CanPath Student Dataset

Dr. Trevor Dummer

National Scientific Co-Director, CanPath
Canadian Cancer Society Chair in Cancer Primary Prevention
Associate Professor, University of British Columbia

Dr. Jennifer Brooks & Dorothy Apedaile

Dalla Lana School of Public Health, University of Toronto

CanPath

Canadian Partnership for Tomorrow's Health

Partenariat canadien pour la santé de demain

Today's Agenda

1) Overview of CanPath
2) Overview of the Student Dataset
3) Example of how it has been used
4) Student project presentation
5) Process to obtain the student dataset

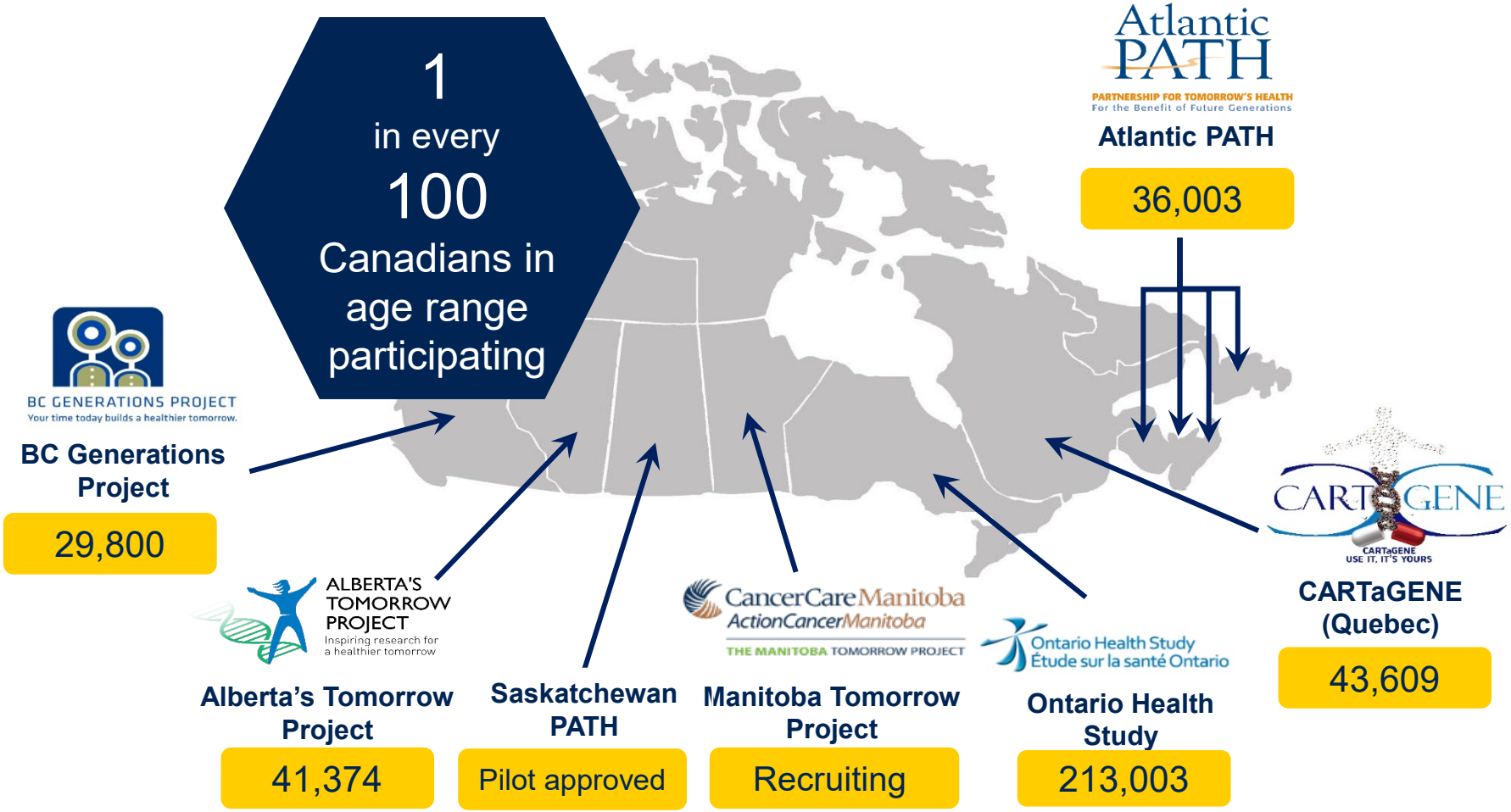
Canada's largest population health research platform

CanPath
Canadian Partnership for Tomorrow's Health

CanPath is a **population-health research platform** for assessing the effect of genetics, behaviours, family health history and environment on chronic diseases.

We aim to provide a national platform for population-level health research in Canada and globally.

CanPath brings together seven regional cohorts across ten provinces



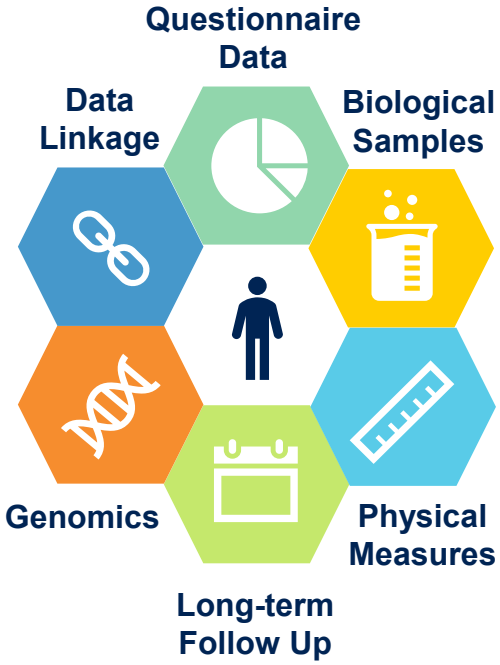
CanPath

Canada's largest population health research platform

330,000+ Canadian Adults (Age 30-74)

