPROGESTERONE ACET	2221292
8	MEDROXYPROGESTERONE ACET	2229839
9	MEDROXYPROGESTERONE ACET	2244727
10	MEDROXYPROGESTERONE ACET	2246628
11	MEDROXYPROGESTERONE ACET	729973
12	MEDROXYPROGESTERONE ACET	2010933
13	MEDROXYPROGESTERONE ACET	2148579
14	MEDROXYPROGESTERONE ACET	2221306
15	MEDROXYPROGESTERONE ACET	2229840
16	MEDROXYPROGESTERONE ACET	2246629
17	MEDROXYPROGESTERONE ACET	30945
18	MEDROXYPROGESTERONE ACET	2267640
19	MEDROXYPROGESTERONE ACET	2267640
20	MEDROXYPROGESTERONE ACET	2267640
21	MEDROXYPROGESTERONE ACET	2267640
22	MEDROXYPROGESTERONE ACET	2267640
23	MEDROXYPROGESTERONE ACET	2267640
24	MEDROXYPROGESTERONE ACET	2267640
25	MEDROXYPROGESTERONE ACET	2267640
26	MEDROXYPROGESTERONE ACET	2267640
27	MEDROXYPROGESTERONE ACET	2267640
28	MEDROXYPROGESTERONE ACET	2267640
29	MEDROXYPROGESTERONE ACET	2267640
30	MEDROXYPROGESTERONE ACET	2267640
31	MEDROXYPROGESTERONE ACET	2267640
32	MEDROXYPROGESTERONE ACET	2267640
33	Transmasculine	
34	Testosterone	
35	TESTOSTERONE	2249499
36	TESTOSTERONE CYPIONATE	30783
37	TESTOSTERONE PROPIONATE	1977571
38	TESTOSTERONE CYPIONATE	1977601
39	TESTOSTERONE CYPIONATE	2220318
40	TESTOSTERONE CYPIONATE	2246063
41	TESTOSTERONE ENANTHATE	29246
42	TESTOSTERONE ENANTHATE	716936
43	TESTOSTERONE ENANTHATE	739944
44	TESTOSTERONE UNDECANOATE	782327
45	TESTOSTERONE ENANTHATE/ESTRAD	108278
46	TESTOSTERONE ENANTHATE/ESTRAD	2061031
47	TESTOSTERONE	2239653
48	TESTOSTERONE	2245346
49	TESTOSTERONE	2245345
50	TESTOSTERONE	2245972
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		

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
	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-9: 410 Acute myocardial infarction 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
	Chronic heart failure	1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	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 ICD-10: I50 Heart failure
	Stroke- hospital	1 or more hospitalizations with relevant ICD codes	ICD-9: 428 Heart failure 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

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

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

Chronic Obstructive
Pulmonary Disease*

1 or more
hospitalization or 2 or
more medical visits
within 365 days

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.

Supplementary material

Hypertension*

Excludes gestational diabetes.

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
*Cancer case definition details
available from the British
Columbia Cancer Agency:
[http://www.bccancer.bc.ca/health-
info/types-of-cancer](http://www.bccancer.bc.ca/health-info/types-of-cancer)*

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)

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 Amnesic 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

1			
2			
3			
4			
5			ICD-9:
6			715 Osteoarthritis and allied
7			disorders
8	Personality disorder	1 or more	ICD-9:
9		hospitalizations or	301.x Personality disorders
10		medical visits with a	
11		relevant diagnosis	ICD-10:
12		within 365 days	F60.x Specified personality
13			disorders
14			F62 Enduring personality
15			changes, not attributable to brain
16			damage and disease
17			F68.1 Intentional production or
18			feigning of symptoms or
19			disabilities, either physical or
20			psychological
21			F68.8 Other specified disorders or
22			adult personality and behaviour
23			F69 Unspecified disorder or adult
24			personality and behaviour
25			ICD-9:
26			295.x Schizophrenic disorders
27	Schizophrenia related	1 or more medical visit	297.0 Paranoid state, simple
28	disorder	or hospitalizations with	297.1 Delusional disorder
29		relevant diagnoses	297.2 Paraphrenia
30		within 365 days	297.3 Shared psychotic disorder
31			
32			ICD-10:
33			F20.x Paranoid schizophrenia
34			F21.x Schizotypal disorder
35			F23.2 Acute schizophrenia-like
36			psychotic disorder
37			F25.x Schizoaffective disorders
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			

* 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)

<p>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27</p> <p>Participants</p>	<p>6</p>	<p>(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> - 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 <i>Cross-sectional study</i> - Give the eligibility criteria, and the sources and methods of selection of participants</p> <p>(b) <i>Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> - For matched studies, give matching criteria and the number of controls per case</p>		<p>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.</p> <p>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.</p> <p>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.</p>	<p>6.1: Methods (pp 6-7); Supplementary Material</p> <p>6.2: Methods (pp 6-7)</p> <p>6.3: Figure 1</p>
<p>28 29 30 31 32 33 34</p> <p>Variables</p>	<p>7</p>	<p>Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.</p>		<p>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.</p>	<p>7.1: Methods (pp 6-7); Supplementary Material</p>
<p>35 36 37 38 39 40 41 42</p> <p>Data sources/ measurement</p>	<p>8</p>	<p>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</p>			<p>Methods (pp 6-7)</p>

1 2 3 4 5 6 7 8 9 10	Bias	9	Describe any efforts to address potential sources of bias			Discussion (pp 9, 11)
11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34	Study size	10	Explain how the study size was arrived at			Methods (page 6); Figure 1
35 36 37 38 39 40 41 42 43 44 45 46 47	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) <i>Cohort study</i> - If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> - If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> - 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 (<i>e.g.</i> , 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>i.e.</i> , 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 (<i>e.g.</i> , 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) <i>Cohort study</i> - summarise follow-up time (<i>e.g.</i> , average and total amount)			Results (page 8)
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time <i>Case-control study</i> - Report numbers in each exposure			Results

		category, or summary measures of exposure <i>Cross-sectional study</i> - 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, Guttman 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.

*Checklist is protected under Creative Commons Attribution ([CC BY](https://creativecommons.org/licenses/by/4.0/)) license.

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:	<i>BMJ Open</i>
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

SCHOLARONE™
Manuscripts



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.

1
2
3 **1 TITLE PAGE**
4 2

5 3 **Title:** Development of a computable phenotype to identify a transgender sample for health
6 4 research purposes: A feasibility study in a large linked provincial healthcare
7 5 administrative cohort in British Columbia, Canada
8 6

9 7 **Authors:** Rich AJ^{1,2}, Poteat T³, Koehoorn M¹, Li J², Ye M², Sereda, P², Salway T⁴, Hogg RS^{2,4}
10 8
11 9

12 **Affiliations:**

- 13 10 1. School of Population and Public Health, University of British Columbia, Vancouver, Canada
14 11 2. British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada
15 12 3. Department of Social Medicine, University of North Carolina- Chapel Hill, Chapel Hill, USA
16 13 4. Faculty of Health Sciences, Simon Fraser University, Burnaby, Canada
17 14

18 15 **Corresponding author:**

19 16 Ashleigh J Rich
20 17 ajrich@mail.ubc.ca
21 18

22 19 2206 East Mall
23 20 Vancouver, BC
24 21 V6T 1Z3
25 22 Canada
26 23
27 24
28 25
29 26
30 27
31 28
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 ABSTRACT

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 1 healthcare setting, using a computable phenotype, and capacity to estimate the prevalence
4 2 of transgender status among the population living with HIV in the province.

- 5 3
6 4
7 4
8 5
9 6
10 7
11 8
12 8
13 9
14 10
15 11
16 12
- 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.

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1 INTRODUCTION

2 Limited data on transgender people

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

15
16 The erasures or exclusions of transgender persons in health studies may be explained, in part, by
17 methodological challenges. Specific to electronic health record (EHR) data, a 2017 report
18 identified only one transgender person among 38,5820 cancer cases in a Minnesota cancer
19 registry,[9] clearly an undercount given that 0.4% of the US general population and 0.6% (95%
20 credibility intervals: 0.5%-0.7%) of the Minnesota population is estimated to be
21 transgender.[10,11] This highlights the need for improved gender ascertainment and transgender
22 inclusion in research relying on patient records and administrative data. The establishment of
23 best practices for measuring transgender status in survey research, such as the two-step method
24 (measuring sex assigned at birth and current gender identity), points to a way forward for
25 transgender-inclusive population health research.[12,13] However, innovative research methods
26 are needed to identify transgender people in studies that rely on existing data sources (in
27 particular EHR) and that optimize the use of transgender respondents' data in non-transgender
28 specific research.

