

Ensemble de données pour étudiants

CanPath

Dr Trevor Dummer

Codirecteur scientifique national, CanPath

Titulaire de la chaire de la Société canadienne du cancer en prévention primaire du cancer

Professeur agrégé, Université de la Colombie-Britannique

Dre Jennifer Brooks et Dorothy Apedaile

École de santé publique Dalla Lana, Université de Toronto



CanPath

Canadian Partnership
for Tomorrow's Health

Partenariat canadien
pour la santé de demain





Ordre du jour

- 1) Aperçu de CanPath
- 2) Aperçu de l'Ensemble de données pour étudiants
- 3) Exemple de la façon dont il a été utilisé
- 4) Présentation d'un projet d'étudiant
- 5) Processus pour obtenir l'Ensemble de données pour étudiants

La plus vaste plateforme de recherche sur la santé de la population au Canada



Partenariat canadien
pour la santé de demain

CanPath est une **plateforme de recherche sur la santé de la population** visant à évaluer l'effet de la génétique, des comportements, des antécédents de santé familiaux et de l'environnement sur les maladies chroniques.

Nous visons à fournir une plate-forme nationale pour la recherche sur la santé des populations au Canada et dans le monde.

CanPath est composé de sept cohortes régionales dans dix provinces

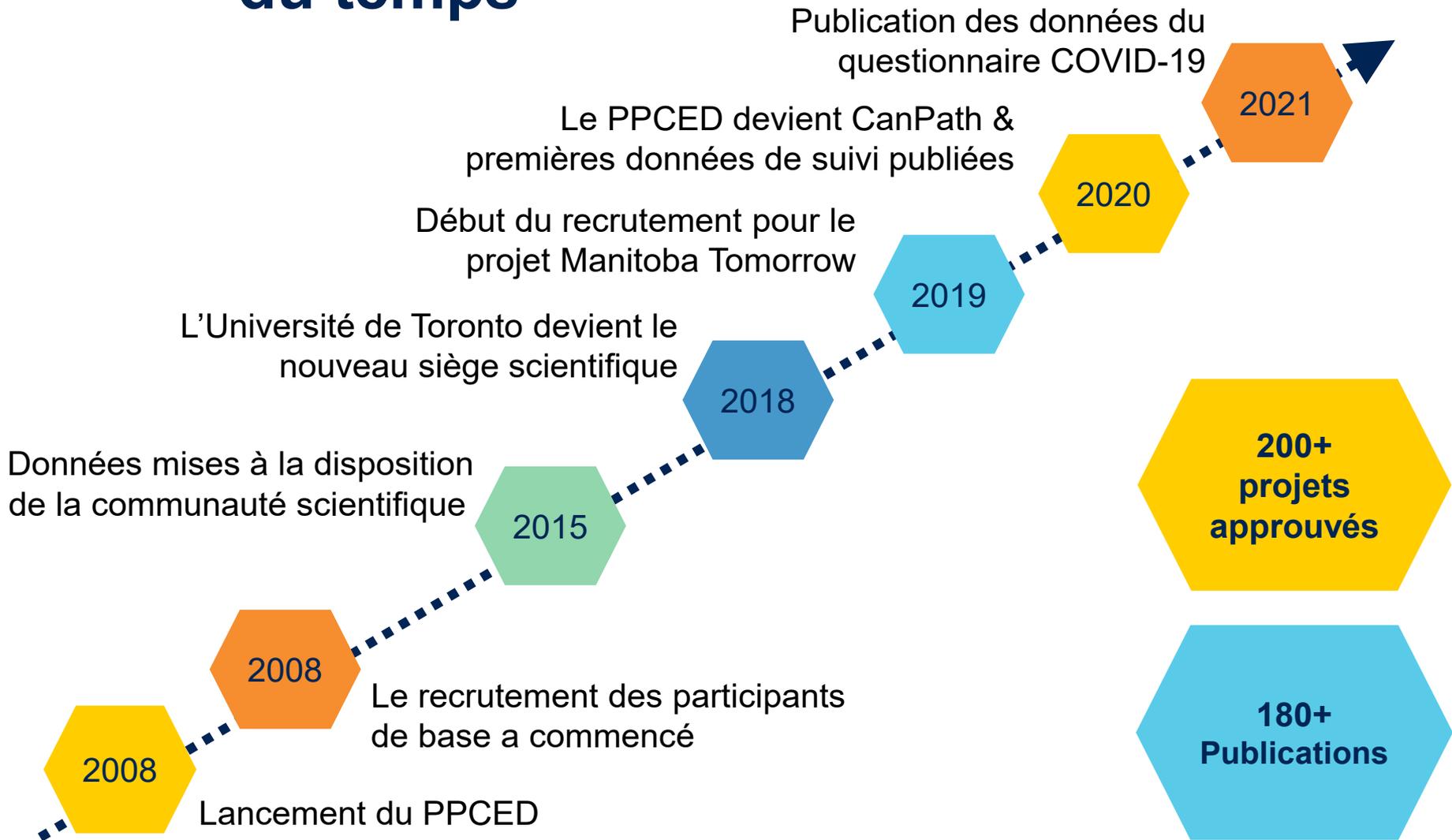


La plus vaste plateforme de recherche sur la santé de la population au Canada



CanPath a permis l'étude de la **biologie, du comportement et de l'environnement** de plus de 330 000 Canadiens depuis plus d'une décennie et continue de révéler les causes cachées des maladies chroniques courantes et rares et du cancer.

