Unveiling the Power of AI in Public Health: Insights from Diverse Case Studies and Future Directions

Hind Sbihi and Mike Irvine

November 21, 2023

BC Centre for Disease Control
Provincial Health Services Authority

Land Acknowledgement

We are gathered on the traditional, ancestral, and unceded territory of the Coast Salish peoples–Sḵwx̱wú7mesh (Squamish), Stó:lō and Səl̓ílwətaʔ/Selilwitulh (Tsleil-Waututh) and xʷməϑkʷəy̓əm (Musqueam) Nations.

BC Centre for Disease Control
Provincial Health Services Authority

Acknowledgements

Thank you to the people of British Columbia, whose data are integrated in the cohorts used for this presentation, and for whom this work is intended to benefit.

Ministry of Health and all data stewards

BC Centre for Disease Control
Provincial Health Services Authority

BRITISH COLUMBIA Ministry of Health

Provincial Health Services Authority
Province-wide solutions.
Better health.

PHSA teams:

Data & Analytics Services (BCCDC)

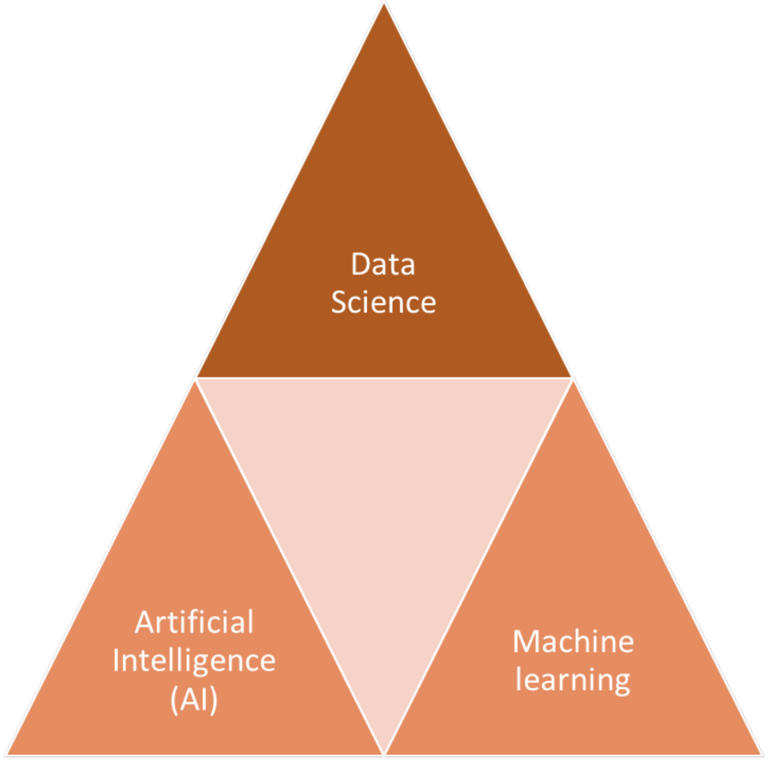
Data Analytics, Evaluation, and Reporting (PHSA)

BC-HTC team

Disclaimer

- All inferences, opinions, and conclusions drawn in this presentation are those of the authors and do not necessarily reflect the opinions or policies of the Data Steward(s)

AI in public health

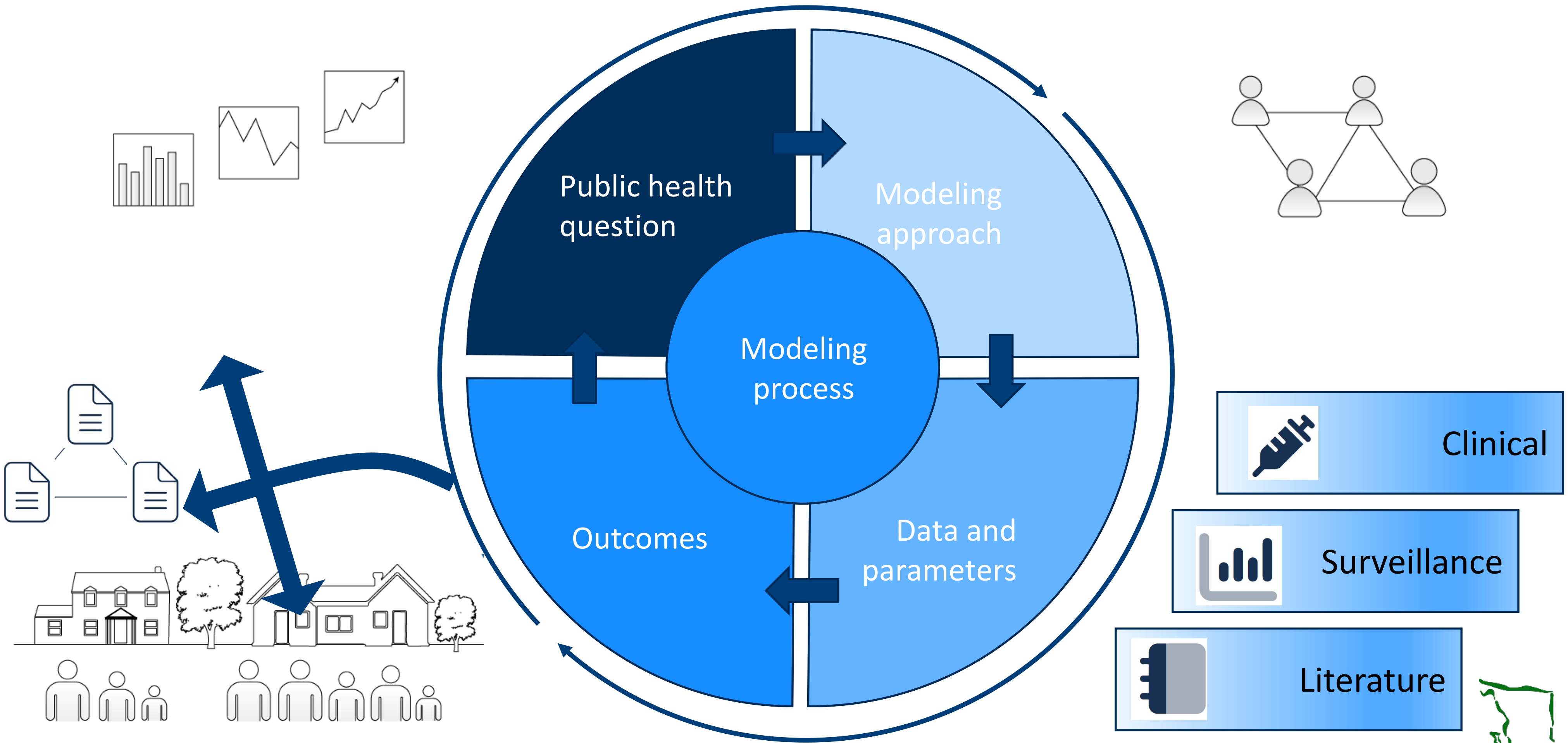


Data Science is focused on producing **insights**

Machine Learning is focused on generating **predictions**

AI is focused on making **decisions** and taking **action**

AI relies on underlying data, context, key assumptions, and communication



Provincial Health Services Authority

Adapted from: Bedson, Jamie, et al. "A review and agenda for integrated disease models including social and behavioural factors." *Nature human behaviour* 5.7 (2021): 834-846.