+

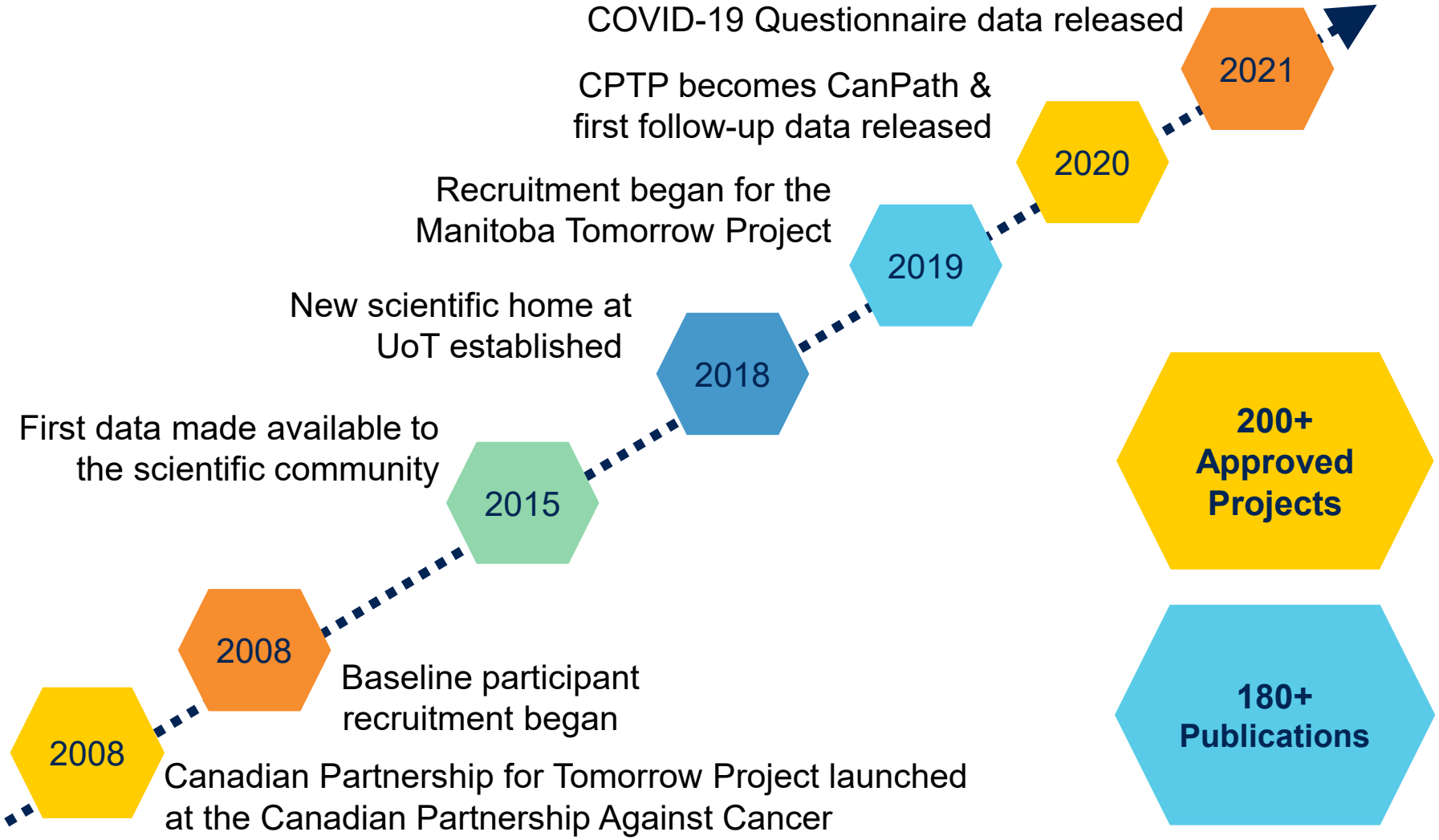
+



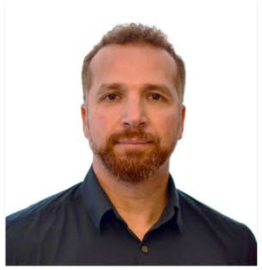
CanPath has **enabled the study of the biology, behaviour and environments of over 330,000 Canadians for over a decade** and continues to reveal hidden causes of common and rare chronic diseases and cancer.

CanPath

CanPath's value is growing over time



National Leadership Team



Philip Awadalla
***National Scientific Director,* CanPath;**
Ontario Health Study

John McLaughlin
***Executive Director,* CanPath**

Trevor Dummer
***National Scientific Co-Director,* CanPath;**
BC Generations Project