La valeur de CanPath augmente au fil du temps



Équipe de direction nationale



Philip Awadalla
Directeur scientifique national;
Étude sur la santé Ontario



John McLaughlin
Directeur exécutif



Trevor Dummer
co-directeur scientifique national;
Projet BC Generations



Parveen Bhatti
Projet BC Generations



Shandra Harman
Projet Alberta Tomorrow



Jennifer Vena
Projet Alberta Tomorrow



Riaz Alvi
Saskatchewan PATH



Donna Turner
Projet Manitoba Tomorrow



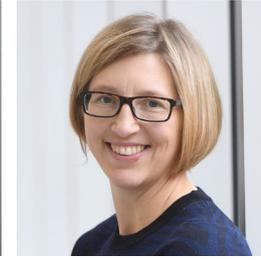
Philippe Broët
CARTaGENE



Simon Gravel
CARTaGENE



Guillaume Lettre
CARTaGENE



Robin Urquhart
La VOIE Atlantique



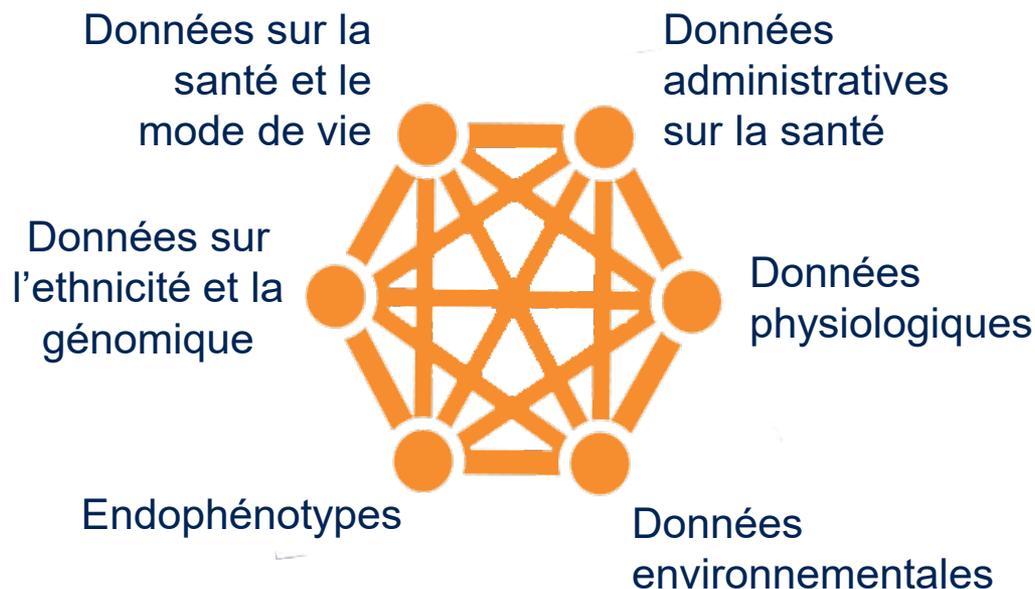
Jason Hicks
La VOIE Atlantique

CanPath permet une recherche rétrospective et prospective



- Les participants à CanPath seront suivis pendant plus de 50 ans (une décennie écoulée déjà!).
- Les cohortes longitudinales permettent aux scientifiques de faire des recherches sur la santé des gens, **aujourd'hui et pour des années à venir.**

CanPath saisit un large éventail de données pour permettre l'étude de la santé et de la maladie



CanPath relie les données personnelles, comportementales, environnementales, biologiques et sur le système de santé afin d'étudier les causes et les déterminants du cancer et des maladies chroniques.

Tous les participants à CanPath ont rempli des questionnaires de base détaillés sur la santé et les facteurs de risque



