



# Context is everything: a recommendation for selecting case definitions when using linked administrative health data in HIV research

Program ID: EPHP1.10

Track: Epidemiology and Public Health Sciences

Taylor McLinden, Ni Gusti Ayu Nanditha, Andreea Bratu, Martin St-Jean, Paul Sereda, Viviane D. Lima, Robert S. Hogg, Rolando Barrios



: [tmclinden@cfenet.ubc.ca](mailto:tmclinden@cfenet.ubc.ca)

CAHR 2020 Virtual

May 1 & 2, 2020

Hope, Victories and Perseverance beyond 2020



Conflicts of interest: none



- Healthcare data collected for **non-research** purposes: e.g.,
  - Outpatient physician billings ('MSP')
  - Hospitalizations ('Discharge Abstract Database')
  - Prescription drug data ('PharmaNet')
    - **Source:** BC Ministry of Health
- With **administrative data**, the researcher does **NOT**:
  - Generate research questions / hypotheses *prior* to data collection
  - Decide what is measured (collected)
  - Create the measurement tools (e.g., questionnaires)
  - Administer the measurement tools (no interaction with participants)





# Secondary data

- **Secondary data analyses:** linked administrative health data studies
  - Participants do not routinely fill out questionnaires / surveys \*
  - Often use admin data to define **measurements** (e.g., health conditions)



HIV DATA II cont'd Centre-Patient ID-Visit# \_\_\_\_\_

Has the patient ever had an HIV genotype done?  Yes  No  
 If **YES**, which date(s) and please attach report (s): \_\_\_\_\_ / \_\_\_\_\_ / \_\_\_\_\_  
 \_\_\_\_\_ / \_\_\_\_\_ / \_\_\_\_\_  
 day month year

Please make sure the following tests are done at baseline

**Lymphocyte Subsets**

CD4	Value	_____	cells/ $\mu$ L	CD4%	_____
CD8	Value	_____	cells/ $\mu$ L	CD8%	_____
CD4/CD8 Ratio	Value	_____			
CD38	Value	_____	cells/ $\mu$ L		

**Virology**

HIV RNA  
 Date \_\_\_\_\_ / \_\_\_\_\_ / \_\_\_\_\_  
 Value \_\_\_\_\_ Type of test \_\_\_\_\_

**Tick off an answer for each diagnosis**  
 Does the patient have any of the following diagnoses?

Cardiovascular disease	<input type="checkbox"/> Yes <input type="checkbox"/> No	Thyroid disease	<input type="checkbox"/> Yes <input type="checkbox"/> No
Hypercholesterolemia	<input type="checkbox"/> Yes <input type="checkbox"/> No	Lipodystrophy	<input type="checkbox"/> Yes <input type="checkbox"/> No
Autoimmune disease	<input type="checkbox"/> Yes <input type="checkbox"/> No	Psoriasis	<input type="checkbox"/> Yes <input type="checkbox"/> No
Hypertension	<input type="checkbox"/> Yes <input type="checkbox"/> No	Diabetes	<input type="checkbox"/> Yes <input type="checkbox"/> No

**Has the patient ever had any of the following psychiatric diagnoses?**

Depression	<input type="checkbox"/> Yes <input type="checkbox"/> No
Bipolar Disorder	<input type="checkbox"/> Yes <input type="checkbox"/> No
Schizophrenia	<input type="checkbox"/> Yes <input type="checkbox"/> No
Personality Disorder	<input type="checkbox"/> Yes <input type="checkbox"/> No

Other: \_\_\_\_\_

01/06/06 P.18  
0097247122

~~'Tick off an answer for each diagnosis:'~~

~~'Does the patient have any of the following diagnoses?'~~

~~'Has the patient ever had any of the following psychiatric diagnoses?'~~

\* **COAST & STOP HIV:**

Include other linked datasets that contain primary data (from questionnaires)

However: many measurements are still derived entirely from administrative data





# Case definitions

- **Case definition:** collection of **codes** (e.g., ICD-9, ICD-10-CA, DIN) that are used to measure the presence / absence of a condition

- **Frequency** (e.g., two ICD-9 codes)
- **Time-window** (e.g., within one year)
- **Datasets** (e.g., within the outpatient data)

- **Example:** “The case definition for depression requires 1 hospitalization or 2 physician visits in one year with the following ICD codes...”

- **Note:** **50B** = BC-specific diagnostic code used in outpatient data (‘Anxiety / Depression’)

**CASE DEFINITION – DEPRESSION**

Case Definition Type: Cumulative Prevalence  Episodic Prevalence  Annual Prevalence  Annual Service Utilization Prevalence   
Annual Prevalence with 365 Days Follow-up

Case Definition: The case definition for Depression requires one hospitalization or two physician visits in one year with ICD code(s) specified below. The case definition applies to persons aged 1 and older.

Signed-off BC Case Definition: Yes  - V2017 No

Algorithm: 1H or 2P in 1Y with ICD code(s) specified below.