Parveen Bhatti
BC Generations Project

Shandra Harman
Alberta's Tomorrow Project

Jennifer Vena
Alberta's Tomorrow Project

Riaz Alvi
Saskatchewan PATH

Donna Turner
The Manitoba Tomorrow Project

Philippe Broët
CARTaGENE

Simon Gravel
CARTaGENE

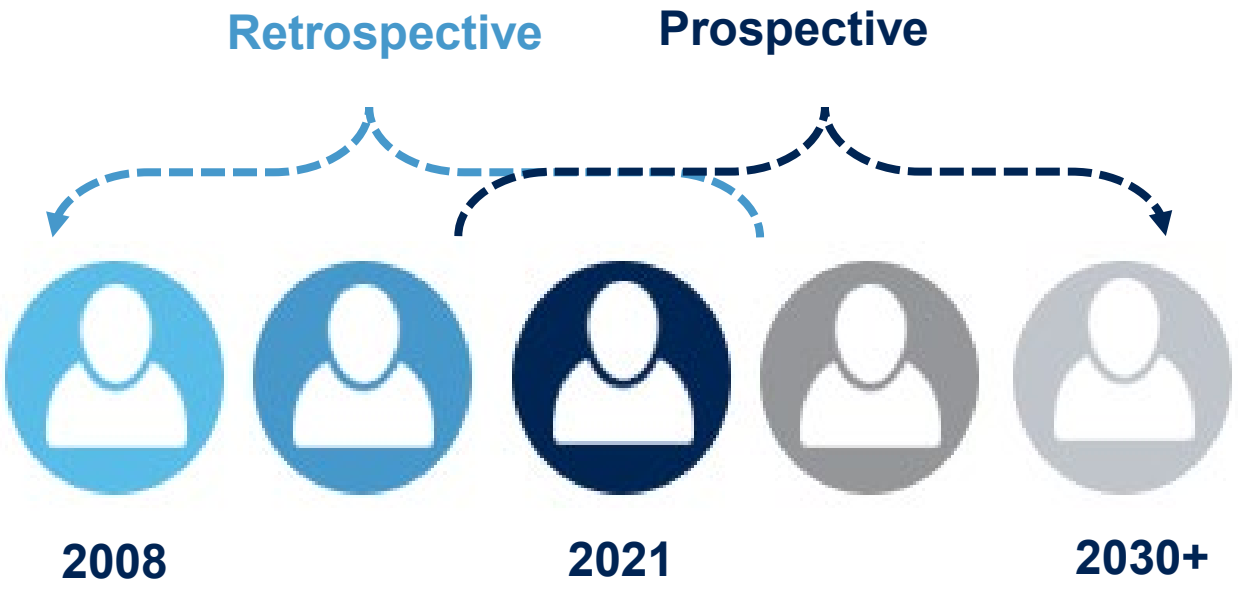
Guillaume Lettre
CARTaGENE

Robin Urquhart
Atlantic PATH

Jason Hicks
Atlantic PATH

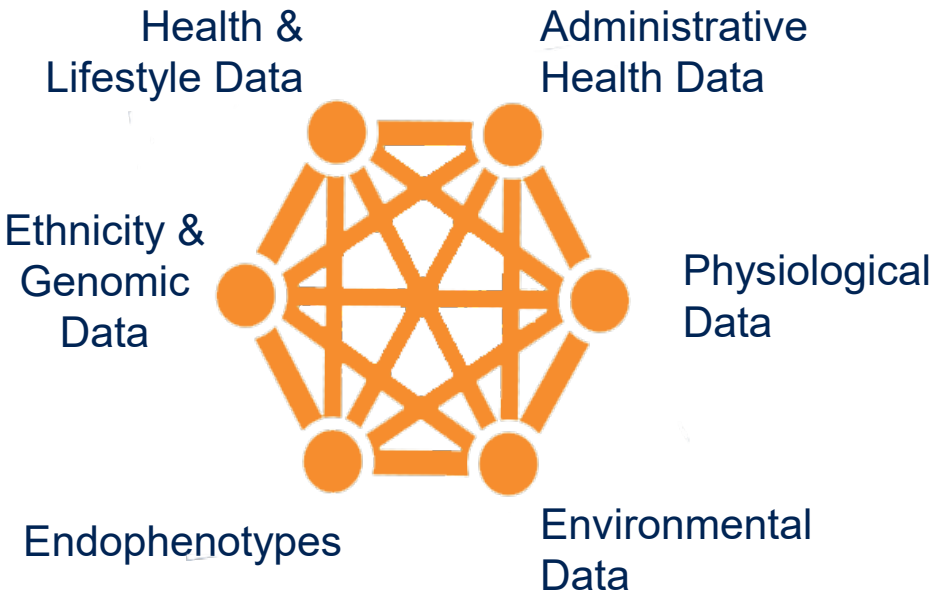
Led by a National Coordinating Centre (NCC) based at the Dalla Lana School of Public Health, University of Toronto, and Scientific Directors of the regional cohorts

CanPath enables both retrospective and prospective research



- CanPath participants will be followed for over 50 years (~ a decade of follow-up already!)
- Longitudinal cohorts enable scientists to perform health-related research for **today and for those in years to come**

CanPath has captured a broad range of data to enable the study of health and disease



CanPath is **linking personal, behavioural, environmental, health system and biological data** to investigate cancer and chronic disease causes and determinants.

All CanPath participants completed detailed baseline health and risk factor questionnaires


Participant demographics


Health status


Medical history


Prescribed medications


Family health history


Anthropometric measurements


Working status


Household income


Behaviours (sleep, alcohol, tobacco)

Harmonized Baseline Questionnaire data available - see the CanPath Portal

Collecting biological data

Venous Blood
(145,760)

Saliva
(18,799)