Données démographiques des participants



État de santé



Antécédents médicaux



Médicaments prescrits



Antécédents de santé familiaux



Mesures anthropométriques



Situation professionnelle



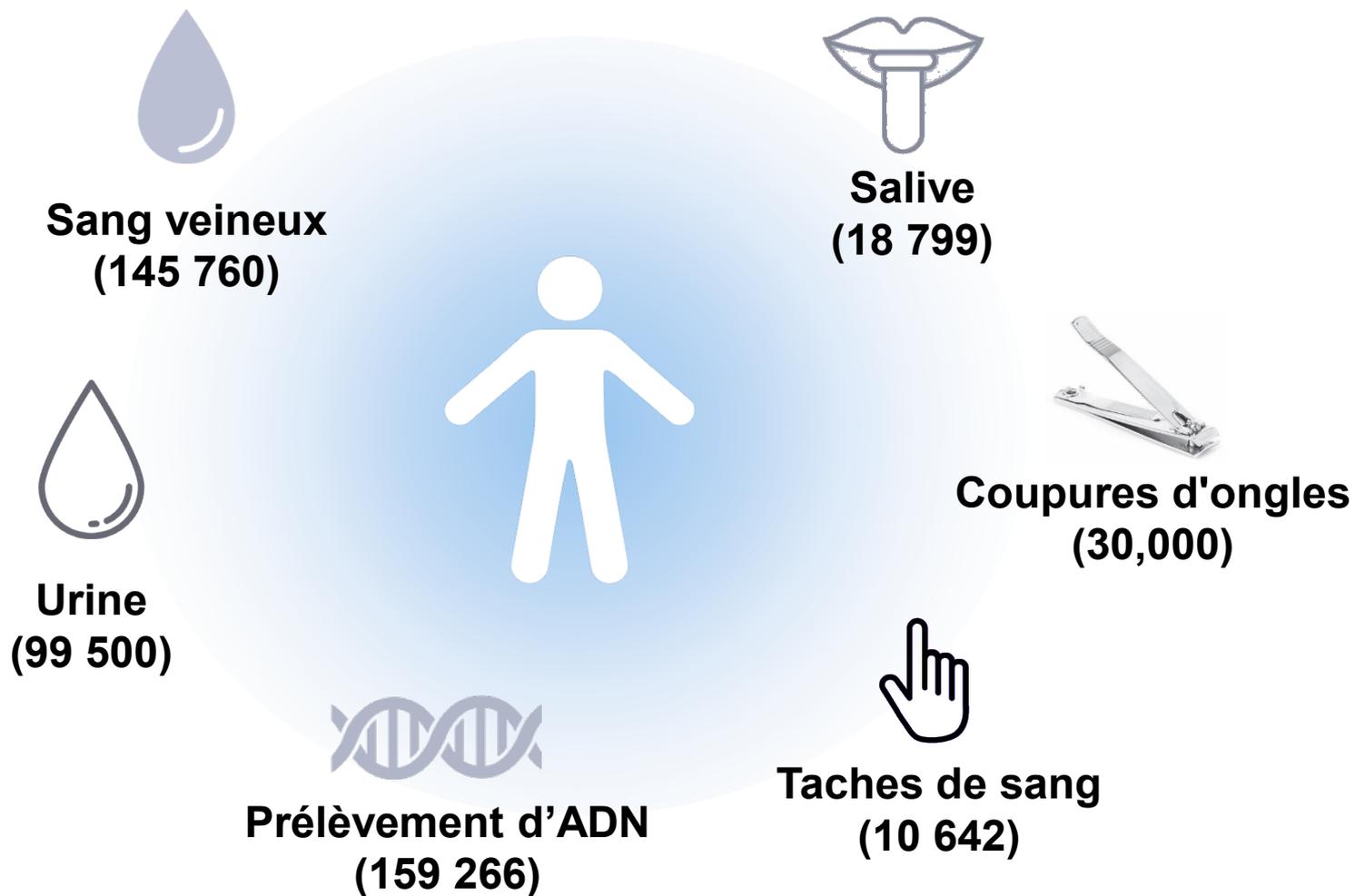
Revenu du ménage



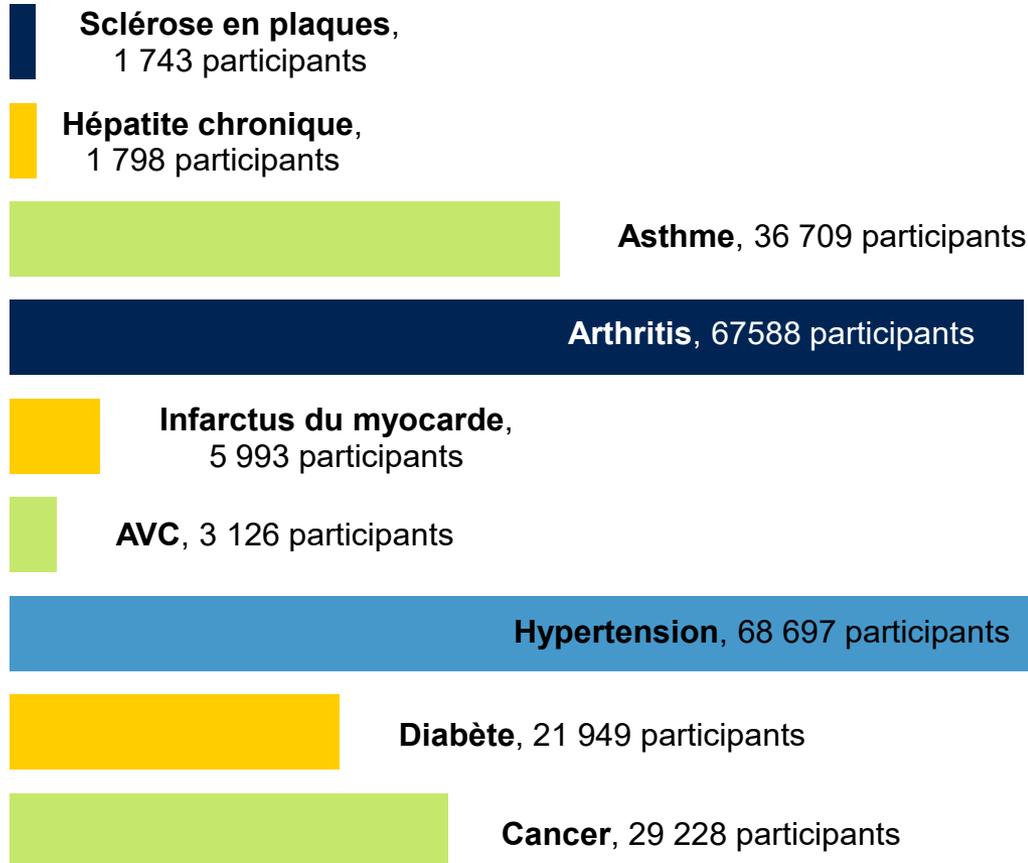
Comportement (sommeil, consommation d'alcool, de tabac)

**Données
harmonisées du
questionnaire de
base disponibles
- voir le Portail
CanPath**

Échantillonnage longitudinal des données biologiques



Maladies et affections courantes signalées dans la cohorte



Incluant les participants atteints de diverses maladies et affections chroniques (au point de départ)

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 LL.M.; with the CPTP Regional Cohort Consortium*

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

Plus de 120 000 participants à CanPath ont rempli des questionnaires de suivi détaillés



Données démographiques des participants



État de santé



Santé mentale*



Antécédents médicaux



Médicaments prescrits



Antécédents de santé familiaux



Mesures anthropométriques



Situation professionnelle



Revenu du ménage

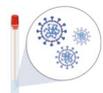


Comportement (sommeil, consommation d'alcool, de tabac, de marijuana* et de cigarettes électroniques*)



Les données du questionnaire de suivi seront disponibles sur le portail CanPath

Le questionnaire CanPath sur la COVID-19 a été conçu pour s'aligner sur les efforts internationaux



Résultat du test de la COVID-19 / infection soupçonnée



Symptômes ressentis (le cas échéant)