Implementation before the era of cloud-based surveillance: **computational phenotyping**

BC Centre for Disease Control
Provincial Health Services Authority

Background

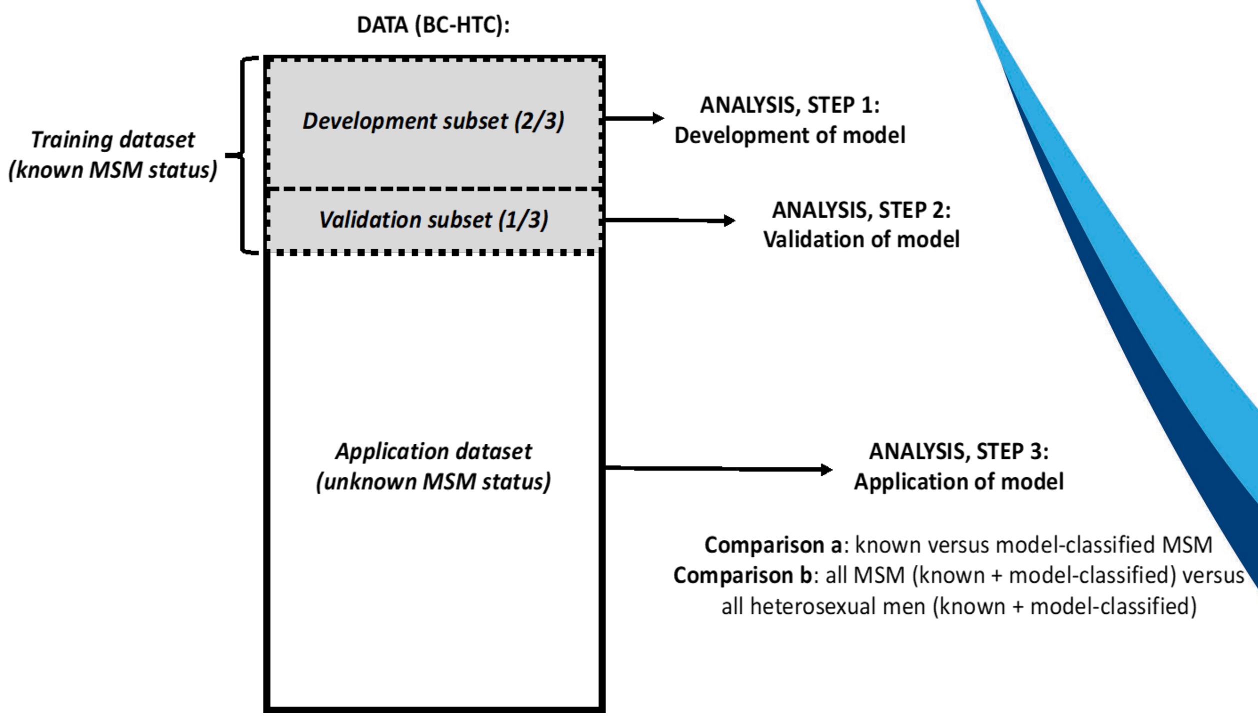
- To identify population groups and behaviors hidden in data
- MSM are disproportionately represented in multiple epidemics (HIV, Hep B, etc.)
- Linked public health databases could be useful in for monitoring trends and evaluating prevention and treatment programs within MSM;
    - however, MSM status is only routinely measured at some encounters (e.g., HIV case reports).
- Computational phenotyping, or predictive modeling, offers a powerful remedy to this limitation

Model development and application

Training (gold standard/self-report) dataset

- STI and HIV surveillance data

with follow-up for risk factors assessed gender identity



Selected potential variables based on expert opinion

- HIV and STI testing frequency,
- HIV and STI diagnoses,
- substance use,
- visits to clinics providing services to MSM,
- prescription for HIV PrEP,
- Residence in an area with high MSM population density.

Use of various ML techniques

- Application to the males in the BC-HTC.
  - Random Forest
  - Support Vector Machines (SVM)
  - Naïve Bayes
  - ***Elastic Net (Ridge/Lasso) regression***
    - Sensitivity: 72.2%
    - Specificity: 93.6%
    - Accuracy: 88%
    - AUC: 92.4%

Injection drug use, gender/sexual orientation, and infection risk

	HBV (OR (95%CI))	HBV HCV (OR (95%CI))	HBV HIV (OR (95%CI))	HBV HIV HCV (OR (95%CI))	HCV (OR (95%CI))	HIV (OR (95%CI))	HIV HCV (OR (95%CI))
Male/No IDU	1.0	1.0	1.0	1.0	1.0	1.0	1.0
MSM/No IDU	1.07 (1.01, 1.14)	0.36 (0.26, 0.5)	**5.24** (4.4, 6.24)	0.44 (0.34, 0.58)	0.28 (0.26, 0.31)	**4.67** (4.37, 4.98)	**2.45** (2.15, 2.79)
Female/No IDU	0.71 (0.7, 0.73)	0.34 (0.32, 0.37)	0.1 (0.08, 0.13)	0.02 (0.02, 0.03)	0.41 (0.4, 0.41)	0.2 (0.19, 0.22)	0.22 (0.2, 0.24)
Male/IDU	0.63 (0.55, 0.73)	**9.17** (8.34, 10.09)	2.25 (1.71, 2.97)	**6.92** (6.29, 7.63)	4.03 (3.88, 4.18)	0.98 (0.83, 1.16)	**11.84** (10.82, 12.95)
MSM/IDU*	NA	NA	NA	NA	NA	NA	NA
Female /IDU	0.52 (0.44, 0.61)	8.12 (7.3, 9.03)	0.56 (0.33, 0.94)	3.91 (3.47, 4.41)	3.41 (3.27, 3.55)	0.47 (0.36, 0.59)	**8.62** (7.77, 9.56)