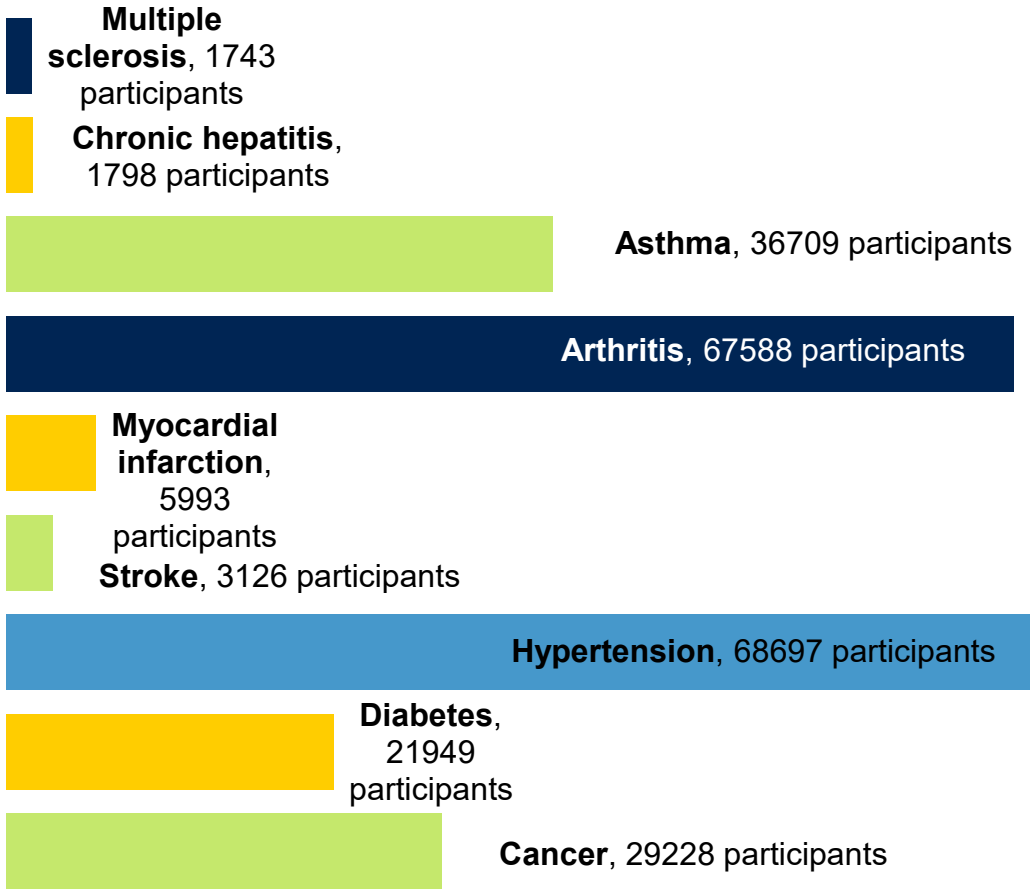
Toenail clippings
(30,000)

Urine
(99,500)

Blood Spots
(10,642)

DNA Source Material
(178,065)

Common prevalent diseases and conditions



Including participants with various chronic diseases and conditions (at baseline)

The Canadian Partnership for Tomorrow Project: a pan-Canadian platform for research on chronic disease prevention

Trevor J.B. Dummer PhD, Philip Awadalla PhD, Catherine Boileau PhD, Camille Craig MSc, Isabel Fortier PhD, Vivek Goel MD, Jason M.T. Hicks MSc, Sébastien Jacquemont MD, Bartha Maria Knoppers PhD, Nhu Le PhD, Treena McDonald MSc, John McLaughlin PhD, Anne-Marie Mes-Masson PhD, Anne-Monique Nuyt MD, Lyle J. Palmer PhD, Louise Parker PhD, Mark Purdue PhD, Paula J. Robson PhD, John J. Spinelli PhD, David Thompson MSc, Jennifer Vena PhD, Ma'n Zawati LLM; with the CPTP Regional Cohort Consortium*

Canadian Medical Association Journal, 2018, **190**: E710-7.

Over 120,000 CanPath Participants have completed detailed follow up questionnaires

- Participant demographics
- Health status
- Mental Health*
- Medical history
- Prescribed medication
- Family health history
- Anthropometric measurements
- Working status
- Household income
- Behaviours (sleep, alcohol, tobacco, marijuana use*, and e-cigarette use*)

Follow-up Questionnaire data are available - see the CanPath Portal

CanPath

Over 101,000 CanPath Participants completed COVID-19 questionnaires


COVID-19 test result/suspected infection


Symptoms experienced (if any)