Le participant a été hospitalisé ou a reçu des soins médicaux



État de santé actuel et facteurs de risque pour la COVID-19



Source potentielle d'exposition



Impact de la pandémie sur la situation professionnelle



Impact de la pandémie sur le bien-être mental, émotionnel, social et financier

101,595

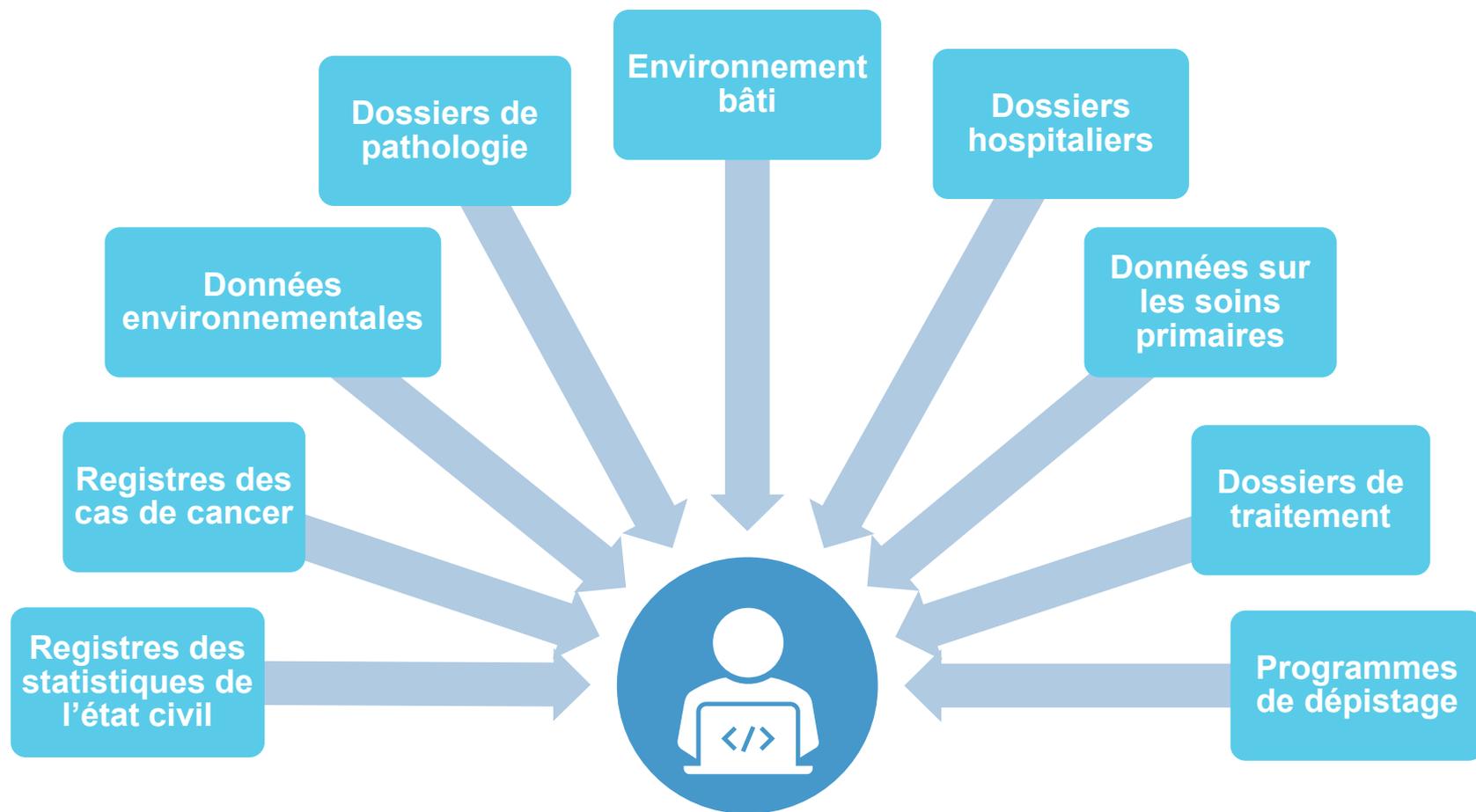
*Questionnaires COVID-19
remplis*

*Les données sont maintenant
disponibles aux chercheurs*

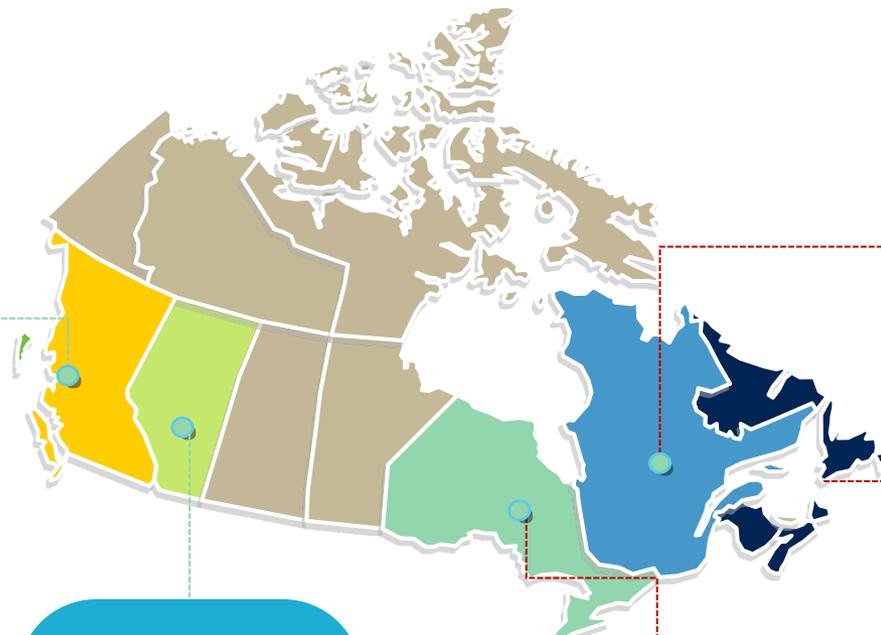
CanPath est membre de
The COVID-19 Host Genetics Initiative