29 Computable phenotypes for transgender health research

30 Previous research in transgender health largely comprises cross-sectional studies, case reports,
31 and qualitative or observational research.[7] Much consists of clinic- or venue-based
32 convenience samples or lack comparison groups.[7,8] The literature is further characterized by
33 inconsistent transgender status measurement,[14] small sample sizes, and focus on the United
34 States (US).[8] In response, researchers have called for advancing transgender health research
35 methods - namely ascertainment of high-quality samples via systematic approaches - including
36 for general population-based and health systems-based studies.[15] One opportunity for the
37 advancement of transgender health research methods is the emerging use of computable
38 phenotypes (CPs)[16] or case ascertainment algorithms, to identify transgender samples in
39 healthcare utilization data. A computable phenotype is an algorithm for identifying a clinical
40 feature, condition, or set of characteristics that can be determined directly from EHR and other
41 ancillary health care data systems (e.g., disease registries, insurance claims data) data.[17] CPs
42 are developed using a combination of data elements (e.g., sociodemographic variables, clinical
43 diagnoses) and value sets (i.e., the selection of a set of relevant values for each data element).
44 Development of CPs using standardized methods and definitions enables identification and
45 inclusion of transgender persons in research, as well as replication of analyses across data
46

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

3
4
5
6
7 4 Recently, CP and other EHR-based algorithm methods have been applied in a number of settings
8 primarily in the US to identify transgender samples for health research.[14] Specifically, the
9 STRONG study identified a transgender cohort (n = 6,456) using EHR data from Kaiser
10 Permanente health plan members in California and Georgia, for investigation of general and
11 transgender-specific health outcomes.[18] Blosnich et al identified 3,177 people with a “gender
12 identity disorder” diagnosis among military veterans accessing care through the US Veterans
13 Health Administration healthcare system,[19] for examination of mental health and other
14 outcomes. Researchers with the US Centers for Medicare & Medicaid Services identified 4,098
15 transgender beneficiaries using national Medicare claims data,[20] and researchers at Vanderbilt
16 University identified 234 transgender patients in their university clinic EHR data.[16] While
17 these cohorts represent important opportunities for advancement of transgender health research,
18 these methods have yet to be applied widely outside the US context. This is particularly
19 important as different jurisdictions may vary in medical billing and coding practices, healthcare
20 system patient populations, and representativeness of the general population. Specifically, in
21 Canada, healthcare is delivered through a provincially administered universal healthcare system.
22 As such, research using EHR provides an opportunity to develop methods for population-based,
23 representative estimates of transgender populations within the Canadian context. Coupled with
24 the current absence of gender ascertainment measures in population-based routinely collected
25 data (e.g., census, national government health surveys, etc.) in Canada and many other
26 jurisdictions, this remains an evidence need.

27 **Summary of study rationale**

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

32 **METHODS**

33 **Data Sources and Participants**

34 The Comparative Outcomes and Service Utilization Trends Study (COAST)
35 COAST is a population-based cohort study focused on health services utilization research
36 questions among all people known to be living with HIV (PLWH) in the province of British
37 Columbia (BC) and a 10% random sample comparison group of the HIV-negative general
38 population.[21] The COAST cohort comprises individual-level, longitudinal data from PLWH
39 who have ever accessed HIV treatment in BC between 1996 and 2013, provided by Population
40 Data BC (PopDataBC)[21] via data linkage between two provincial data sources, by personal
41 health number: the Drug Treatment Program (DTP) [22] and the Ministry of Health. PopDataBC
42 provides infrastructure for access to, and linkage of, longitudinal and individual-level
43 administrative health data for all BC residents.[23]. The HIV-negative general population cohort
44 was drawn randomly from the Ministry of Health registry data by PopDataBC. The COAST study
45 has received approval from the University of British Columbia/Providence Health Care
46 Research Ethics Board (#H09-02905) and Simon Fraser University Office of Research Ethics
47 (#2013 s0566). The study complies with the BC Freedom of Information and Protection of

1
2
3 1 Privacy Act (FIPPA) and did not require informed consent as it is conducted using
4 2 retrospective administrative and anonymized data for research and statistical purposes
5 3 only. No patients or public were involved in this study.
6 4

5 Drug Treatment Program

6 In BC, antiretroviral therapy (ART) is provided to PLWH at no cost to the patient, and distributed
7 through the DTP.[22] The DTP contributed a provider-reported measure of transgender status for
8 COAST.
9

10 Ministry of Health

11 Ministry of Health data available via COAST included insured medical service billing records for
12 outpatient visits,[24,25] hospital (in-patient) visits,[26] prescription medications,[27,28] and vital
13 statistics.[29]
14

15 **Measures & Analyses**

16 Transgender computable phenotypes

17 Identification of transgender cases was tested in COAST using International Classification of
18 Disease (ICD) codes (9th and 10th editions) and exogenous sex hormone prescription use.
19 Transgender-specific diagnoses in medical and hospital billing records included the ICD-9 codes
20 302.5 Trans-sexualism with unspecified history, 302.51 Trans-sexualism with asexual history,
21 302.52 Trans-sexualism with homosexual history, 302.53 Trans-sexualism with heterosexual
22 history, 302.6 Gender Identity Disorder in children, 302.85 Gender Identity Disorder in
23 adolescents or adults; and ICD-10 codes F64.0 Gender Identity Disorder of childhood, F64.2
24 Gender Identity Disorder of childhood, F64.8 Other Gender Identity Disorder, and F64.9 Gender
25 Identity Disorder unspecified. The full list of androgen blockers and exogenous sex hormone
26 prescriptions included in analyses is available in the supplementary material.
27

28 Concordance

29 To assess face validity and utility of diagnosis and prescription data over time in CP
30 development (i.e. whether the identified transgender sample had exogenous sex hormone
31 prescription use and other diagnoses patterns consistent with that of transgender populations in
32 other studies), concordance analyses evaluated the presence of at least one included diagnosis
33 and prescription during the COAST study follow-up period with the presence of at least one
34 included diagnosis and prescription in the last study year. Concordance was assessed between
35 transgender-specific diagnoses, exogenous sex hormone and androgen blocker prescriptions, and
36 non-transgender specific diagnoses (ICD-9 259.9 Unspecified Endocrine Disorder and ICD-10
37 E34.9 Endocrine Disorder, Unspecified [see supplementary material]). Endocrine disorder
38 diagnosis codes are sometimes preferred by medical providers treating transgender people in
39 response to historic exclusions of transgender-specific care from insurance coverage and to
40 combat the stigma of transgender-specific diagnosis codes that have historically been classified
41 as psychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM).[30]
42 Exogenous sex hormone use, while common in transgender populations,[5,31] is not
43 transgender-specific. Cisgender populations also use androgen blocker and sex hormone
44 prescriptions (e.g. estrogen to treat menopausal symptoms in cisgender women, spironolactone is
45 used for hypertension), thus exogenous sex hormone and androgen blocker prescription use
56
57
58
59
60

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

39

1 Validation

2 While no one CP consistently performed well across all validation metrics, the CP with the best
 3 overall performance across test statistics was based on having received at least one transgender-
 4 specific diagnosis and at least one androgen blocker/exogenous sex hormone prescription over
 5 the study follow-up period (Table 2). This CP had high specificity (99.8%, 95% CI: 99.6-99.8),
 6 low sensitivity (27.5%, 95% CI: 17.8-39.8), low to moderate Kappa coefficients (0.3, 95% CI:
 7 0.2-0.5) and moderate positive predictive values (43.2%, 95% CI: 28.7-58.9). This CP also had
 8 the second longest mean follow-up time (mean: 136.3, range: 21.0-203.0), similar overall to the
 9 other CP groups (mean: 136.5, range: 21.0-203.0; mean: 117.1, range: 24.0-198.0; mean: 130.4,
 10 range: 69.0-198.0; respectively).