McKee G 2018 Eclinical Medicine
DOI:https://doi.org/10.1016/j.eclinm.2018.10.006

ML applications for public health

- Linked administrative datasets could be used for:
  - Characterizing syndemics of social conditions, infections and outcomes
  - Public health surveillance and Program planning and monitoring
- Machine learning techniques could enable characterizing population groups and characteristics in linked datasets for informing optimal suite of prevention and care services

Implementation in the era of cloud based surveillance platforms: **Long Covid**

BC Centre for Disease Control
Provincial Health Services Authority

British Columbia Ministry of Health

BC Centre for Disease Control
Provincial Health Services Authority

Provincial Health Services Authority
Province-wide solutions. Better health.

BC COVID-19 COHORT

COVID-19 Related Data

D	COVID-19 Case Surveillance	2020 onward
D	COVID-19 Hospital Census (PCMS)	2020 onward
D	COVID-19 Lab Tests & Sequencing	2020 onward
D	COVID-19 Vaccinations	2020 onward

Population/Demographic*

Y	Chronic Disease Registry	2008 – 2019/20
M	Client Roster	2008 onward
Y	Health System Matrix	2018/19
Y	Population Grouper Data	2008 onward
C	Socioeconomic Census Data (DA-level)	2016
D	Vital Statistics	2008 onward

Administrative/Laboratory*

D W	Emergency Department Visits (Health Authority, NACRS)	2020 onward 2011 onward
W	Hospitalizations (DAD)	2008 onward
M	Medical Visits (MSP)	2008 onward
W	Medications (PharmaNet)	2008 onward
D	Influenza Lab Tests	2008 onward
D	Private and Public Lab Tests (PLIS)	2020 onward
W	811 Call Data (respiratory related)	2014 onward

Refresh Cycles Daily Weekly Monthly Yearly Census Year

Data Stewards Regional Health Authority; Provincial Health Services Authority; Ministry of Health; Statistics Canada

DA = Dissemination Area; **DAD** = Discharge Abstracts Database; **MSP** = Medical Services Plan;
NACRS = National Ambulatory Care Reporting System; **PCMS** = Provincial COVID-19 Monitoring Solution;
PLIS = Provincial Laboratory Information Solution

*contain data for entire BC population

Implementation in the era of cloud based surveillance platforms: **Long Covid**

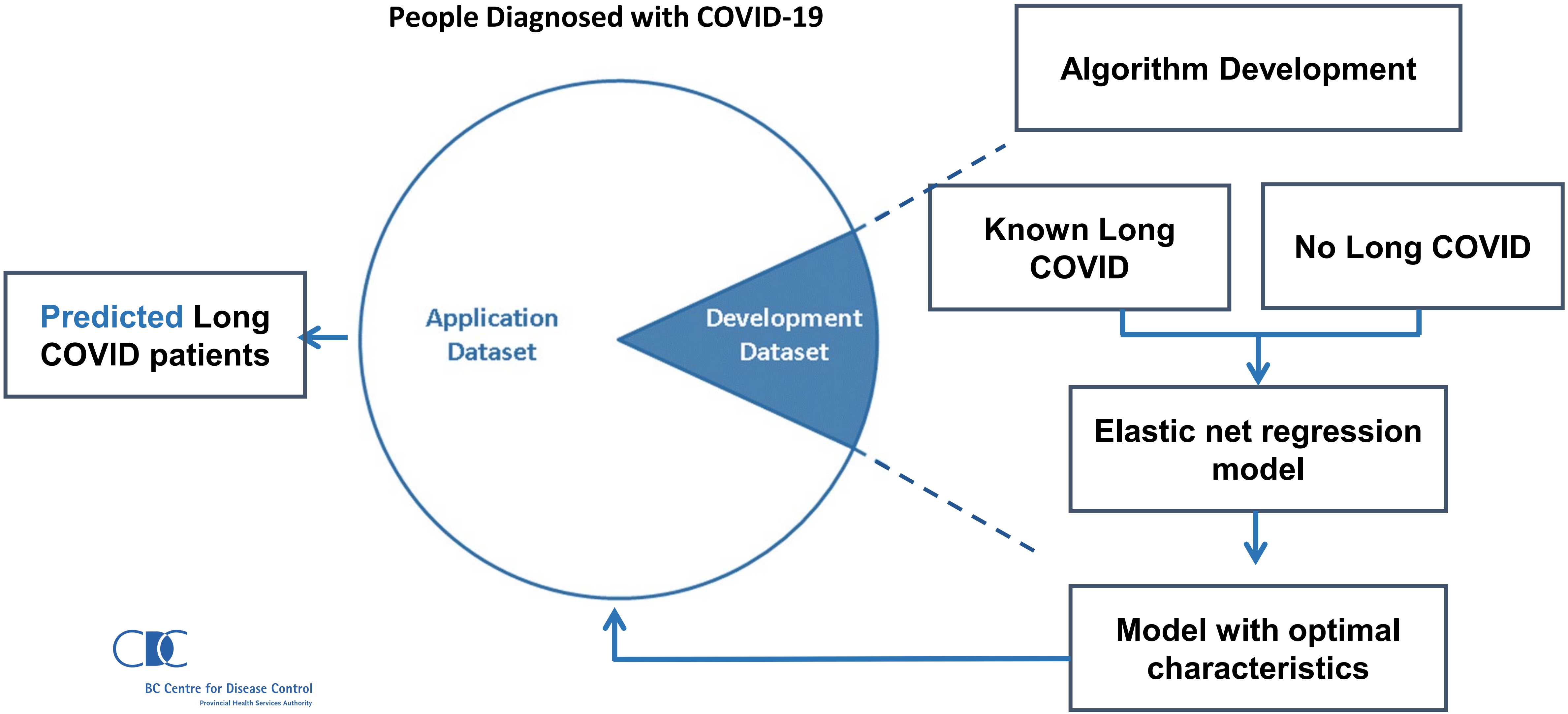
- Healthcare utilization data could provide critical information on the long COVID burden, which could inform care planning

- However, not all patients are diagnosed or assigned long COVID diagnostic codes

- We developed an algorithm to identify individuals with long COVID using population-level health administrative data from British Columbia (BC), Canada

- The algorithm will be refined for use across Canada

ML Approach



BC Centre for Disease Control
Provincial Health Services Authority

Identifying long COVID patients in the BCC19C using the characteristics of **known long** COVID patients within the cohort

1) **DAD, NACRS** (ICD-10-CA diagnostic code for "post-COVID condition")

Post COVID-19 condition

A post COVID-19 condition is classified to emergency use code U07.4 *Post COVID-19 condition* when a physician/primary care provider has documented a relationship between or association with a specific condition or symptom and past COVID-19. For the purpose of ICD-10-CA code assignment, a post COVID-19 condition is classified to a set of codes.