Le couplage des données nous permet d'évaluer notre cohorte en temps réel



Des couplages de données administratives sur la santé peuvent être faits au sein des cohortes régionales



Projet BC Generations

- Ministère de la santé de la Colombie-Britannique
- Agence des statistiques de l'état civil de la Colombie-Britannique
- BC Cancer

Projet Alberta Tomorrow

- Services de santé de l'Alberta
- Santé Alberta
- Registre du cancer de l'Alberta

Étude sur la santé Ontario

- ICES
- Action cancer Ontario

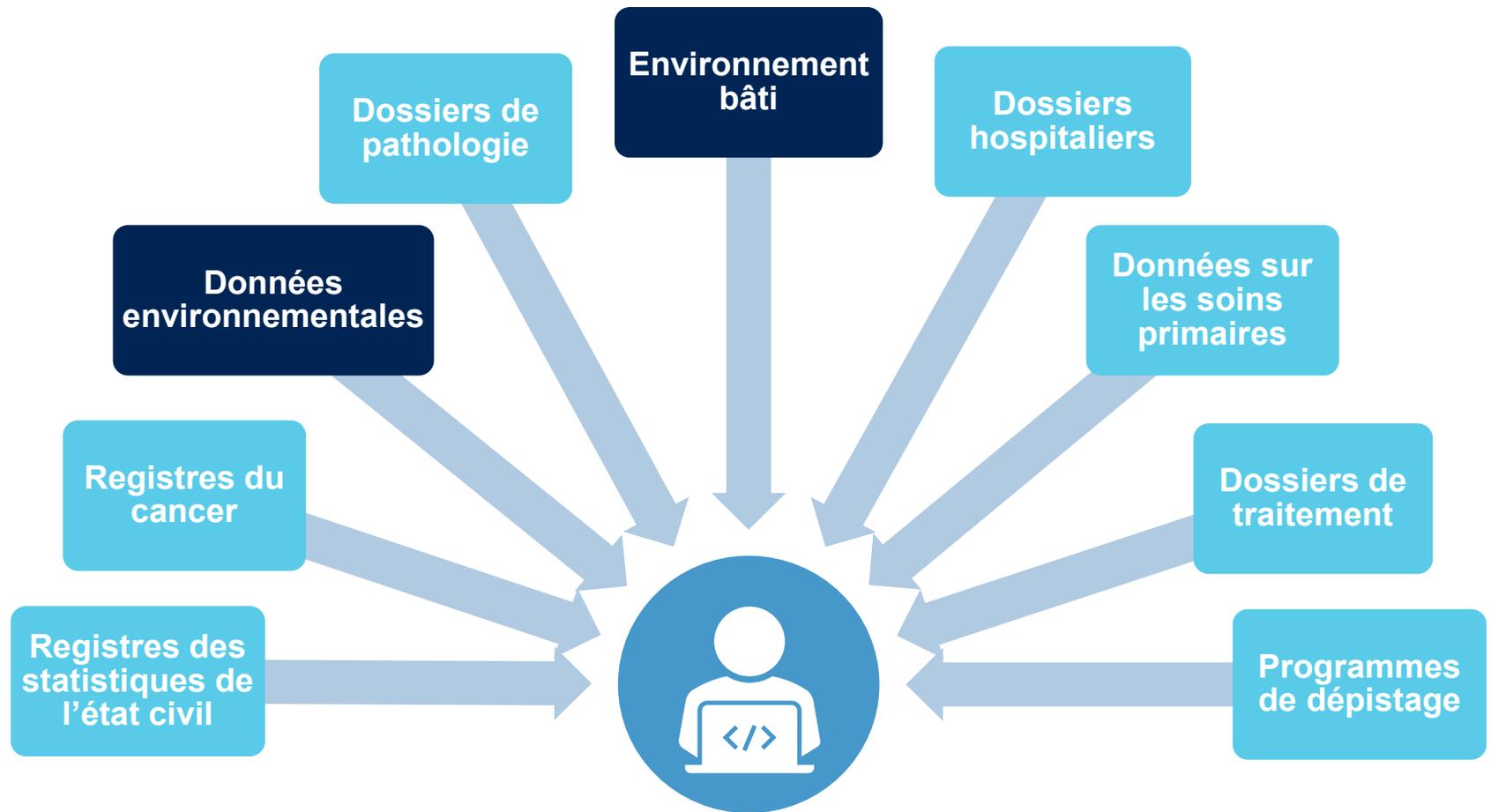
CARTaGENE

- Institut de la statistique du Québec (ISQ)
- Registres des tumeurs

La VOIE Atlantique

- Registre des cas de cancer de Terre-Neuve et Labrador

Le couplage des données nous permet d'évaluer notre cohorte en temps réel



Consortium canadien de recherche en santé environnementale urbaine

- Les données de tous les participants à CanPath ont été couplées à celles de CANUE sur les expositions environnementales.
- Chaque endroit au Canada peut être décrit par un ensemble complexe de facteurs environnementaux.
- CANUE renforce la capacité d'étudier comment ces multiples **facteurs environnementaux sont liés à un large éventail de résultats sur le plan de la santé.**





CanPath

Canadian Partnership
for Tomorrow's Health

**Ensemble de données
pour étudiants**

Objet de l'Ensemble de données pour étudiants

- Offrir aux étudiants l'occasion unique d'acquérir une **expérience pratique** en travaillant avec des données de CanPath et en faisant l'analyse de données
- Fournir une ressource accessible aux enseignants des établissements canadiens
- Faire connaître les données de CanPath aux futurs chercheurs canadiens
- Protéger la vie privée et la confidentialité des participants à CanPath :
 - **Les données de CanPath sont anonymisées pour assurer la confidentialité des données des participants.** Des directives strictes s'appliquent à l'accès aux données. L'Ensemble de données pour étudiants vient ajouter un élément supplémentaire pour permettre l'utilisation d'un échantillon de données dans un contexte d'enseignement.