11
 12 **Table 2. Validation measures of transgender computable phenotype (CP) with provider-
 13 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)

14 Transgender phenotype

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

17 Demographics and chronic conditions

18 Demographic characteristics and chronic conditions for the 188 transgender people identified via
 19 the best-performing CP are presented in Figures 2 to 4. Transgender people were geographical
 20 located throughout BC health regions. The Vancouver Coastal Health Authority region, which
 21 includes the largest municipal area in BC, had the highest concentration of transgender people
 22 (44.2%) while the Northern Health Authority region - a predominantly rural and remote area of
 23 the province - had the lowest (1.6%).^[36] The HIV-positive group had a higher median age than
 24 the HIV-negative group (35 [Q1, Q3: 30,42] and 30 [Q1, Q3: 19,42], respectively). For the HIV-
 25 negative sample, the largest proportion of transgender people were aged 19 to 29 years (44.5%)
 26 and the smallest proportion aged 55 years and older (<3%). For the HIV-positive sample, the
 27 largest proportion were aged 30 to 34 years (25.5%) and the smallest proportion aged 55 years
 28 and older (<2%).

1
2
3 1 [Figures 2 and 3 here]
4 2

5 3 Overall, HIV-positive transgender people had a higher prevalence of at least one chronic
6 4 condition (other than HIV) compared to HIV-negative transgender people (88.2% versus 85.4%,
7 5 respectively), and of two or more chronic conditions (76.5% versus 52.6%, respectively).
8 6 Specific chronic disease differences between transgender people living with and without HIV
9 7 were most notable for a higher prevalence among the HIV-positive cohort of cardiovascular
10 8 disease, chronic kidney disease, osteoarthritis, schizophrenia and personality disorders, and
11 9 chronic liver disease, but a lower prevalence for hypertension.
12
13

14 10
15 11 [Figure 4 here]
16 12

17 13 **DISCUSSION**

18 14 This study demonstrates the feasibility of identification of a sample of transgender people in a
19 15 large linked provincial healthcare administrative database, using a CP based on prescriptions and
20 16 diagnoses. Among a growing number of studies using EHR and CP methods to identify
21 17 transgender samples for health research purposes, this is the first to do so in Canada., to
22 18 independently validate the CP using a ‘gold standard’ of provider-reported transgender status,
23 19 and the only to use population-based data.
24
25

26 21 **Concordance**

27 22 There was high concordance between transgender-specific diagnoses and exogenous sex
28 23 hormone or androgen blocker prescription use in this study. That nearly half of those with at
29 24 least one transgender-specific diagnosis had been dispensed hormones or blockers in the past
30 25 year is consistent with findings from US and Canadian studies (48.9% and 43.0%,
31 26 respectively)[20,32,33] - suggesting face validity for the current CP.
32
33

34 28 **CP development and validation**

35 29 The best-performing CP overall successfully identified cisgender people who were truly
36 30 cisgender (specificity) and correctly identified transgender people who were truly transgender
37 31 (0.2% false positive rate, results not shown). However, the selected CP had relatively low
38 32 sensitivity, missing approximately 72.5% of ‘true’ transgender people in COAST, as identified
39 33 by the gold standard provider-based measure. Though a relatively small proportion of the ‘true’
40 34 transgender sample was identified in this study, the impact on future analyses comparing health
41 35 outcomes for transgender and cisgender groups is likely negligible, as even the large proportion
42 36 of ‘true’ transgender people misclassified as cisgender (approximate n=496) is a very small
43 37 proportion of the total COAST sample. At worst, this misclassification would bias results related
44 38 to disparities between transgender and cisgender health toward the null, producing a conservative
45 39 attenuated effect in COAST, and other such administrative datasets. Further, as discussed below,
46 40 gender identity classification will likely greatly improve as transgender care shifts further into
47 41 the fee-for-service system in BC. As in other Canadian administrative data studies, low
48 42 sensitivity may be explained in part by provider and system billing preferences using 3-digit ICD
49 43 diagnosis coding instead of the more specific 4-digit coding, and inconsistencies in the BC
50 44 billing management system.[37] Despite the low sensitivity, CP development in this study with
51 45 high specificity offers an advancement for transgender health research. A measure that correctly
52 46 identifies cases for transgender samples in research with good success translates to better
53
54
55
56
57
58
59
60

1
2
3 1 opportunities to include transgender people in health studies and to investigate their health
4 2 relative to other groups. While future research may lead to improvements in CP development, the
5 3 CP identified in the current study with good specificity, albeit relatively poor sensitivity, has
6 4 important utility in advancing opportunities in transgender health research. Additionally, while
7 5 differential follow-up time can affect algorithm performance, the similar mean and range follow-
8 6 up time for all CPs in this study suggests that differential follow-up time was not an important
9 7 source of bias in this study.
10 8

11 9
12 10 The limited agreement between the CP and provider-report transgender status may be due to the
13 11 widely varying transgender status prevalence depending on study design and ascertainment
14 12 measures used.[14] In the BC context, the CP and the DTP measures are assessing transgender
15 13 status in different ways and for different purposes. In the DTP, transgender status is ascertained
16 14 in the context of HIV diagnosis and ART prescribing, during which demographics and HIV
17 15 transmission risk factors are recorded. This differs from recording diagnoses in EHR for those
18 16 accessing transgender-specific care as utilized in the CP. This may explain the lower PPV for the
19 17 best-performing CP compared to the CP based on recent transgender diagnoses, suggesting the
20 18 DTP provider-reported transgender status measure has better coverage for recent cases and the
21 19 potential for use of recent diagnosis over ever to be beneficial in future CP development.
22 20 Ultimately, a single CP may not be sufficient for all intended purposes and the best applicable
23 21 CP (using different types of diagnoses, prescriptions or procedures) may differ depending on the
24 22 intended healthcare, health research, or health policy application.[17]
25 23
26 24

27 25
28 26 There is limited literature on EHR-based studies with the ability to validate an administrative
29 27 transgender measure using a ‘gold standard’ comparison measure.[16] The two previous studies
30 28 that have developed and validated algorithms to identify transgender individuals have both been
31 29 conducted in non-representative samples in the US, one using Medicare data[38] and one in a
32 30 university medical center.[16] Similar to the current study, the Medicare study found high
33 31 specificity when comparing an EHR-based and a two-step survey-based transgender measure.
34 32 However, the Medicare study found that the EHR measure performed consistently well with high
35 33 sensitivity and a high Kappa statistic, unlike in the current study. Using chart review as the ‘gold
36 34 standard’ for comparison of transgender status, Ehrenfeld et al. found a low false positive rate for
37 35 their best-performing algorithm (3%), though not as low as the false positive rate in the current
38 36 study. The overall high levels of agreement for transgender measures in the two previous studies
39 37 is likely a function of the lack of independence between the ‘gold standard’ and the CP or
40 38 algorithm measures. Specifically, only those classified as transgender in the Medicare EHR data
41 39 were offered survey participation to complete the two-step ‘gold standard’ survey measure, and
42 40 only those cases identified as transgender in the university clinic EHR were included in chart
43 41 review. Thus, previous studies could assess agreement between the two measures, but not
44 42 robustly validate either. In the current study, the DTP provider-based transgender status measure
45 43 is independent and thus could be used for robust CP validation.
46 44
47 45
48 46
49 47
50 48
51 49
52 50
53 51
54 52
55 53
56 54
57 55
58 56
59 57
60 58

59 42 While not possible to incorporate free-text records in case-finding algorithms in the current study
60 43 as only structured EHR data is linked through COAST, it is worth noting the opportunities
44 44 potentiated by use of NLP and machine learning approaches as methods for identifying
45 45 transgender samples in EHR data as this research area continues to grow. Outside of transgender
46 46 health, the use of NLP and machine learning to mine unstructured free-text EHR data has

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

7 **Transgender status prevalence & ascertainment**

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

20
21 Using the broadest CP algorithm (any transgender-specific diagnosis ever, n=56) and those
22 identified by provider-report together (total n=106), the total transgender PLWH sample would
23 represent as high as 0.88% (range: 0.73-1.1%) of the prevalent HIV infections in BC in
24 2014.[40] This overrepresentation of transgender people among PLWH is consistent with
25 evidence of a disproportionate HIV burden for transgender populations globally,[5,41,42] as well
26 as in line with the only other available data on the proportion of PLWH who are transgender,
27 from US national surveillance data (2012 data: 1.1%, 95%CI: 0.8-1.4).[43]