Examples of diagnoses where a physician/primary care provider has documented a relationship between or association with a specific condition or symptom and past COVID-19 include

- Post COVID-19 viral cough;
- Post COVID-19 deconditioning;
- Post viral intermittent shortness of breath post COVID-19 infection;
- Deep vein thrombosis secondary to past COVID-19; and
- Early pneumonia/post COVID-19.

2) **Post-COVID-19 Recovery Clinic** [PCRC] (Providence Care, VCH, Fraser Health, PHSA)

Section of PCRC referral form:

REASON FOR REFERRAL ** Required field. - will be used for triage purposes*

Category A	Category B	Category C
☐ Hospitalization for COVID-19 ☐ 2 or more ER presentations following diagnosis of COVID-19 ☐ New evidence of end organ impairment without identifiable cause: (check all that apply) ☐ cardio ☐ neuro ☐ resp ☐ renal ☐ musculoskeletal	☐ NYHA dyspnea scale 3 or higher (new finding) ☐ Inability to return to work or school post diagnosis of COVID-19 for 12 or more weeks ☐ Functional deterioration post diagnosis of COVID-19 (dependence on ADLs or iADLs) for 12 or more weeks	☐ Unexplained, persistent symptoms for more than 12 weeks post symptom-onset, thought to be related to COVID-19

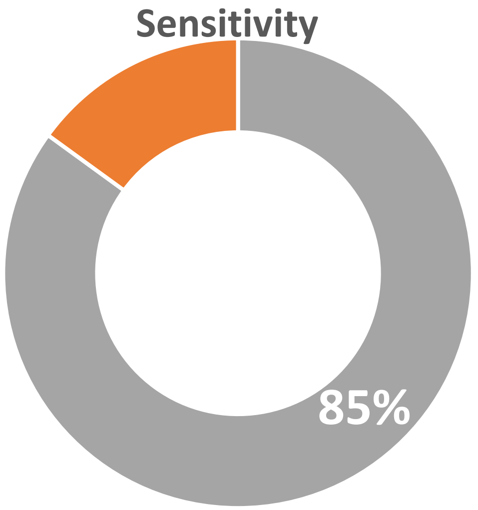
Referral Criteria, Referring Clinician Checklist on reverse.

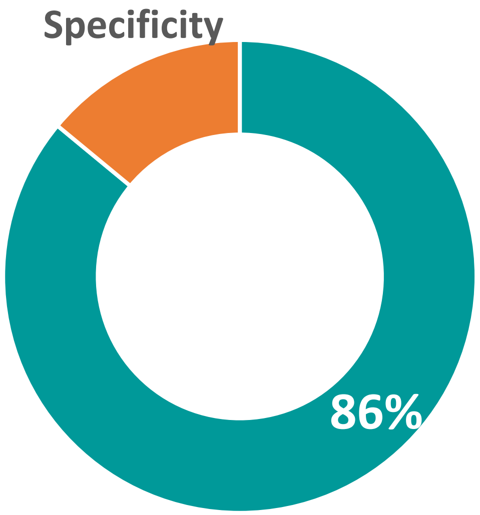
BC Centre for Disease Control
Provincial Health Services Authority

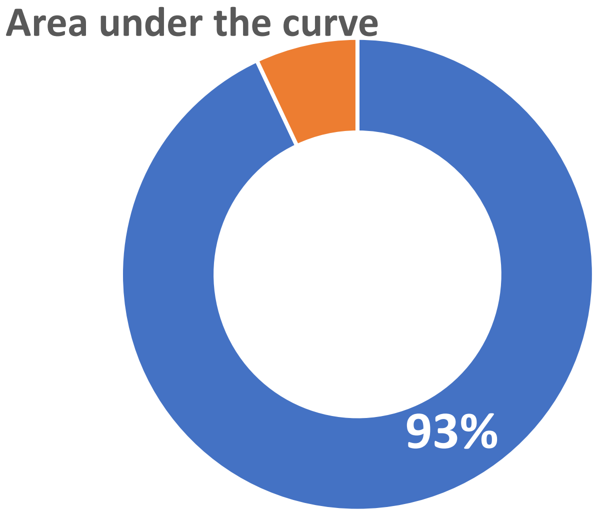
Binka et al, OFID, 2022

Refinement to Enhance Application

- Reduction in number of variables in the model
- Application of data reduction technique-
  - Principal component analysis of mixed data
  - Correlation of variables through independence test
  - Clinical assessment by two physicians
- Refined model based on ML, statistical techniques and clinical input