Développement de l'Ensemble de données pour étudiants

- Pour assurer la confidentialité des données CanPath, un logiciel a été utilisé pour créer un ensemble de données synthétiques au niveau individuel qui préserve les informations statistiques importantes (c.-à-d., les relations entre les variables).
- L'ensemble de données synthétiques est essentiellement un échantillon aléatoire de données CanPath où les informations sur les participants ont été réorganisées
- Créé à l'aide du package R 'synthpop'
 - Cette trousse a été explicitement conçue pour générer des versions synthétiques des données d'enquête longitudinale.



Avantages de l'Ensemble de données pour étudiants de CanPath

- ✓ Large échantillon (plus de 41 000 participants)
- ✓ Données canadiennes réelles au niveau de la population
- ✓ Variété de domaines d'information permettant un large éventail de sujets de recherche
- ✓ **Aucun coût d'accès pour les enseignants**
- ✓ Possibilité pour les étudiants de demander l'accès aux données complètes de CanPath afin de développer leurs travaux et de publier leurs conclusions

Données sur les participants des cinq cohortes fondatrices de CanPath



**Projet BC
Generations**



**Atlantic
PATH**

PARTNERSHIP FOR TOMORROW'S HEALTH
For the Benefit of Future Generations

La VOIE atlantique



**Projet Alberta
Tomorrow**



**Étude sur la santé
Ontario**



CARTaGENE

Les données de base des projets Manitoba Tomorrow et Saskatchewan PATH pourraient être ajoutées une fois le recrutement terminé.

Données de base du questionnaire



Données démographiques des participants



État de santé



Antécédents médicaux



Médicaments prescrits



Antécédents de santé familiaux



Mesures anthropométriques



Situation professionnelle



Revenu du ménage



Comportements (sommeil, alcool, tabac, nutrition)

Données du questionnaire sur d'autres maladies

- L'ensemble harmonisé de données sur les autres maladies renferme des variables recueillies au départ par trois des cinq cohortes de population de CanPath.



- Cet ensemble de données harmonisé renferme des renseignements sur :
 - Les antécédents personnels et familiaux de maladies autres que celles saisies dans le Questionnaire de base sur la santé et les facteurs de risque :
 - Santé auditive
 - Santé visuelle
 - Santé bucco-dentaire

Données de CANUE

Le Consortium canadien de recherche en santé environnementale urbaine (CANUE), financé par les IRSC, recueille et génère des données environnementales normalisées au niveau régional sur :



L'Ensemble de données pour étudiants comprend des variables telles que :

- L'indice de privation matérielle
- L'exposition moyenne annuelle à la pollution de l'air ambiant



Résumé de ce qui est inclus

L'Ensemble de données pour étudiants comprend :

- Un échantillon de plus de 40 000 participants
- 403 variables catégorielles
 - Données du Questionnaire de base de CanPath
 - Données du questionnaire de CanPath sur les autres maladies
 - Données de CANUE
- Données de cinq cohortes couvrant huit provinces

Les variables CanPath englobent :

- Des renseignements socio-démographiques et économiques
- Le mode de vie et les comportements (p. ex., tabagisme, consommation d'alcool, nutrition)
- La perception de la santé
- Certaines maladies autodéclarées (p. ex., hypertension artérielle, arthrite et premier cancer)

À des fins de formation uniquement

- L'Ensemble de données pour étudiants de CanPath est une version synthétique des données de CanPath destinée uniquement à des fins de formation et **ne peut être utilisé à des fins de publication.**
- Les étudiants qui voudraient savoir si les résultats de leur projet peuvent être reproduits à l'aide des données réelles de CanPath en vue d'une publication éventuelle peuvent demander l'accès aux données via la procédure régulière d'accès à CanPath.
- Des frais réduits sont offerts aux étudiants et aux stagiaires qui demandent l'accès aux données et aux échantillons biologiques de CanPath.



Présenter une demande

sur le portail CanPath

Étude pilote sur l'Ensemble de données pour étudiants

Dre Jennifer Brooks

Professeure adjointe d'épidémiologie
École de santé publique Dalla Lana
Université de Toronto



CanPath

Canadian Partnership
for Tomorrow's Health

Partenariat canadien
pour la santé de demain





Aperçu

- Analyse de données catégorielles pour les épidémiologistes
- Projets pour la durée de la session
- Ensemble de données pour étudiants de CanPath
- Comment les étudiants ont utilisé l'ensemble de données
- Commentaires des étudiants



Analyse de données catégorielles

- Cours de deuxième année pour les étudiants au MPH en épidémiologie initialement conçu par la Dre Laura Rosella (DLSPH)
- Conférences hebdomadaires et laboratoires informatiques utilisant SAS
- Projet d'analyse de données sur l'ensemble de la session (par paires)
- Le cours est conçu pour présenter aux étudiants en épidémiologie ayant une certaine expérience en analyse statistique de base les principes et les méthodes d'analyse de données catégoriques pertinentes pour les études épidémiologiques, en mettant l'accent sur l'application et l'interprétation.

Projet pour la durée de la session

- Ce projet offre aux étudiants l'occasion d'appliquer les concepts appris dans des cours magistraux et des tutoriels informatiques.
- Les étudiants travaillent avec un/e partenaire pour élaborer une question de recherche, identifier une source de données, compléter l'analyse des données, et interpréter et présenter les résultats
- Étalés sur l'ensemble de la session :
 - Brève description du projet
 - **Partie 1** : Modèle conceptuel (quelle est leur exposition et le résultat d'intérêt, les facteurs de confusion pertinents, les modificateurs d'effet, les médiateurs)
 - **Partie 2** : Plan d'analyse (y compris les tableaux modèles) et aperçu du manuscrit
 - **Partie 3** : Manuscrit final (format AJE) et code SAS
- Des conférences et des travaux dirigés soutiennent l'exécution du projet.