Participant hospitalized or received medical care


Current health status and risk factors for COVID-19


Potential source of exposure


Impact of pandemic on job status


Impact of the pandemic on mental, emotional, social and financial wellbeing

101,595

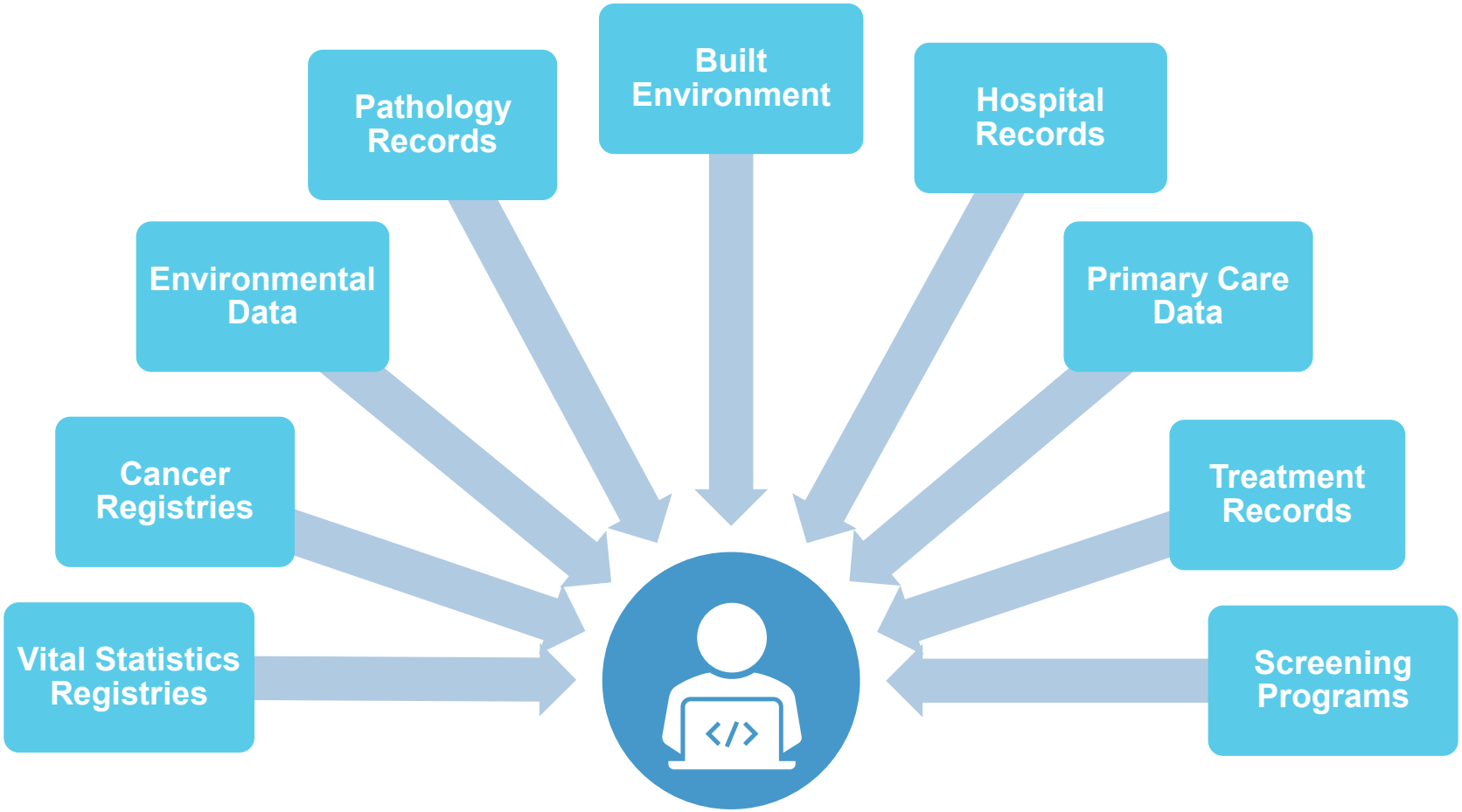
COVID-19 Questionnaires Completed

Data now available to researchers

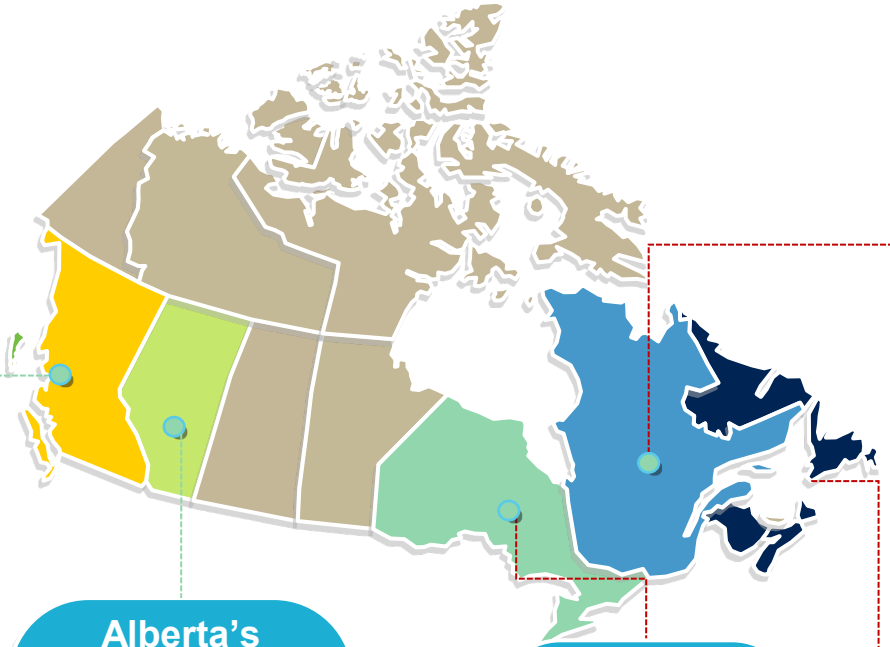
CanPath is a member of
The COVID-19 Host Genetics Initiative