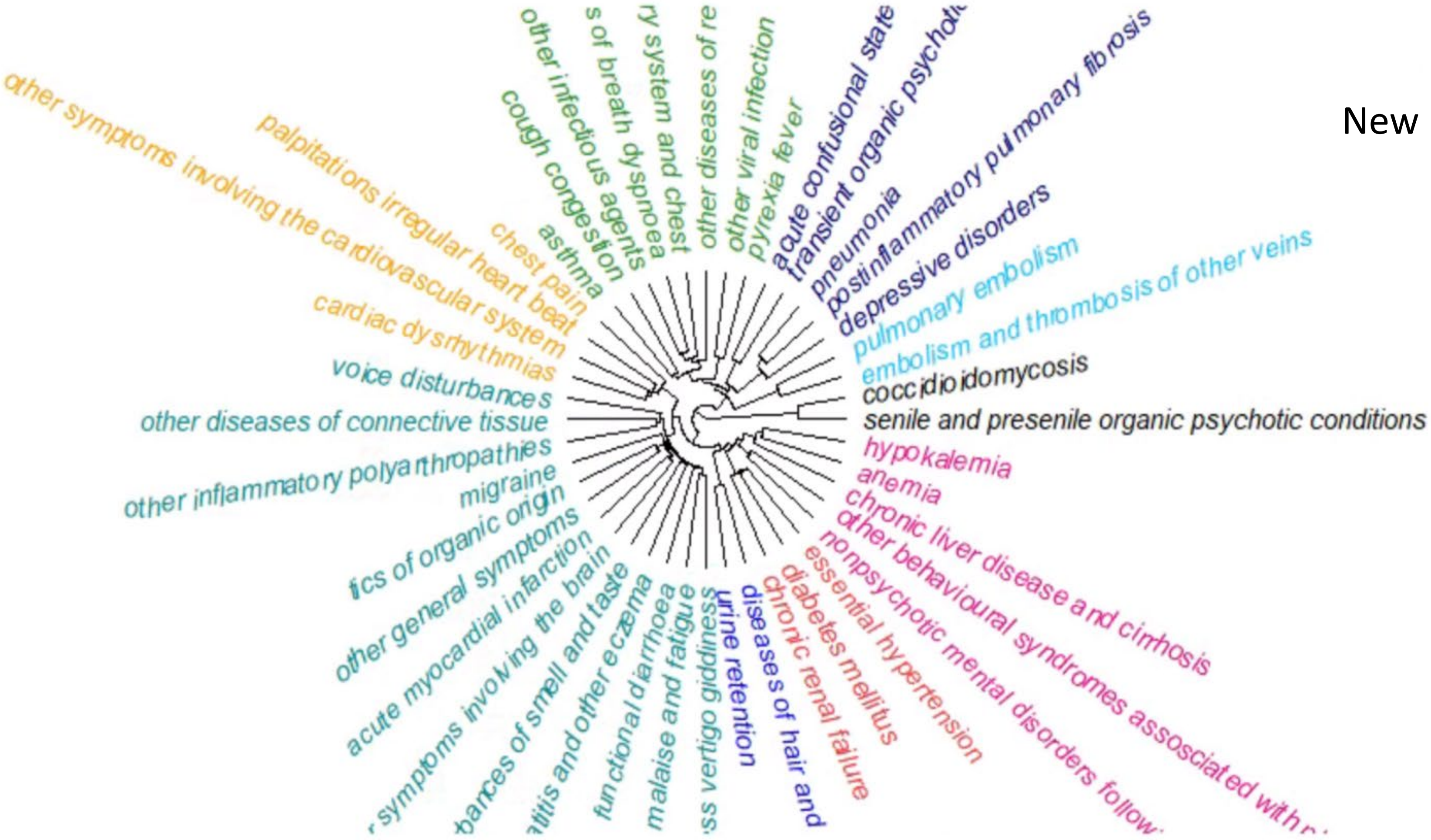


Identified long COVID cases:

52, 512 long COVID patients among 141,381 adult COVID-19 cases

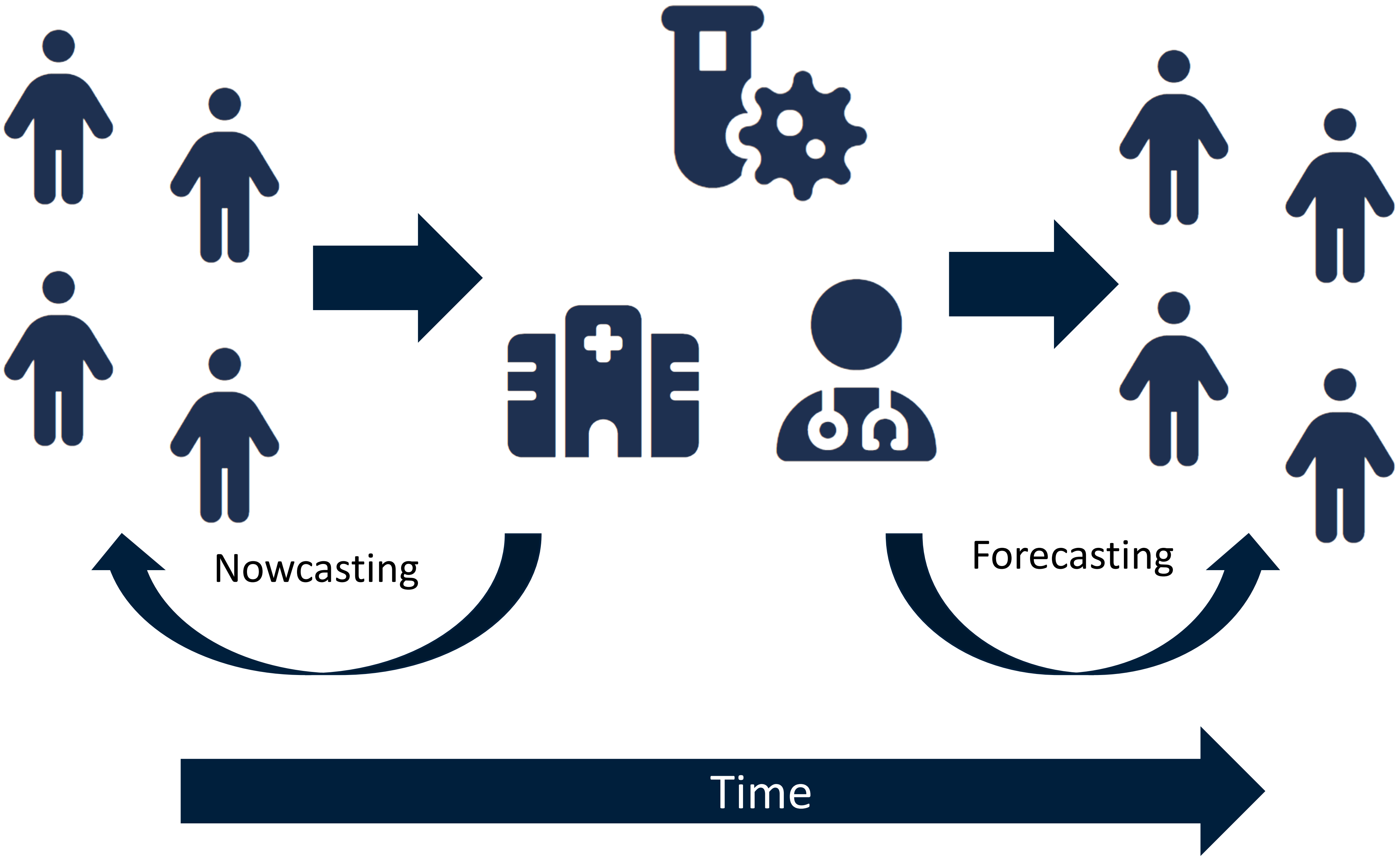
BC Centre for Disease Control
Provincial Health Services Authority