29 **Demographics and chronic conditions**

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

43 **Limitations**

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

1
2
3 1 accessing medical transition care in BC and may only represent 24% to 47% of the total
4 2 transgender population.[33] This study was also limited by the inability to validate the
5 3 transgender CP among the HIV-negative COAST cohort, as a 'gold standard' provider-based
6 4 transgender measure was only available for the HIV-positive cohort. It is possible that the
7 5 transgender CPs would perform differently in populations living without HIV, particularly as
8 6 healthcare contact is higher among populations living with HIV. Additionally, this study should
9 7 be considered in light of the context in which it was conducted, an environment in which
10 8 transgender healthcare delivery in BC is currently shifting from specialized care settings to the
11 9 main primary care fee-for-service settings. Given that COAST only includes fee-for-service data,
12 10 this study was limited by the inability to capture transgender people who access transgender care
13 11 outside the fee-for-service system. However, fortunately, as the shift to the fee-for-service
14 12 system occurs, transgender ascertainment via CPs in BC will likely improve. The administrative
15 13 data used in this study may also be susceptible to coding error (and coding biases/practices)
16 14 across conditions and settings,[49] potentially introducing misclassification bias in terms of
17 15 transgender ascertainment. Finally, chronic condition prevalence data reported in this study
18 16 should be interpreted with caution, given potential selection bias by serostatus in the COAST
19 17 cohort; though any such bias likely resulted in conservative estimates of difference by serostatus
20 18 in this analysis.
21 19

20 20 CONCLUSION

21 21 This study makes a number of important contributions to the literature on innovative methods in
22 22 transgender health. Major contributions include development and validation of a transgender CP,
23 23 using a population-based representative source population, in the Canadian context. Another
24 24 strength is the approximately complete ascertainment of the population of transgender PLWH in
25 25 BC, and capacity to estimate transgender status prevalence among PLWH. In a current funding
26 26 environment of limited support for longitudinal transgender health studies in the US and none to
27 27 date in Canada, this study and the methods employed offer an efficient, replicable and cost-
28 28 effective way forward in creating electronic cohorts for advancing transgender health
29 29 research.[15] Moreover, the recent rollback of sexual orientation and gender identity data
30 30 collection and legal changes in insurance coverage of transgender healthcare in the US potentiate
31 31 decline in accurate claims coding for gender-affirming care.[30] This highlights the utility of
32 32 work in this area from other jurisdictions, particularly those with transgender-inclusive universal
33 33 healthcare systems such as Canada.
34 34

35 35 Future research should build upon the methods developed in this study and explore
36 36 complimentary approaches for gender identity ascertainment in administrative and EHR data,
37 37 such as machine learning approaches, as have been used to develop algorithms based on
38 38 healthcare utilization data in other research areas. Finally, the current study lays the foundation
39 39 for future work with the ability to study transgender health and healthcare use patterns over time,
40 40 with linkage to laboratory data, as well as inclusion of appropriate comparison groups.[15,50]
41 41

42 42 ACKNOWLEDGEMENTS & DISCLAIMER

43 43 The authors thank the COAST study participants, the BC Centre for Excellence in HIV/AIDS,
44 44 the BC Ministry of Health, BC Vital Statistics Agency, PharmaNet and the institutional data
45 45 stewards for granting access to the data, and Population Data BC for facilitating the data linkage
46 46 process. In addition, we would like to thank the COAST core team members and other support
47 47
48 48
49 49
50 50
51 51
52 52
53 53
54 54
55 55
56 56
57 57
58 58
59 59
60 60

1
2
3 1 staff at these institutions for their administrative assistance with the data access and preparation.
4 2 All inferences, opinions, and conclusions drawn in this paper are those of the authors, and do not
5 3 reflect the opinions or policies of the Data Steward(s). There are no conflicts of interest to
6 4 declare.
7 5

6 **COMPETING INTERESTS**

7 None declared.
8

9 **FUNDING**

10 This work was supported by the Canadian Institutes of Health Research, through an Operation
11 Grant [grant number 130419], a Foundation Award to RSH [grant number 143342], and a
12 Doctoral Research Award to AJR [grant number 152382] and support from the British Columbia
13 Centre for Excellence in HIV/AIDS. The DTP receives funding from the provincial government
14 of British Columbia (PharmaCare). The funders had no role in the study design, analysis,
15 interpretation of the data, drafting of the manuscript or in the decision to submit for
16 publication.
17

18 **CONTRIBUTORS**

19 AJR led the study from conceptualization to analysis plan to interpretation, drafting of the first
20 manuscript version, revisions and final version. RSH acquired study data and funding. TP, MK,
21 PS, TS, and RSH all contributed to study design, interpretation of results, and reviewed manuscript
22 versions. JL and MY contributed to study analysis and reviewed manuscript versions. All authors
23 provided critical review of first and subsequent manuscript drafts, approved the final version, and
24 agree to be accountable for the work presented.
25

26 **PATIENT AND PUBLIC INVOLVEMENT**

27 No patients involved.
28

29 **DATA SHARING STATEMENT**

30 The data used for this study are held by the BC Centre for Excellence in HIV/AIDS under the
31 authority of the BC Ministry of Health; as they contain confidential patient health records
32 including HIV serostatus, data cannot be made available to other parties.
33
34
35
36
37

1
2
3 **1 FIGURES LEGENDS**
4

5 **2 Figure 1. Total transgender sample identified using a computable phenotype with**
6 **3 electronic health records**
7
8
9

10 **5 Figure 2. Geographic distribution of transgender people across province, by health**
11 **6 authority***

12 *7 % of transgender individuals with known health authority (n=182)*
13
14

15 **9 Figure 3. Age distribution of transgender sample, by HIV serostatus**
16
17

10

18 **11 Figure 4. Co-morbidities among transgender sample, by HIV serostatus**
19
20