Projet pour la durée de la session : sources de données

- Les étudiants utilisent généralement :
 - **L'Enquête sur la santé dans les collectivités canadiennes (ESCC)**
 - **NHANES**
- Mais aussi :
 - Les données de l'Unité de recherche sur le tabac de l'Ontario
 - La Canadian Study of Diet, Lifestyle and Health (CSDLH)
 - Le National Health Interview Survey
 - L'Enquête longitudinale nationale sur les enfants et les jeunes
 - L'Enquête de surveillance canadienne de la consommation d'alcool et de drogues
 - Les ensembles de données disponibles dans le cadre de collaborations antérieures (projets de stage, assistants de recherche - doit être un projet distinct)
- Bon nombre de ces ensembles de données sont accessibles aux étudiants de l'Université de Toronto via le Centre de données informatiques en sciences humaines et sociales (CHASS).

Projet pour la durée de la session : CER

- Couverture du CER pour les analyses dans le cadre d'un cours
- Analyse secondaire, aucune nouvelle donnée n'est recueillie aux fins de ces analyses
- Les données sont soit accessibles au public (p. ex., NHANES), soit téléchargeables par les étudiants de l'Université de Toronto via le CHASS (p. ex., ESCC)
- Si vous utilisez des données spécifiques au projet, elles doivent être couvertes par le CER via cet IP
- Si des étudiants souhaitent publier leurs résultats, ils doivent soumettre une nouvelle demande au CER en indiquant que le projet va au-delà d'un but éducatif et passe dans le domaine de la recherche.



Types de projets réalisés :

- Les étudiants élaborent leur question de recherche (avec leur partenaire)
- **Les seules exigences sont que :**
 - Le résultat soit catégoriel
 - On doit appliquer une approche analytique introduite durant le cours
 - P. ex., régression logistique, régression logistique multinomiale, méthode log-binomiale, méthode de Poisson, etc.
 - Les expositions et les résultats peuvent être liés à la **santé** (p. ex., état de santé/maladie ou facteur lié au mode de vie), à des facteurs **sociaux** (p. ex., situation matrimoniale, nombre de personnes vivant au foyer, sentiment de bien-être) ou à des facteurs **économiques** (p. ex., l'éducation, le revenu du ménage).

Automne 2020 : Projet pilote de CanPath

- L'ensemble de données comprend :
 - >40 000 participants
 - >400 variables (Questionnaire de base, CANUE)
- J'ai contrôlé l'accès au moyen de One Drive de l'U de T
- Remis avec un dictionnaire de données

- **Les projets des étudiants utilisant les données de CanPath portaient notamment sur :**
 - L'anxiété et les migraines
 - Les fruits et légumes dans l'alimentation et le cancer colorectal
 - Les espaces verts et l'obésité
 - L'exposition passive au tabagisme pendant l'enfance et la SP
 - L'anxiété et la dépendance
 - FIV et MCV
 - L'éducation et la tension artérielle
 - **L'horaire de travail et la consommation excessive d'alcool**



Commentaires des étudiants

- Les étudiants ont répondu à un court sondage (4 des 8 groupes ont répondu) :
 - Ils ont trouvé les données faciles à utiliser
 - Certains ont dû modifier leur question de recherche en fonction des données disponibles
 - Ils ont apprécié la taille de l'échantillon et le nombre de variables
 - Les données étaient clairement formatées; bon dictionnaire de données
 - Ils ont trouvé le site Web utile (p. ex., pour comprendre la structure du questionnaire)
 - Ils le recommanderaient à d'autres étudiants
- Aucun problème majeur identifié
 - Le commentaire le plus courant était la disponibilité de différentes variables (c.-à-d. que l'ensemble de données n'inclut pas tout).

Travail par quarts et consommation excessive d'alcool chez les travailleurs adultes canadiens

Dorothy Apedaile

École de santé publique Dalla Lana



CanPath

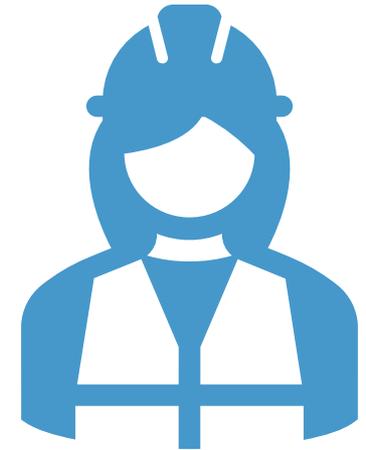
Canadian Partnership
for Tomorrow's Health

Partenariat canadien
pour la santé de demain



Introduction

- Environ le tiers des Canadiens occupant un emploi travaillent des heures atypiques (quarts de travail).
- Le travail par quarts est associé à plusieurs résultats négatifs en matière de santé physique et mentale, y compris les maladies cardiovasculaires, le cancer, la dépression et les troubles du sommeil.
- Certaines études ont révélé que les personnes travaillant par quarts font état d'une consommation d'alcool plus élevée que les autres travailleurs, potentiellement pour aider à trouver le sommeil.
- La consommation excessive d'alcool a été associée à des dommages aux organes et à un risque accru de cancer.
- **Objectif : évaluer le lien entre les horaires de travail et la fréquence de consommation excessive d'alcool chez les travailleurs adultes au Canada.**