Phenotypes clustering within long-covid predictions

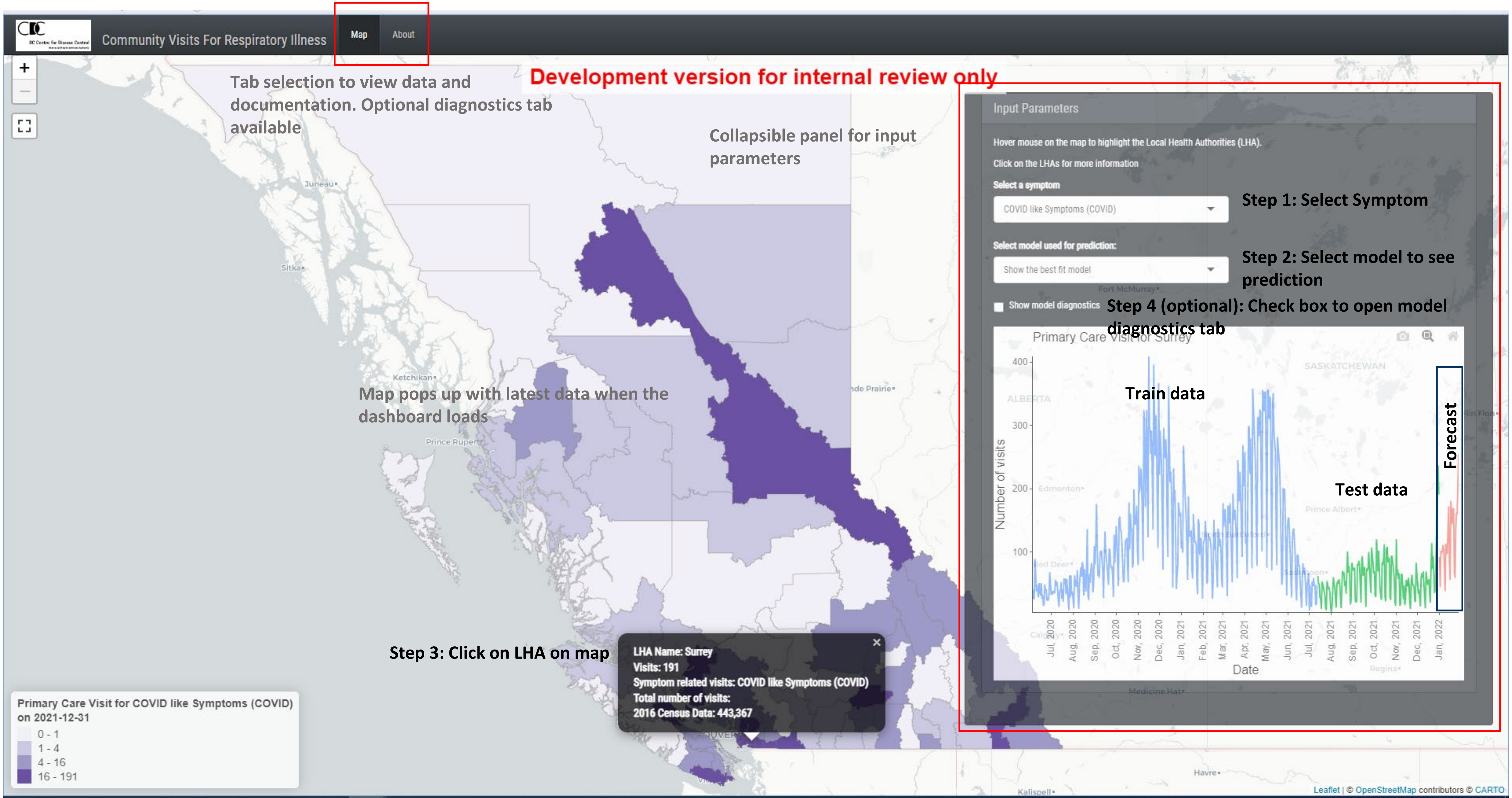


New Model- 57 variables

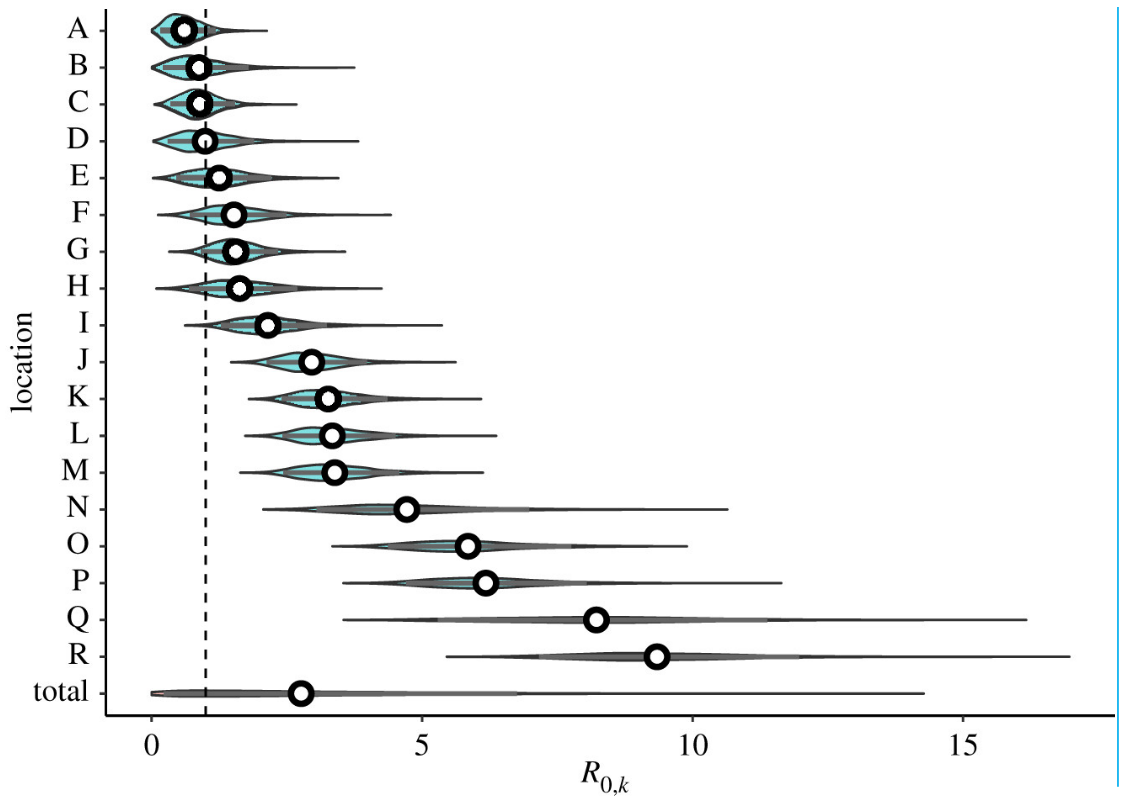
Applications for nowcasting and forecasting