CanPath

Data linkages enable us to evaluate our cohort in real-time



Administrative health linkages can be completed within regional cohorts



BC Generations Project

- BC Ministry of Health
- BC Vital Statistics Agency
- BC Cancer

Alberta's Tomorrow Project

- Alberta Health Services
- Alberta Health
- Alberta Cancer Registry

Ontario Health Study

- ICES
- Cancer Care Ontario

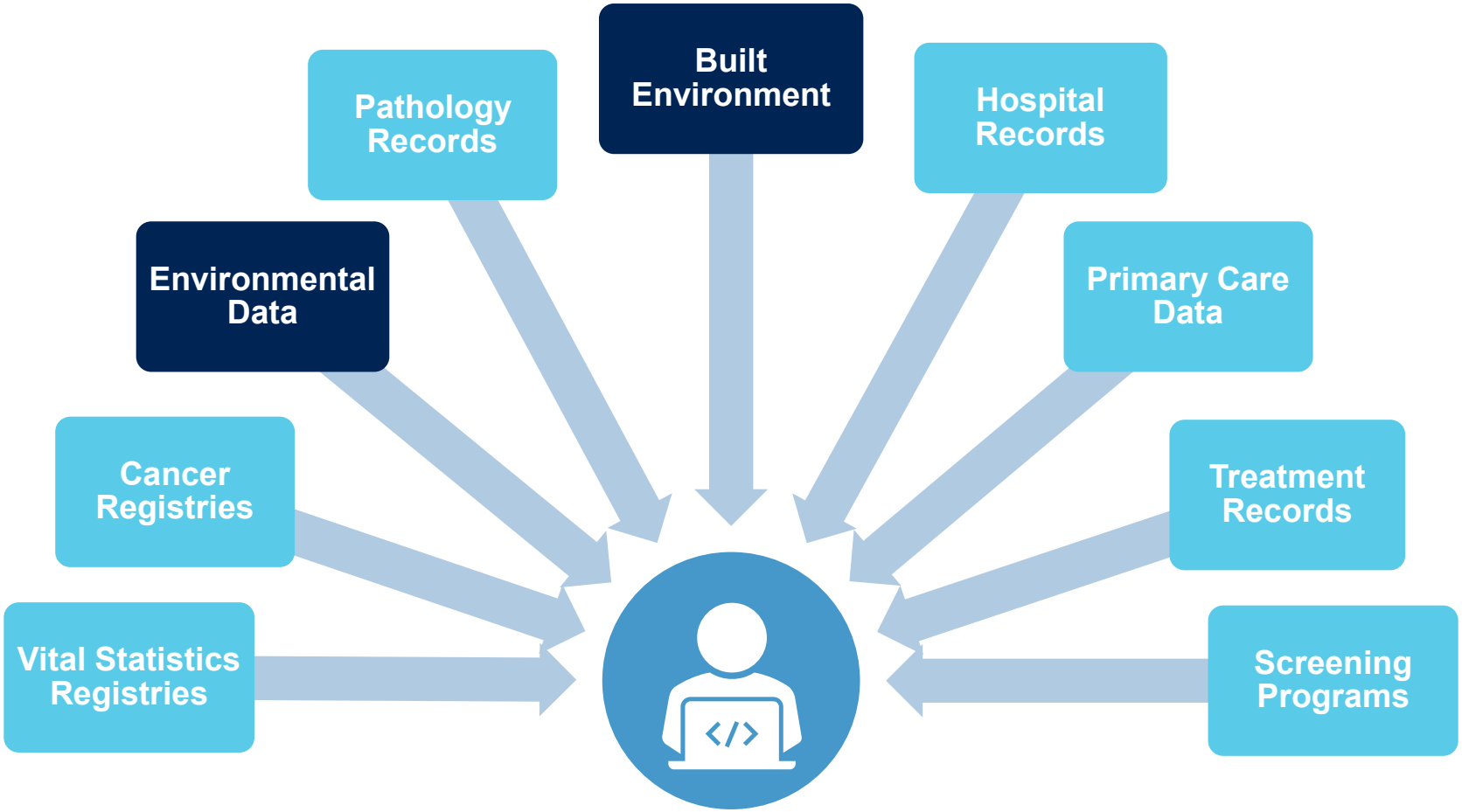
CARTaGENE

- Institut de la statistique du Québec (ISQ)
- Tumor Registries

Atlantic PATH

- Newfoundland & Labrador Cancer Registry

Data linkages enable us to evaluate our cohort in real-time



The Canadian Urban Environmental Health Research Consortium (CANUE)

- All CanPath participants have been linked to CANUE environmental exposures
- Every location in Canada can be described by a complex set of environmental factors
- CANUE is building the capacity to study how these multiple **environmental factors are linked to a wide range of health outcomes**

CanPath

Canadian Partnership for Tomorrow's Health

Student Dataset

Purpose of the Student Dataset

- Provide students the unique opportunity to gain **hands-on experience** working with CanPath data and with data analysis
- Provide an accessible resource for faculty at Canadian institutions
- Raise awareness of CanPath data to future Canadian researchers
- Protect the privacy and confidentiality of CanPath participants
  - **Data in CanPath is de-identified to ensure privacy and confidentiality of participant data.** There are strict guidelines around data access. The student dataset adds an additional element to enable use of a sample of the data in a teaching environment

Development of the Student Dataset

- To ensure the privacy of the CanPath data, software was used to create a synthetic individual-level dataset that preserves important statistical information (i.e. relationships between variables)
- The synthetic dataset is essentially a random sample of CanPath data where participant information has been rearranged
- Created using the R Package ‘synthpop’
  - This package was explicitly designed to generate synthetic versions of longitudinal survey data

Advantages of the CanPath Student Dataset

- ✓ Large sample size (Over 41,000 participants)
- ✓ Real-world population-level Canadian data
- ✓ Variety of areas of information allowing for a wide range of research topics
- ✓ **No access cost to faculty**
- ✓ Potential for students to apply for access to full CanPath data to develop their work and publish their findings

Participant data from the five founding CanPath cohorts

BC Generations Project

Atlantic PATH

Alberta's Tomorrow Project

Ontario Health Study

CARTaGENE

Baseline data from the Manitoba Tomorrow Project and Saskatchewan PATH could be added once recruitment is completed.

Baseline Questionnaire Data


Participant demographics


Health status


Medical history


Prescribed medication


Family health history


Anthropometric measurements


Working status


Household income


Behaviours (sleep, alcohol, tobacco, nutrition)