12

21
22 **13**
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

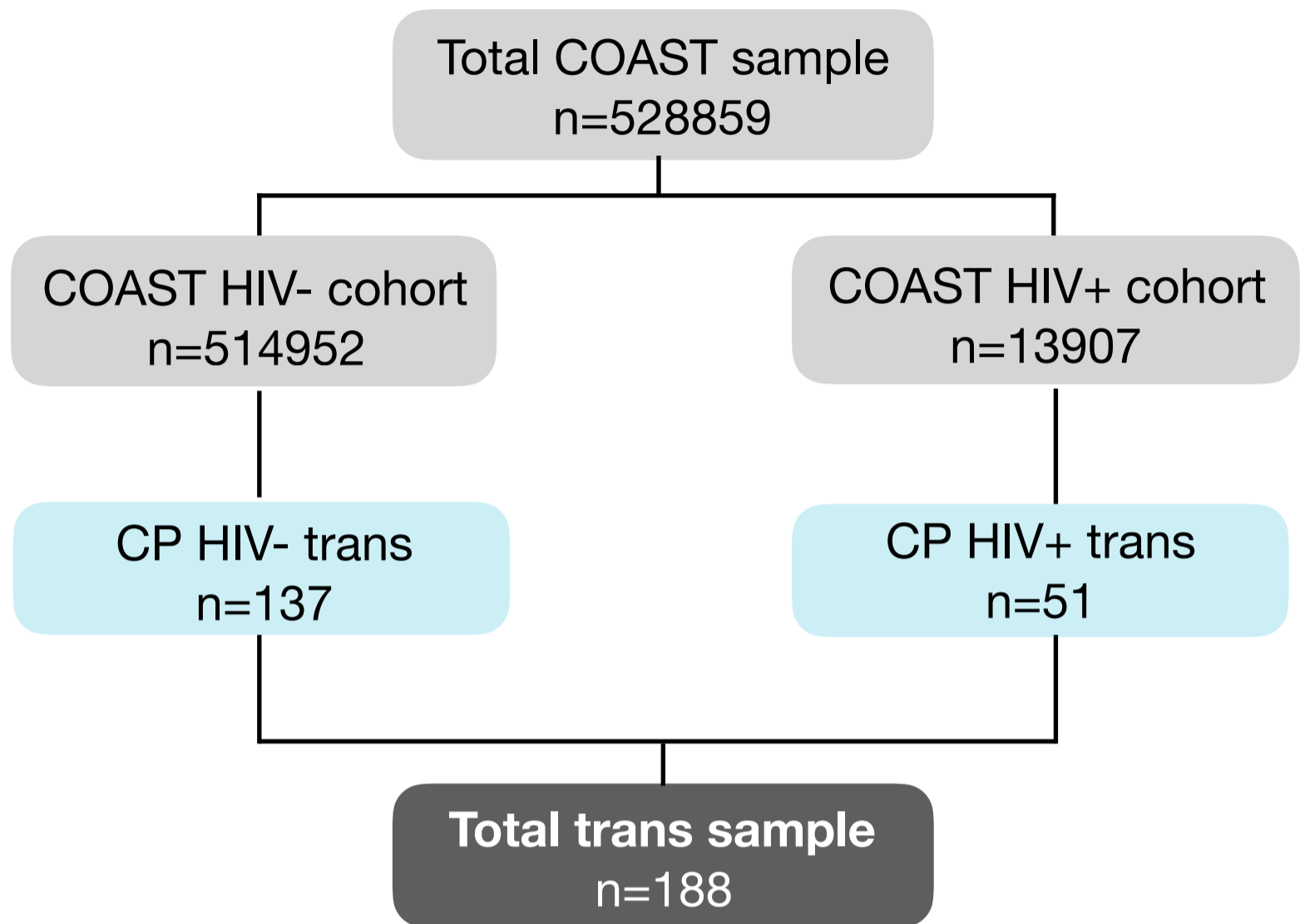
1 REFERENCES

- 2 1. Bauer GR. Making sure everyone counts: Considerations for inclusion, identification, and
3 analysis of transgender and transsexual participants in health surveys. *What Differ Sex Gen*
4 *Make Gen Sex Health Res Casebook*. Vancouver, Canada: Canadian Institutes of Health
5 Research (CIHR), Institute for Gender and Health (IGH); 2012. p. 59–67.
- 6 2. Bauer GR, Hammond R, Travers R, Kaay M, Hohenadel KM, Boyce M. “I don’t think this is
7 theoretical; this is our lives”: How erasure impacts health care for transgender people. *JANAC*.
8 2009;20:348–361.
- 9 3. Poteat T, German D, Flynn C. The conflation of gender and sex: Gaps and opportunities in
10 HIV data among transgender women and MSM. *Glob Public Health*. 2016;11:835–848.
- 11 4. Wanta JW, Unger CA. Review of the transgender literature: Where do we go from here?
12 *Transgender Health*. 2017;2:119–128.
- 13 5. Herbst JH, Jacobs ED, Finlayson TJ, McKleroy VS, Neumann MS, Crepaz N, et al.
14 Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a
15 systematic review. *AIDS Behav*. 2008;12:1–17.
- 16 6. Becasen JS, Denard CL, Mullins MM, Higa DH, Sipe TA. Estimating the prevalence of HIV
17 and sexual behaviors among the US transgender population: A systematic review and meta-
18 analysis, 2006–2017. *Am J Public Health*. 2019;109:e1–8.
- 19 7. Marshall Z, Welch V, Thomas J, Brunger F, Swab M, Shemilt I, et al. Documenting research
20 with transgender and gender diverse people: protocol for an evidence map and thematic analysis.
21 *Syst Rev*. 2017;6:35.
- 22 8. Reisner SL, Poteat T, Keatley J, Cabral M, Mothopeng T, Dunham E, et al. Global health
23 burden and needs of transgender populations: a review. *The Lancet*. 2016;388:412–436.
- 24 9. Cathcart-Rake EJ, Lightner DJ, Quevedo FJ, Ertz DC, Jatoi A. Cancer in transgender patients:
25 One case in 385,820 is indicative of a paucity of data. *J Oncol Pract*. 2017;JOP2017027714.
- 26 10. Meerwijk EL, Sevelius JM. Transgender population size in the United States: A meta-
27 regression of population-based probability samples. *Am J Public Health*. 2017;107:e1–e8.
- 28 11. Flores AR, Herman JL, Gates GJ, Brown TNT. How Many Adults Identify as Transgender in
29 the United States? [Internet]. Los Angeles, CA: Williams Institute; 2016 Jun p. 13. Available
30 from: [https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-](https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf)
31 [Transgender-in-the-United-States.pdf](https://williamsinstitute.law.ucla.edu/wp-content/uploads/How-Many-Adults-Identify-as-Transgender-in-the-United-States.pdf)
- 32 12. The GenIUSS Group. Best Practices for Asking Questions to Identify Transgender and Other
33 Gender Minority Respondents on Population-Based Surveys. Los Angeles, CA: The Williams
34 Institute; 2014.

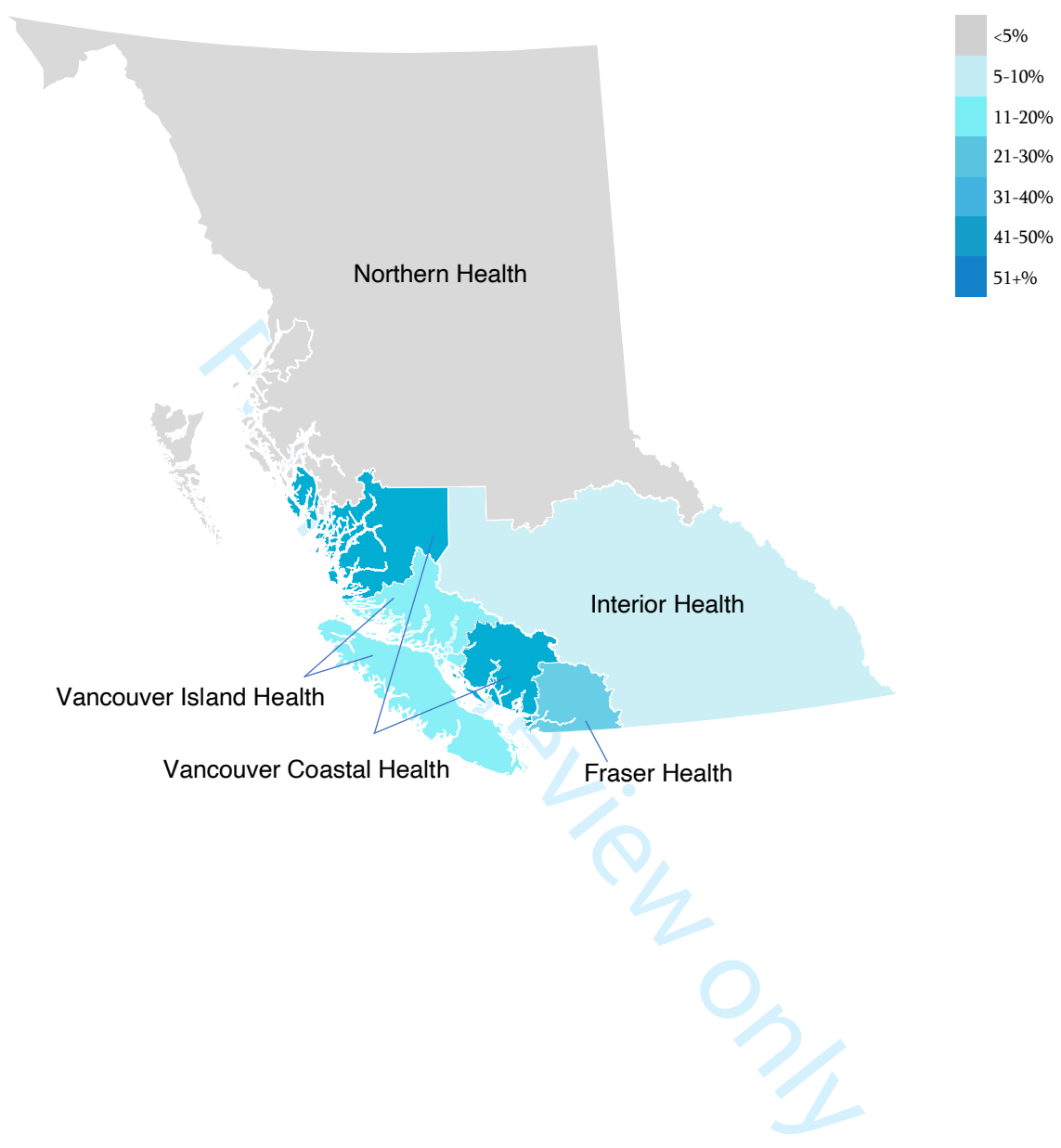
13. Bauer GR, Braimoh J, Scheim AI, Dharma C. Transgender-inclusive measures of sex/gender for population surveys: Mixed-methods evaluation and recommendations. *PLOS ONE*. 2017;12:e0178043.
14. Collin L, Reisner SL, Tangpricha V, Goodman M. Prevalence of transgender depends on the “case” definition: 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 Textbook 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. HIV/AIDS. 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>