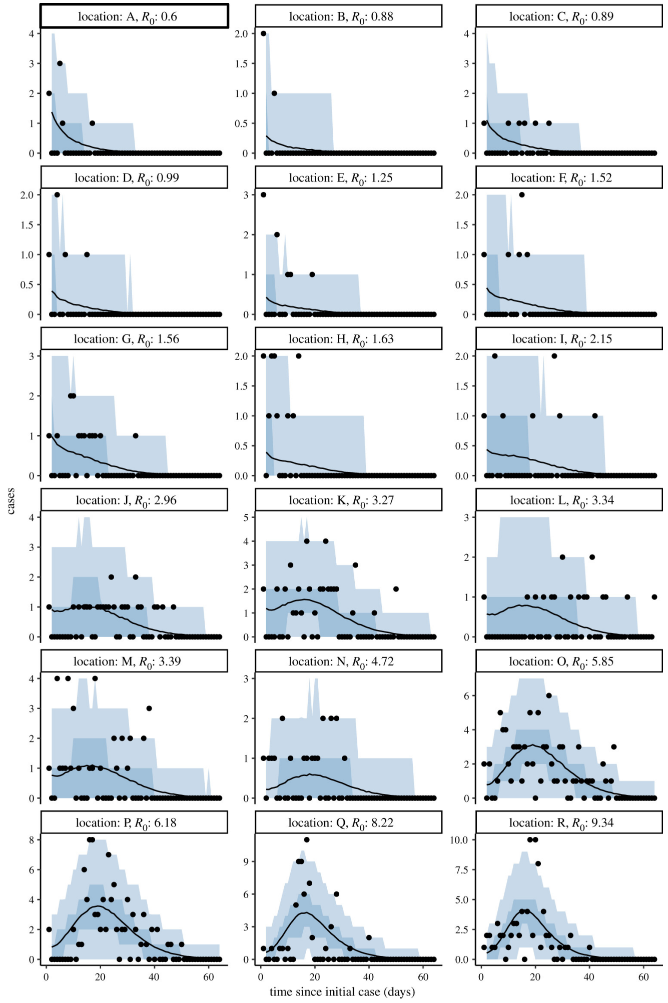

Nowcasting
Forecasting
Time

Forecasting for MSP viral respiratory infection visits



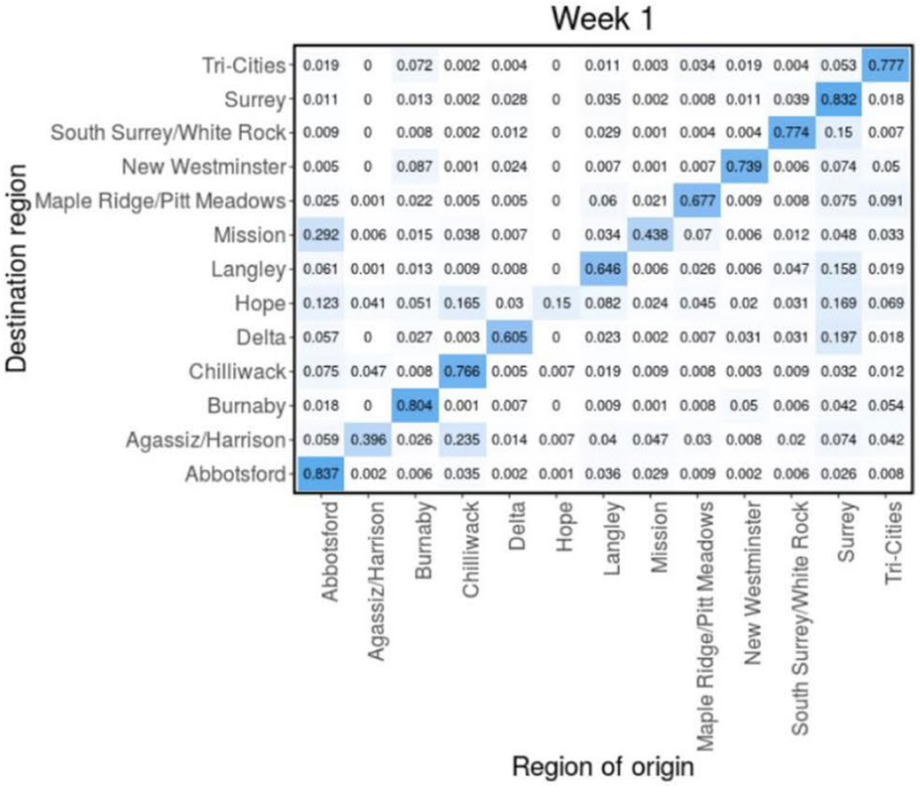
Quantifying the transmissibility of COVID-19 in long-term care