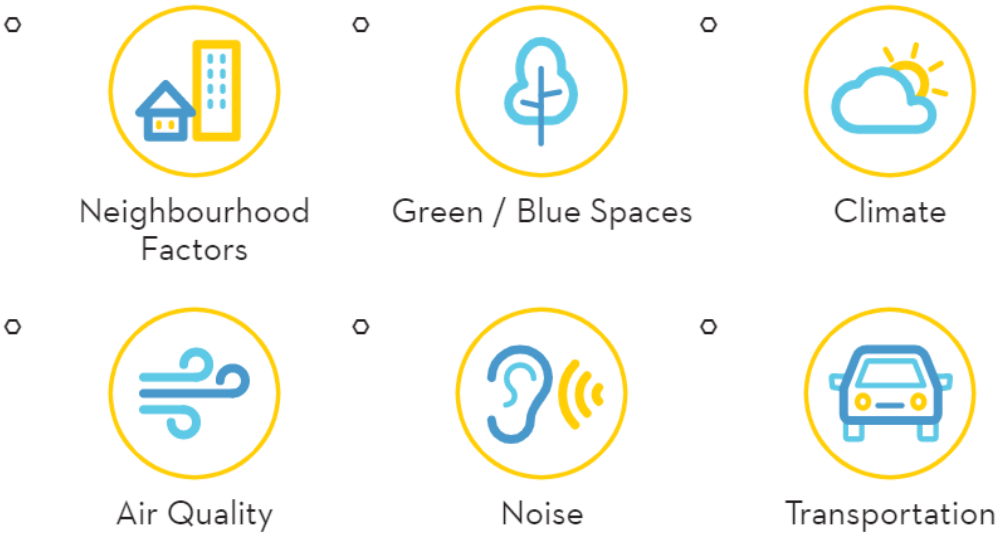
Additional Diseases Questionnaire Data

- The Additional Diseases harmonized dataset includes variables collected at baseline by three of the five CanPath population-based cohorts

- This harmonized dataset includes information on:
  - Personal and family history of diseases other than those captured in the Baseline Health and Risk Factor Questionnaire
  - Hearing health
  - Visual health
  - Oral health

CANUE Data

The CIHR-funded Canadian Urban Environmental Health Research Consortium (CANUE) collates and generates standardized area-level environmental data on:



The Student Dataset includes variables such as:

- Material deprivation index
- Annual average exposure to ambient air pollution

Summary of what's included

The student dataset includes:

- A sample of over 40,000 participants
- 403 categorical variables
  - CanPath Baseline Questionnaire Data
  - CanPath Additional Diseases Questionnaire Data
  - CANUE Data
- Data from 5 cohorts spanning 8 provinces

CanPath variables include:

- Socio-demographic and economic information
- Lifestyle and behaviour (e.g. tobacco use, alcohol use, nutrition)
- Perception of health
- Select self-reported diseases (e.g. high blood pressure, arthritis, and first cancer

For training purposes only

- The CanPath Student Dataset is a synthetic version of the CanPath data and is for training purposes only and **cannot be used for publication**.
- Students interested in finding out if their project results can be replicated using the real CanPath data for potential publication can apply for data access through the regular CanPath Access Process
- A reduced fee is available to students and trainees applying for access to CanPath data and biosamples

Submit Access Application

through the CanPath Portal

Student Dataset Pilot Study

Dr. Jennifer Brooks
Assistant Professor of Epidemiology
Dalla Lana School of Public Health
University of Toronto

CanPath
Canadian Partnership for Tomorrow's Health
Partenariat canadien pour la santé de demain

Outline

- Categorical Data Analysis for Epidemiologists
- Term-long projects
- CanPath student data set
- How the students used the dataset
- Student feedback

Categorical Data Analysis

- Second year course for MPH Epidemiology Students initially designed by Dr. Laura Rosella (DLSPH)
- Weekly lectures and computer labs using SAS
- Term-long data analysis project (done in pairs)
- Course is designed to introduce epidemiology students with some background in basic statistical analysis to the principals and methods of categorical data analysis relevant to epidemiological studies, with an emphasis on application and interpretation.

Term-long Project

- This project provides students with the opportunity to apply concepts learned in lectures and computer tutorials.
- Students work with a partner to develop a research question, identify a data source, complete the data analysis, interpret and present the results

- Scaffolded over the course of the term:
  - Brief description of the project
  - **Part 1**: Conceptual model (what is their exposure and outcome of interest, relevant confounders, effect modifiers, mediators)
  - **Part 2**: Analytic plan (including shell tables), manuscript outline
  - **Part 3**: Final manuscript (AJE format) and SAS code

- Lectures and tutorials support the execution of the project.

Term-long Project: Data Sources

- Students typically use:
  - **Canadian community health survey (CCHS)**
  - **NHANES**
- But also:
  - Ontario Tobacco Research Unit data
  - The Canadian Study of Diet, Lifestyle and Health (CSDLH)
  - National Health Interview Survey
  - National Longitudinal Survey of Children and Youth
  - Canadian Alcohol and Drug Use Monitoring Survey
  - Data sets available through prior collaborations (practicum projects, research assistants – must be a distinct project)

- Many of these datasets are available to U of T students through the Computing in the Humanities and Social Sciences (CHASS) Data Centre

Term-long Project: REB

- Blanket REB for a course-based analyses
- Secondary analysis, no new data is being collected for the purpose of these analyses
- Data is either publicly available (e.g., NHANES) or available for download by U of T students through CHASS (e.g., CCHS)
- If using project specific data, then must be covered by REB through that PI
- If they wish to publish their findings, they have to submit a new REB indicating that the project is moving beyond educational purposes to that of research

Types of projects done:

- Students develop their research question (with their partner)

- **The only requirements are that**:
  - The outcome is categorical

  - Must apply an analytic approach introduced in lecture
    - E.g., logistic regression, multinomial logistic regression, log binomial, Poisson, etc.

  - Exposures and outcomes can be **health**-related (e.g., health condition/disease or lifestyle factor), **social** factors (e.g., marital status, number of individuals living in the home, sense of well being) or **economic** factors (e.g., education, household income).