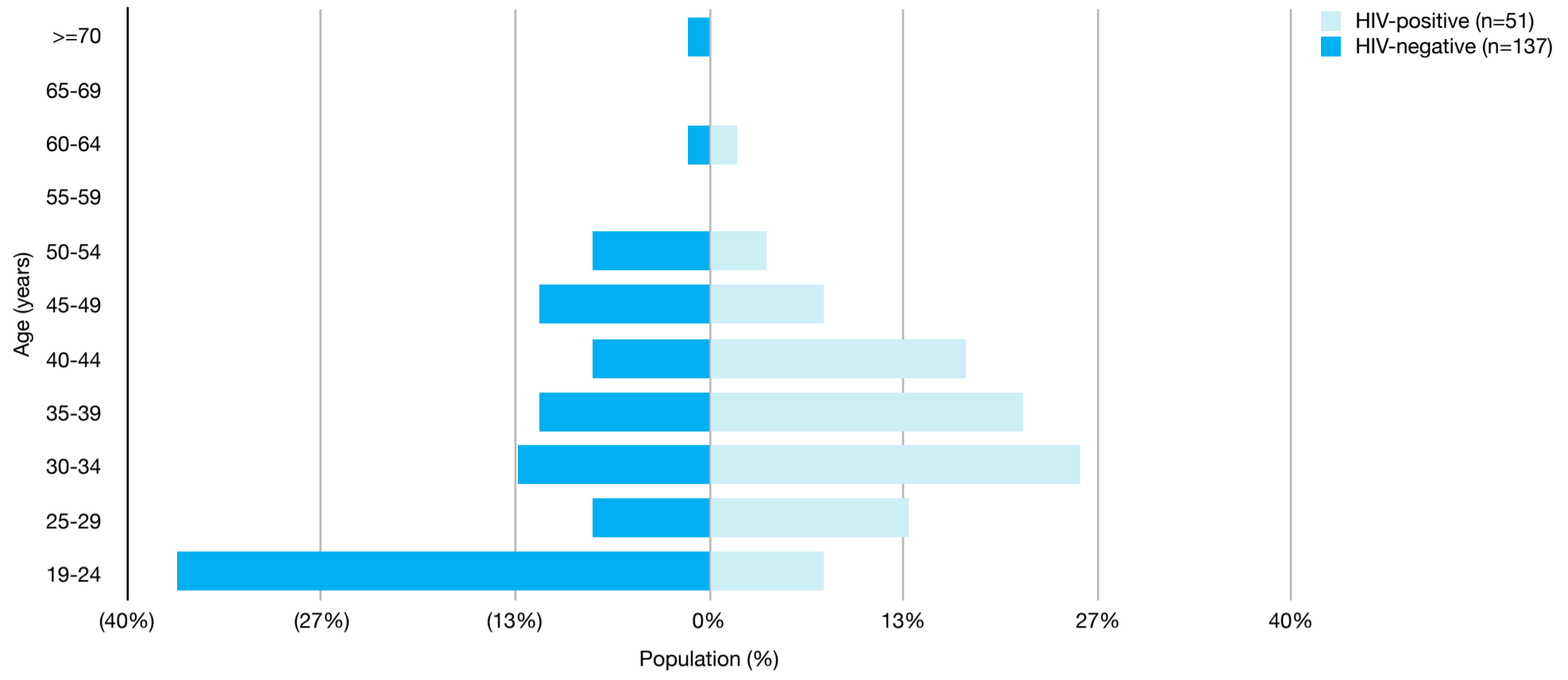
- 1
2
3 1 26. Canadian Institute for Health Information (2014). Discharge Abstract Database (Hospital
4 2 Separations) [Internet]. Popul. Data BC. 2014 [cited 2019 Sep 27]. Available from:
5 3 <http://www.popdata.bc.ca/data>
6
7
8 4 27. British Columbia Ministry of Health (2014). PharmaCare [Internet]. Popul. Data BC. 2014
9 5 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca/data>
10
11 6 28. British Columbia Ministry of Health (2014). PharmaNet [Internet]. Popul. Data BC. 2014
12 7 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca/data>
13
14 8 29. British Columbia Vital Statistics Agency (2014). Vital Statistics Deaths [Internet]. Popul.
15 9 Data BC. 2014 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca>
16
17
18 10 30. Arnold JD, Nelson AE, Loubier EM. Trends in Insurance Coverage for Gender-Affirming
19 11 Surgeries. *JAMA Surg.* 2018;153:972.
20
21 12 31. Radix A, Sevelius J, Deutsch MB. Transgender women, hormonal therapy and HIV
22 13 treatment: a comprehensive review of the literature and recommendations for best practices. *J Int*
23 14 *AIDS Soc.* 2016;19:20810.
24
25 15 32. Rotondi NK, Bauer GR, Scanlon K, Kaay M, Travers R, Travers A. Nonprescribed hormone
26 16 use and self-performed surgeries: “do-it-yourself” transitions in transgender communities in
27 17 Ontario, Canada. *Am J Public Health.* 2013;103:1830–1836.
28
29
30 18 33. Scheim AI, Bauer GR. Sex and gender diversity among transgender persons in Ontario,
31 19 Canada: results from a respondent-driven sampling survey. *J Sex Res.* 2015;52:1–14.
32
33 20 34. Chronic Disease Information Working Group. BC Chronic Disease and Selected Procedure
34 21 Case Definitions [Internet]. British Columbia, Canada: BC Ministry of Health; 2019 Apr p. 1–
35 22 132. Available from: [https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-](https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf)
36 23 [access/chronic-disease-registries-case-definitions.pdf](https://www2.gov.bc.ca/assets/gov/health/conducting-health-research/data-access/chronic-disease-registries-case-definitions.pdf)
37
38
39 24 35. British Columbia Cancer Agency. BC Cancer Registry Data [Internet]. Popul. Data BC. 2014
40 25 [cited 2019 Sep 27]. Available from: <http://www.popdata.bc.ca/data>
41
42 26 36. BC Stats. Sub-provincial Population Estimates [Internet]. 2018 [cited 2019 Feb 1]. Available
43 27 from: <file://localhost/Users/AJRich/Zotero/storage/5VM2LBUN/PopulationEstimates.html>
44
45 28 37. Hu W. Diagnostic Codes in MSP Claim Data. Program Monitoring and Information
46 29 Management Branch, Resource Management Division, Medical Services Plan; 1996 p. 39.
47
48 30 38. Gerth J, Becerra-Culqui T, Bradlyn A, Getahun D, Hunkeler EM, Lash TL, et al. Agreement
49 31 between medical records and self-reports: Implications for transgender health research. *Rev*
50 32 *Endocr Metab Disord.* 2018;19:263–269.
51
52
53 33 39. Liao KP, Cai T, Savova GK, Murphy SN, Karlson EW, Ananthakrishnan AN, et al.
54 34 Development of phenotype algorithms using electronic medical records and incorporating natural
55
56
57
58
59
60