Méthodes

- **Résultat : Fréquence de la consommation excessive d'alcool au cours des 12 derniers mois**
 - Non-buveur
 - Jamais (référence)
 - Une fois par mois ou moins
 - Plus d'une fois par mois
- **Exposition : horaire de travail**
 - Travailleurs de jour réguliers
 - Travailleurs par postes (quart de soir, quart de nuit, quart tournant, quart fractionné, irrégulier/sur appel)
- **Analyse statistique : régression logistique multinomiale**
 - Stratifiée par sexe
 - Ajustement des covariables basé sur un ensemble de variables identifiées grâce à une revue de la littérature et une sélection ciblée à l'aide de la méthode Hosmer-Lemeshow-Sturdivant
 - Analyse de sensibilité excluant les non-buveurs

Résultats

- Taille finale de l'échantillon de **17 064** après s'être limité aux participants qui travaillent et présentent des données complètes sur les résultats et l'exposition.
- 94 % des participants ont déclaré avoir consommé de l'alcool au cours de l'année précédente, et 55 % ont déclaré avoir consommé de l'alcool de façon excessive au moins une fois au cours de l'année précédente.
- 25 % des participants étaient des travailleurs par quarts.

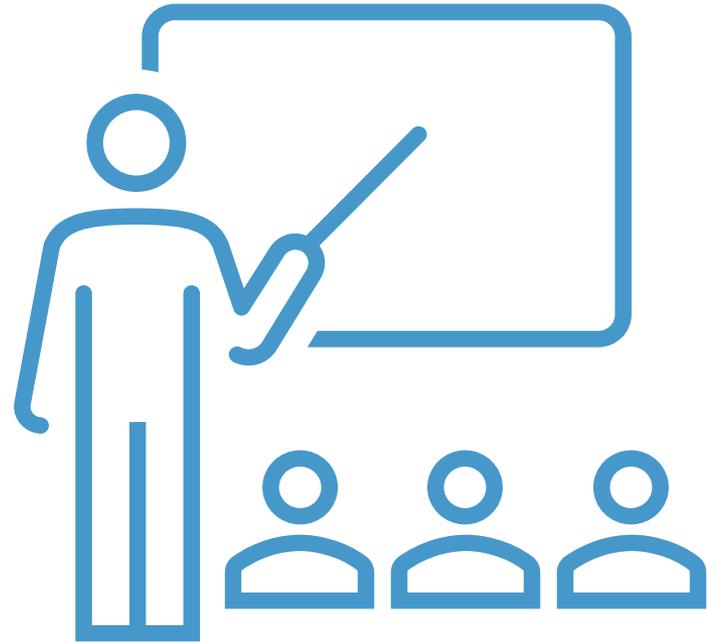
	A bu excessivement ≤1 fois / mois^a (aOR^b, IC : 95 %)	A bu excessivement >1 fois / mois^a (aOR^b, IC : 95 %)	N'a pas bu du tout^a (aOR^b, IC : 95 %)
Hommes Travail par quarts c. horaire de jour régulier	0,97 (0,83, 1,14)	1,27 (1,04, 1,54)	1.46 (1.12, 1.91)
Femmes Travail par quarts c. horaire de jour régulier	1,02 (0,93, 1,11)	1,09 (0,96, 1,25)	0,99 (0,81, 1,21)

^a Groupe de référence : « n'a pas bu excessivement d'alcool »

^b Ajusté en fonction de la situation matrimoniale, de l'âge (non linéaire), du niveau de scolarité le plus élevé atteint, du nombre d'enfants dans le ménage, du pays de naissance, des difficultés à trouver le sommeil et de la région

Leçons apprises

- Expérience de nettoyage d'un ensemble de données « réel » et d'opérationnalisation de variables d'exposition et de résultats complexes
- Expérience de traitement des données manquantes
- Excellente occasion de se familiariser avec les données de base de CanPath





CanPath

Canadian Partnership
for Tomorrow's Health

Comment obtenir l'ensemble de données

Procédure pour l'accès à l'Ensemble de données pour étudiants

Critères d'admissibilité :

- Le candidat doit enseigner dans une université ou un collège canadien.
- L'ensemble de données est demandé en vue d'être utilisé dans le cadre d'un cours d'enseignement.
- Les objectifs du cours sont pertinents à l'objet, à la vision et à la mission de CanPath (<https://canpath.ca/fr/objectif/>).
- L'ensemble de données CanPath s'aligne sur les objectifs et les méthodes du cours.

Procédure pour l'accès à l'Ensemble de données pour étudiants

Documents requis :

1. Formulaire de demande rempli ([disponible sur le site Web de CanPath](#))
2. Copie de la demande au CER
3. Lettre de décision du CER *ou* preuve d'exemption
4. Bref CV du candidat (2 pages)
5. Syllabus



- Les demandes dûment remplies et les pièces justificatives peuvent être soumises par courrier électronique à access@canpath.ca.
- Les candidatures seront examinées dans un délai de deux semaines.
- Les candidats approuvés devront signer une entente d'accès aux données pour obtenir l'ensemble de données.



L'ensemble de données pour étudiants

L'ensemble de données de CanPath pour étudiants offre les étudiants l'occasion unique d'acquérir une expérience pratique de l'utilisation des données de CanPath.

SUR CETTE PAGE :

Qu'est-ce que l'ensemble de données pour étudiants?

Quelles données sont incluses?

Processus d'accès à l'ensemble de données pour étudiants

Soumettez une demande Questions?





Besoin d'aide? Questions?

Courriel : access@canpath.ca



Merci aux participants de CanPath des six cohortes régionales qui ont généreusement donné de leur temps et fourni des renseignements et des échantillons biologiques. **Le succès de CanPath repose sur l'engagement continu des participants.**

Merci à nos commanditaires et hôtes!



<https://canpath.ca/fr/>



CanPath

Canadian Partnership
for Tomorrow's Health

Partenariat canadien
pour la santé de demain