Fall of 2020: CanPath Pilot

- Data set includes:
  - >40,000 participants
  - >400 variables (baseline questionnaire, CANUE)
- I controlled access through U of T One Drive
- Came with a data dictionary

- **Student projects using CanPath data included:**
  - Anxiety and migraines
  - Dietary fruits and vegetables and CRC
  - Green space and obesity
  - Passive exposure to cigarette smoking during childhood and MS
  - Anxiety and addiction
  - IVF and CVD
  - Education and blood pressure
  - **Work schedule and binge drinking**

Student Feedback

- Students completed a short survey (4 of 8 groups responded):
    - Found the data easy to use
    - Some had to modify their research question based on available data
    - Appreciated the sample size and number of variables
    - Data was clearly formatted, good data dictionary
    - Found website helpful (e.g., understanding the structure of the questionnaire)
    - They would recommend to other students
- No major issues identified
    - Most common comment was availability of different variables (i.e., the dataset doesn't include everything)

Shift Work and Binge Drinking Among Working Canadian Adults

Dorothy Apedaile
Dalla Lana School of Public Health

CanPath
Canadian Partnership for Tomorrow's Health
Partenariat canadien pour la santé de demain

Introduction

- Approximately one third of employed Canadians work non-standard hours (shifts)
- Shift work is associated with several negative physical and mental health outcomes, including cardiovascular disease, cancer, depression, and sleep disruptions

- Some studies have found that shift workers report higher alcohol use than non-shift workers, potentially as a sleep aid
- Binge drinking has been associated with organ damage and increased risk of cancer

- **Objective**: **estimate the association between work schedule and the frequency of binge drinking in working Canadian adults**

Methods

- **Outcome: Frequency of binge drinking in the past 12 months**
  - Non-drinker
  - Never (reference)
  - Once a month or less
  - More than once a month
- **Exposure: work schedule**
  - Regular day workers
  - Shift workers (evening shift, night shift, rotating shift, split shift, irregular/on call)
- **Statistical analysis: Multinomial logistic regression**
  - Stratified by sex
  - Covariate adjustment based on a set of variables identified through literature review and purposeful selection using the Hosmer-Lemeshow-Sturdivant method
  - Sensitivity analysis excluding the non-drinkers

Results

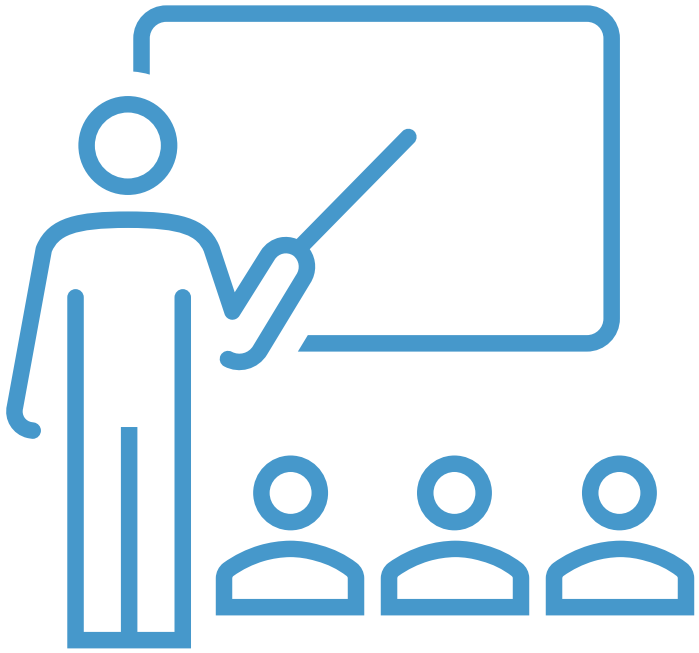
- Final sample size of **17,064** after restricting to working participants with complete outcome and exposure data
- 94% of participants reported drinking in the past year, 55% reported binge drinking at lease once in the past year
- 25% of participants were shift workers

	Binge drank ≤1 times a month[a] (aOR[b], 95% CI)	**Binge drank >1 times a month[a]** (aOR[b], 95% CI)	**Did not drink at all[a]** (aOR[b], 95% CI)
Men Shift work vs Regular Day schedule	0.97 (0.83, 1.14)	1.27 (1.04, 1.54)	1.46 (1.12, 1.91)
Women Shift work vs Regular Day schedule	1.02 (0.93, 1.11)	1.09 (0.96, 1.25)	0.99 (0.81, 1.21)

[a] Reference group: "did not binge drink"
[b] Adjusted for marital status, age (non-linear), highest education completed, number of children in household, country of birth, sleeping difficulties, and region

Lessons Learned

- Experience cleaning a “real” dataset and operationalizing complicated exposure and outcome variables
- Experience dealing with missing data
- Great opportunity to get familiar with the baseline CanPath data



CanPath

Canadian Partnership for Tomorrow's Health

How to Obtain the Dataset

Student Dataset Access Process

Eligibility Criteria:

- Applicant must be an instructor at a Canadian university or college
- The dataset is being requested for use in an academic course
- The course objectives are relevant to CanPath's purpose, vision and mission (https://canpath.ca/mission/)
- The CanPath dataset aligns with the course objectives and methods

Student Dataset Access Process

Required Documents

1. Completed Application Form (available on CanPath website)
2. Copy of REB Application
3. REB decision letter *or* proof of exemption
4. Brief CV of Applicant (2 pages)
5. Course Syllabus

- Completed applications and supporting documents can be submitted by email to access@canpath.ca.
- Applications will be reviewed within two weeks.
- Approved applicants will need to sign a data access agreement to obtain the dataset.

https://canpath.ca/student-dataset/

Student Dataset

CanPath's Student Dataset provides students the unique opportunity to gain hands-on experience working with CanPath data.

ON THIS PAGE:

What is the Student Dataset?

CanPath has developed a Student Dataset that provides students the unique opportunity to gain hands-on experience working with CanPath data. The CanPath Student Dataset is a synthetic dataset that was manipulated to mimic

Need Help? Questions?

Email access@canpath.ca

Thank you to CanPath participants across the regional cohorts who generously donate their time, information and biological samples. **CanPath is a success because of the participants' ongoing commitment.**

Thank you to CanPath's sponsors and hosts!

CanPath.ca

CanPath

Canadian Partnership
for Tomorrow's Health

Partenariat canadien
pour la santé de demain