Notes: None

Age Restriction: Age 1 +

Data Source(s):

Data Source	ICD Code/Procedure Code/Rx	ICD Code/Procedure Code Position	Diagnosis Type	Hospital Care Level
Hospital ICD-10 <b>Hospital ICD-10</b>	F32, F33 <b>F32, F33</b>	First Position Only <input type="checkbox"/> All Positions <input checked="" type="checkbox"/> Others <input type="checkbox"/> N/A <input type="checkbox"/>	M-Most Responsible Diagnosis <input checked="" type="checkbox"/> 1-Pre-Admit Comorbidity <input checked="" type="checkbox"/> 2-Post-Admit Comorbidity <input checked="" type="checkbox"/> 3-Secondary Diagnosis <input checked="" type="checkbox"/> 4-Morphology Code <input checked="" type="checkbox"/> 5-Admitting Diagnosis <input checked="" type="checkbox"/> 6-Proxy Most Responsible Diagnosis <input checked="" type="checkbox"/> 9-External Cause of Injury Code <input checked="" type="checkbox"/> 0-Newborn <input checked="" type="checkbox"/> W,X,Y- Service Transfer Diagnosis <input checked="" type="checkbox"/>	Acute Care <input checked="" type="checkbox"/> Rehabilitation <input checked="" type="checkbox"/> Day Surgery <input checked="" type="checkbox"/>
Hospital ICD-9 <b>Hospital ICD-9</b>	296, 311 <b>296, 311</b>	First Position Only <input type="checkbox"/> All Positions <input checked="" type="checkbox"/> Others <input type="checkbox"/> N/A <input type="checkbox"/>	M-Most Responsible Diagnosis <input checked="" type="checkbox"/> 1-Pre-Admit Comorbidity <input checked="" type="checkbox"/> 2-Post-Admit Comorbidity <input checked="" type="checkbox"/> 3-Secondary Diagnosis <input checked="" type="checkbox"/> 4-Morphology Code <input checked="" type="checkbox"/> 5-Admitting Diagnosis <input checked="" type="checkbox"/> 6-Proxy Most Responsible Diagnosis <input checked="" type="checkbox"/> 9-External Cause of Injury Code <input checked="" type="checkbox"/> 0-Newborn <input checked="" type="checkbox"/> W,X,Y- Service Transfer Diagnosis <input checked="" type="checkbox"/>	Acute Care <input checked="" type="checkbox"/> Rehabilitation <input checked="" type="checkbox"/> Day Surgery <input checked="" type="checkbox"/>
Physician Claims ICD-9 <b>Outpatient ICD-9</b>	296, 311, 50B (see note) <b>296, 311, 50B</b>	First Position Only <input type="checkbox"/> All Positions <input checked="" type="checkbox"/> Others <input type="checkbox"/> N/A <input type="checkbox"/>	N/A	N/A
PharmaNet Drug History	N/A	N/A	N/A	N/A

Note: 50B is a BC Specific diagnostic code used by physician billings (MSP). It is used for both anxiety and depression.

ICD-9/10	Description
F32	Depressive episode
F33	Recurrent depressive disorder
296	Affective psychoses
311	Depressive disorder, not elsewhere classified
MSP DX Code 50B	Anxiety/Depression

Procedure Code: N/A  
Drug List: N/A

**F32: Depressive episode**  
**F33: Recurrent depressive**  
**296: Affective psychoses**  
**311: Depressive disorder**  
**50B: Anxiety / Depression**



# Recommendation

## Abstract

**Background:** Despite not being collected for research purposes, administrative data are increasingly being used in epidemiology. In British Columbia (BC), the Comparative Outcomes And Service Utilization Trends (COAST) and STOP HIV/AIDS studies are based on linkages between HIV-related clinical data and provincial administrative datasets. In such studies, we define health outcomes using 'case definitions': e.g., a depression definition may require one hospitalization (ICD-9: 296/311, ICD-10: F32/F33) or two outpatient (ICD-9: 296/311/'50B') codes within one year. The objective of this abstract is to promote discussion regarding the use of administrative data in HIV research.

**Methods:** Given the expanding use of administrative data, our centre created an 'Administrative Data Working Group.' This group is tasked with equipping researchers with the knowledge to rigorously analyze such data. To date, we have focused our discussions on the selection of case definitions.

**Results:** We put forth the following **recommendation**: when available, researchers should use case definitions that are tailored to their jurisdiction; we use definitions from the BC Ministry of Health (BC-MoH). Unlike definitions based on combinations of codes from other settings, the BC-MoH's definitions consider factors that, if ignored, introduce misclassification bias. For example, known differences in the **detail** of outpatient (limited to three-digits) and hospitalization (four- or five-digits) ICD codes in BC are incorporated. **Inaccuracies** in coding are also considered: e.g., the BC-MoH's definitions often require two outpatient codes to appear within a single year. Lastly, **BC-specific codes**, such as the '50B' code for 'anxiety/depression,' are included; omission of '50B' results in substantial under-ascertainment.

**Conclusion:** **Context is everything**: the BC-MoH's definitions incorporate information that is critical to the validity of administrative data-derived measures in our jurisdiction. While published validation studies from other settings may present highly sensitive and specific definitions, we have found their direct applicability to the BC context to be limited.

## Context is everything

- **BC Ministry of Health** definitions incorporate information critical to the validity of admin data-derived measures in BC
  - i.e., Differences in the **detail & accuracy** of ICD diagnostic codes in outpatient **vs.** hospital data
  - i.e., Inclusion of **BC-specific** codes
- **Conclusion:** think carefully before using a definition from another country, province, etc., in your studies