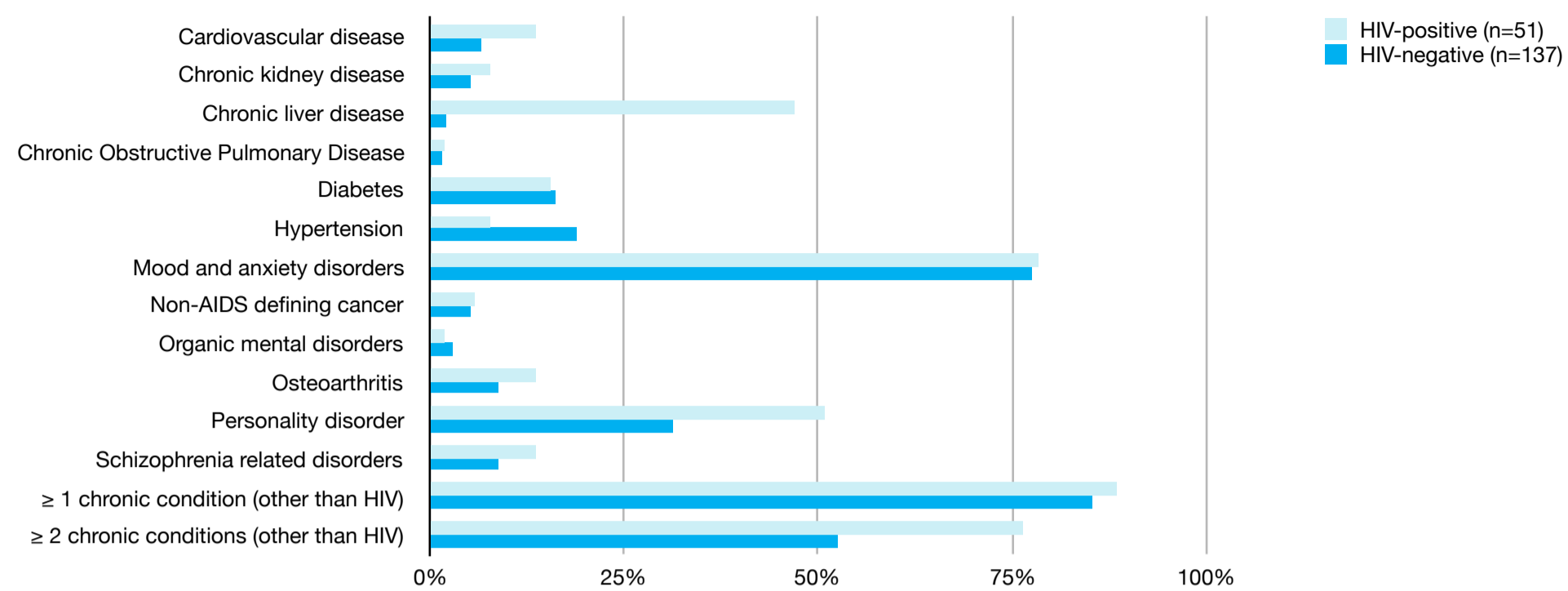
- 1 language processing. *The BMJ* [Internet]. 2015 [cited 2020 Oct 22];350. Available from:
2 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4707569/>
- 3 40. BC Centre for Disease Control. *HIV in British Columbia: Annual Surveillance Report 2015*.
4 Vancouver, Canada: BC Centre for Disease Control; 2017.
- 5 41. Poteat T, Scheim A, Xavier J, Reisner S, Baral S. Global epidemiology of HIV infection and
6 related syndemics affecting transgender people. *J Acquir Immune Defic Syndr*. 2016;72 Suppl
7 3:S210–9.
- 8 42. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of
9 HIV in transgender women: a systematic review and meta-analysis. *Lancet Infect Dis*.
10 2013;13:214–222.
- 11 43. Centers for Disease Control and Prevention. Behavioral and Clinical Characteristics of
12 Persons Receiving Medical Care for HIV Infection—Medical Monitoring Project, United States,
13 2012 [Internet]. Washington, D.C.: Centers for Disease Control; 2015 Aug p. 42. Report No.: 12.
14 Available from: [https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-](https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf)
15 [HSSR_MMP_2012.pdf](https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-HSSR_MMP_2012.pdf)
- 16 44. Grabovac I, Veronese N, Stefanac S, Haider S, Jackson SE, Koyanagi A, et al. Human
17 Immunodeficiency Virus Infection and Diverse Physical Health Outcomes: An Umbrella Review
18 of Meta-analyses of Observational Studies. *Clin Infect Dis*. 2019;ciz539.
- 19 45. Nasi M, De Biasi S, Gibellini L, Bianchini E, Pecorini S, Bacca V, et al. Ageing and
20 inflammation in patients with HIV infection: Ageing, inflammation and HIV infection. *Clin Exp*
21 *Immunol*. 2017;187:44–52.
- 22 46. Judge C., O’Donovan C., Callaghan G., Gaoatswe G., O’Shea D. Gender dysphoria -
23 prevalence and co-morbidities in an Irish adult population. *Front Endocrinol*. 2014;5:87.
- 24 47. Cheung AS, Ooi O, Leemaqz S, Cundill P, Silberstein N, Bretherton I, et al.
25 Sociodemographic and clinical characteristics of transgender adults in Australia. *Transgender*
26 *Health*. 2018;3:229–38.
- 27 48. Swartz JA, Ducheny K, Holloway T, Stokes L, Willis S, Kuhns LM. A latent class analysis
28 of chronic health conditions among HIV-Positive transgender women of color. *AIDS Behav*
29 [Internet]. 2019 [cited 2019 Oct 2]; Available from: [http://link.springer.com/10.1007/s10461-](http://link.springer.com/10.1007/s10461-019-02543-3)
30 [019-02543-3](http://link.springer.com/10.1007/s10461-019-02543-3)
- 31 49. Fisher ES, Whaley FS, Krushat WM, Malenka DJ, Fleming C, Baron JA, et al. The accuracy
32 of Medicare’s hospital claims data: progress has been made, but problems remain. *Am J Public*
33 *Health*. 2011;
- 34 50. Reisner SL, Radix A, Deutsch MB. Integrated and gender-affirming transgender clinical care
35 and research. *J Acquir Immune Defic Syndr*. 2016;72 Suppl 3:S235–42.



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60







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		
Spirolactone		
	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

1		
2		
3		
4	ETHYNODIOL D-ETHINYL ESTRADIOL	28630
5	ETHYNODIOL D-ETHINYL ESTRADIOL	469327
6	NORETHINDRONE-MESTRANOL	22608
7	NORETHINDRONE-MESTRANOL	22659
8		
9	NORETHINDRONE A-E ESTRADIOL	297143
10	NORETHINDRONE A-E ESTRADIOL	315966
11	NORETHINDRONE-ETHINYL ESTRAD	317047
12	NORETHINDRONE-ETHINYL ESTRAD	372846
13	NORETHINDRONE-ETHINYL ESTRAD	373265
14	NORETHINDRONE-ETHINYL ESTRAD	531006
15	NORETHINDRONE-ETHINYL ESTRAD	538590
16	NORETHINDRONE-ETHINYL ESTRAD	602957
17	NORETHINDRONE-ETHINYL ESTRAD	620947
18	NORETHINDRONE-ETHINYL ESTRAD	2187086
19	NORETHINDRONE-ETHINYL ESTRAD	2187108
20	NORETHINDRONE-ETHINYL ESTRAD	2189054
21	NORETHINDRONE-ETHINYL ESTRAD	34207
22	NORGESTREL-ETHINYL ESTRADIOL	300640
23	NORGESTREL-ETHINYL ESTRADIOL	579386
24	LEVONORGESTREL-ETH ESTRA	707600
25	LEVONORGESTREL-ETH ESTRA	782416
26	LEVONORGESTREL-ETH ESTRA	782432
27	LEVONORGESTREL-ETH ESTRA	2042320
28	LEVONORGESTREL-ETH ESTRA	2043033
29	NORGESTREL-ETHINYL ESTRADIOL	2043033
30	LEVONORGESTREL-ETH ESTRA	2043726
31	NORGESTREL-ETHINYL ESTRADIOL	2258560
32	NORGESTIMATE-ETHINYL ESTRADIOL	2258560
33	NORETHINDRONE-MESTRANOL	30333
34	NORETHINDRONE-MESTRANOL	30341
35	NORETHINDRONE-MESTRANOL	30341
36	LEVONORGESTREL-ETH ESTRA	2236974
37	LEVONORGESTREL-ETH ESTRA	2236974
38	ETHYNODIOL D-ETHINYL ESTRADIOL	471526
39	ETHYNODIOL D-ETHINYL ESTRADIOL	471526
40	NORETHINDRONE-ETHINYL ESTRAD	340731
41	NORETHINDRONE-ETHINYL ESTRAD	340731
42	NORETHINDRONE-MESTRANOL	340758
43	NORETHINDRONE-MESTRANOL	340758
44	NORETHINDRONE A-E ESTRADIOL	343838
45	NORETHINDRONE A-E ESTRADIOL	343838
46	NORETHINDRONE A-E ESTRADIOL	353027
47	NORETHINDRONE A-E ESTRADIOL	353027
48	NORETHINDRONE-ETHINYL ESTRAD	372838
49	NORETHINDRONE-ETHINYL ESTRAD	372838
50	NORETHINDRONE-ETHINYL ESTRAD	373273
51	NORETHINDRONE-ETHINYL ESTRAD	373273
52	NORETHINDRONE-ETHINYL ESTRAD	531014
53	NORETHINDRONE-ETHINYL ESTRAD	531014
54	NORETHINDRONE-ETHINYL ESTRAD	602965
55	NORETHINDRONE-ETHINYL ESTRAD	602965
56	NORETHINDRONE-ETHINYL ESTRAD	695734
57	NORETHINDRONE-ETHINYL ESTRAD	695734
58	NORETHINDRONE-ETHINYL ESTRAD	2187094
59	NORETHINDRONE-ETHINYL ESTRAD	2187094
60		

Supplementary material

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1		
2		
3	MEDROXYPROGESTERONE ACET	2246627
4	MEDROXYPROGESTERONE ACET	30937
5	MEDROXYPROGESTERONE ACET	2010739
6	MEDROXYPROGESTERONE ACET	2148560
7	MEDROXYPROGESTERONE ACET	2221292
8	MEDROXYPROGESTERONE ACET	2229839
9	MEDROXYPROGESTERONE ACET	2244727
10	MEDROXYPROGESTERONE ACET	2246628
11	MEDROXYPROGESTERONE ACET	729973
12	MEDROXYPROGESTERONE ACET	2010933
13	MEDROXYPROGESTERONE ACET	2148579
14	MEDROXYPROGESTERONE ACET	2221306
15	MEDROXYPROGESTERONE ACET	2229840
16	MEDROXYPROGESTERONE ACET	2246629
17	MEDROXYPROGESTERONE ACET	30945
18	MEDROXYPROGESTERONE ACET	2267640
19	MEDROXYPROGESTERONE ACET	2267640
20	MEDROXYPROGESTERONE ACET	2267640
21	MEDROXYPROGESTERONE ACET	2267640
22	MEDROXYPROGESTERONE ACET	2267640
23	MEDROXYPROGESTERONE ACET	2267640
24	MEDROXYPROGESTERONE ACET	2267640
25	MEDROXYPROGESTERONE ACET	2267640
26	MEDROXYPROGESTERONE ACET	2267640
27	MEDROXYPROGESTERONE ACET	2267640
28	MEDROXYPROGESTERONE ACET	2267640
29	MEDROXYPROGESTERONE ACET	2267640
30	MEDROXYPROGESTERONE ACET	2267640
31	MEDROXYPROGESTERONE ACET	2267640
32	MEDROXYPROGESTERONE ACET	2267640
33	Transmasculine	
34	Testosterone	
35	TESTOSTERONE	2249499
36	TESTOSTERONE CYPIONATE	30783
37	TESTOSTERONE PROPIONATE	1977571
38	TESTOSTERONE CYPIONATE	1977601
39	TESTOSTERONE CYPIONATE	2220318
40	TESTOSTERONE CYPIONATE	2246063
41	TESTOSTERONE ENANTHATE	29246
42	TESTOSTERONE ENANTHATE	716936
43	TESTOSTERONE ENANTHATE	739944
44	TESTOSTERONE UNDECANOATE	782327
45	TESTOSTERONE ENANTHATE/ESTRAD	108278
46	TESTOSTERONE ENANTHATE/ESTRAD	2061031
47	TESTOSTERONE	2239653
48	TESTOSTERONE	2245346
49	TESTOSTERONE	2245345
50	TESTOSTERONE	2245972
51		
52		
53		
54		
55		
56		
57		
58		
59		
60		

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
	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-9: 410 Acute myocardial infarction 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
	Chronic heart failure	1 or more hospitalizations or 2 or more medical visits in 365 days with relevant ICD codes	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 ICD-10: I50 Heart failure
	Stroke- hospital	1 or more hospitalizations with relevant ICD codes	ICD-9: 428 Heart failure 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

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

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

Chronic Obstructive
 Pulmonary Disease*

1 or more hospitalization or 2 or more medical visits within 365 days

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.

Supplementary material

Hypertension*

Excludes gestational diabetes.

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

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

F68 Other disorders of adult personality & behavior

ICD-9:

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

Supplementary material

296 Affective psychoses 300
Neurotic disorders 311 Depressive
disorder, not elsewhere classified

MSP DX Code:
50B Anxiety/Depression
*Cancer case definition details
available from the British
Columbia Cancer Agency:
[http://www.bccancer.bc.ca/health-
info/types-of-cancer](http://www.bccancer.bc.ca/health-info/types-of-cancer)*

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)

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 Amnesic 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
ICD-10:
M15 Polyarthrosis
M16 Coxarthrosis [arthrosis of
hip]
M17 Gonarthrosis [arthrosis of
knee]
M18 Arthrosis of first
carpometacarpal joint
M19 Other arthrosis

Osteoarthritis* 1 or more
hospitalization or 2 or
more medical visits in
365 days with a
relevant ICD code

Supplementary material

1			
2			
3			
4			
5			ICD-9:
6			715 Osteoarthritis and allied
7			disorders
8	Personality disorder	1 or more	ICD-9:
9		hospitalizations or	301.x Personality disorders
10		medical visits with a	
11		relevant diagnosis	ICD-10:
12		within 365 days	F60.x Specified personality
13			disorders
14			F62 Enduring personality
15			changes, not attributable to brain
16			damage and disease
17			F68.1 Intentional production or
18			feigning of symptoms or
19			disabilities, either physical or
20			psychological
21			F68.8 Other specified disorders or
22			adult personality and behaviour
23			F69 Unspecified disorder or adult
24			personality and behaviour
25			ICD-9:
26			295.x Schizophrenic disorders
27	Schizophrenia related	1 or more medical visit	297.0 Paranoid state, simple
28	disorder	or hospitalizations with	297.1 Delusional disorder
29		relevant diagnoses	297.2 Paraphrenia
30		within 365 days	297.3 Shared psychotic disorder
31			
32			ICD-10:
33			F20.x Paranoid schizophrenia
34			F21.x Schizotypal disorder
35			F23.2 Acute schizophrenia-like
36			psychotic disorder
37			F25.x Schizoaffective disorders
38			
39			
40			
41			
42			

* 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)

<p>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27</p> <p>Participants</p>	<p>6</p>	<p>(a) <i>Cohort study</i> - Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> - 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 <i>Cross-sectional study</i> - Give the eligibility criteria, and the sources and methods of selection of participants</p> <p>(b) <i>Cohort study</i> - For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> - For matched studies, give matching criteria and the number of controls per case</p>		<p>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.</p> <p>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.</p> <p>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.</p>	<p>6.1: Methods (pp 6-7); Supplementary Material</p> <p>6.2: Methods (pp 6-7)</p> <p>6.3: Figure 1</p>
<p>28 29 30 31 32 33 34</p> <p>Variables</p>	<p>7</p>	<p>Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.</p>		<p>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.</p>	<p>7.1: Methods (pp 6-7); Supplementary Material</p>
<p>35 36 37 38 39 40 41 42 43 44 45 46 47</p> <p>Data sources/ measurement</p>	<p>8</p>	<p>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</p>			<p>Methods (pp 6-7)</p>

1 2 3 4 5 6 7 8 9 10	Bias	9	Describe any efforts to address potential sources of bias			Discussion (pp 9, 11)
11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34	Study size	10	Explain how the study size was arrived at			Methods (page 6); Figure 1
35 36 37 38 39 40 41 42 43 44 45 46 47	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) <i>Cohort study</i> - If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> - If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> - 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 (<i>e.g.</i> , 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>i.e.</i> , 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 (<i>e.g.</i> , 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) <i>Cohort study</i> - summarise follow-up time (<i>e.g.</i> , average and total amount)			Results (page 8)
Outcome data	15	<i>Cohort study</i> - Report numbers of outcome events or summary measures over time <i>Case-control study</i> - Report numbers in each exposure			Results

		category, or summary measures of exposure <i>Cross-sectional study</i> - 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, Guttman 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.

*Checklist is protected under Creative Commons Attribution ([CC BY](https://creativecommons.org/licenses/by/4.0/)) license.