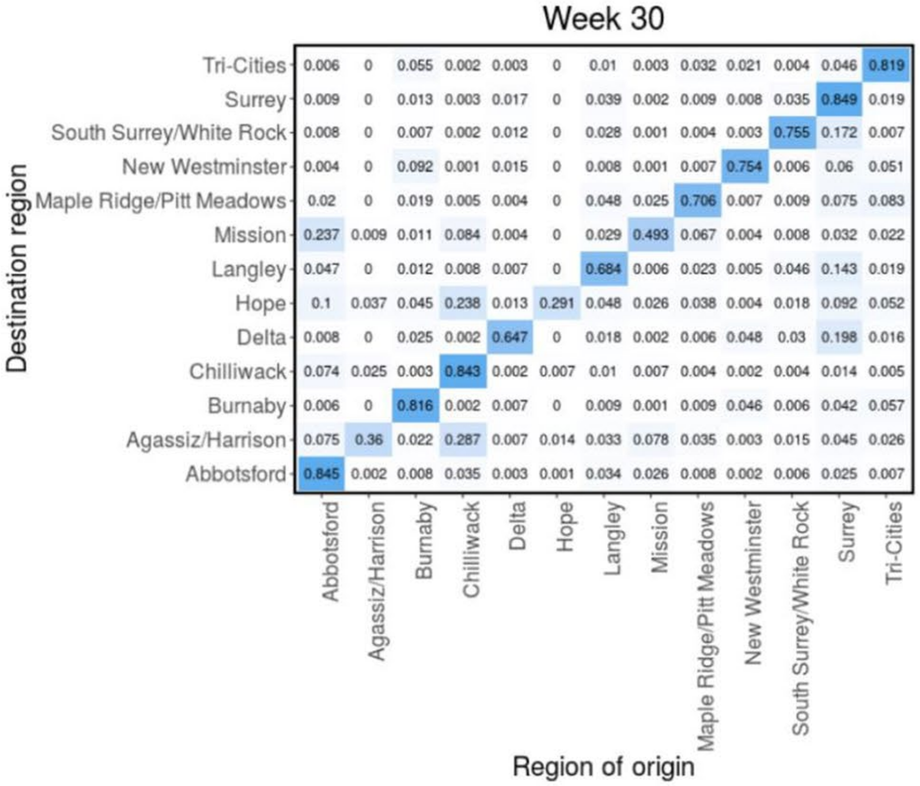




Stockdale, Jessica E., et al. "Quantifying transmissibility of SARS-CoV-2 and impact of intervention within long-term healthcare facilities." *Royal Society Open Science* 9.1 (2022): 211710.

Utilizing Telus mobile data to understand population mobility and the spread of COVID-19 during the pandemic

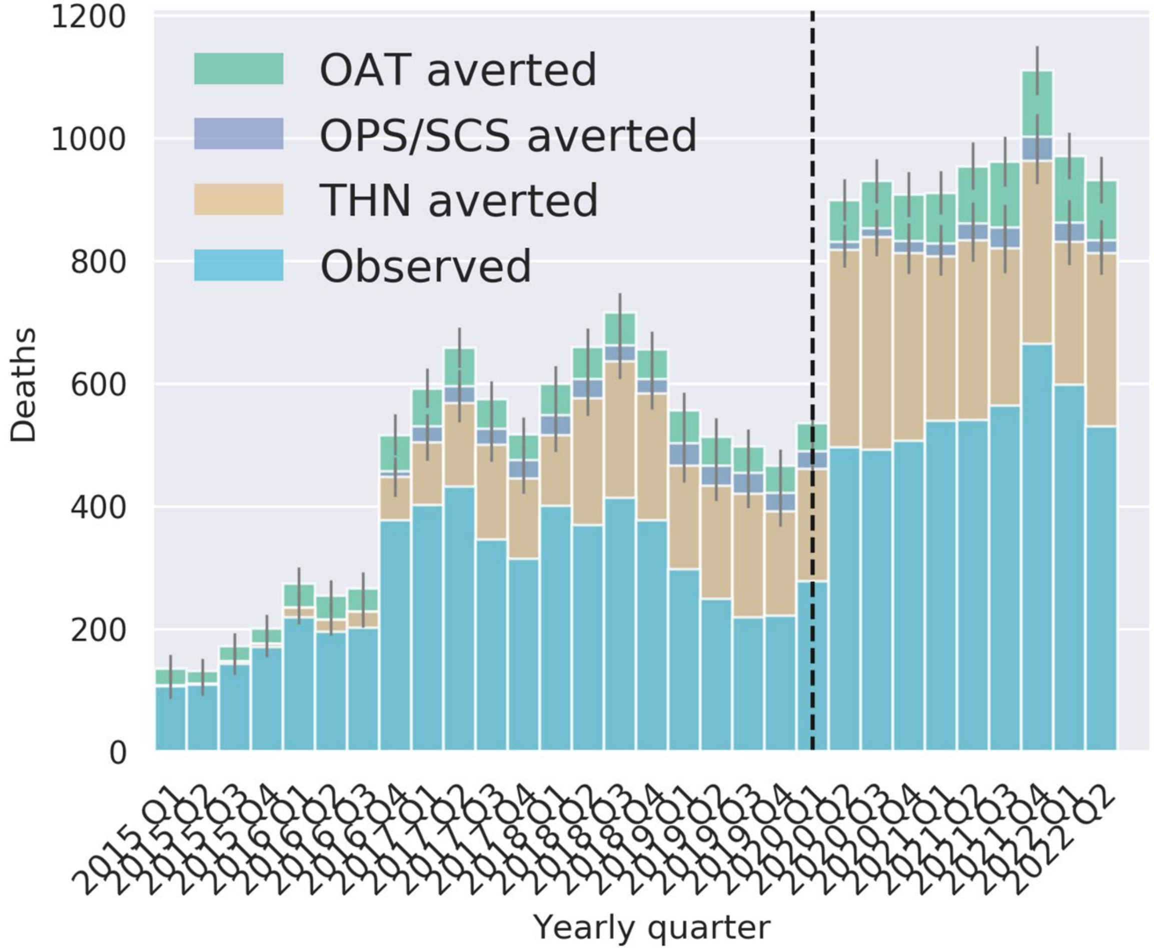




Iyaniwura SA, Ringa N, Adu PA, Mak S, Janjua NZ, Irvine MA, et al. (2023) Understanding the impact of mobility on COVID-19 spread: A hybrid gravity-metapopulation model of COVID-19. PLoS Comput Biol 19(5): e1011123. https://doi.org/10.1371/journal.pcbi.1011123

Generating insights into interventions in response to the Unregulated Drug Poisoning Emergency

- Approach incorporates data from treatment, overdose prevention services, urinalysis, and other sources
- Estimated impact of THN, OPS/SCS, and OAT on the total number of deaths averted for British Columbia from January 2015 to June 2022. Dashed line indicates start of COVID-19 public health emergency in BC.
- Note: THN deaths averted based on reported kits used and is an underestimate of the total impact of the program



THN: take home naloxone
OPS/SCS: Overdose prevention services/ supervised consumption sites
OAT: Opioid-agonist therapy

The future of AI in public health

- Advancements in machine learning allow for the combination of disparate forms of data including mobility, text, geographic and images
- COVID-19 pandemic accelerated applications of modeling combining expert understanding with data models
- Future includes:
  - Wastewater integration
  - NLP & epidemiology
  - Public health genomics

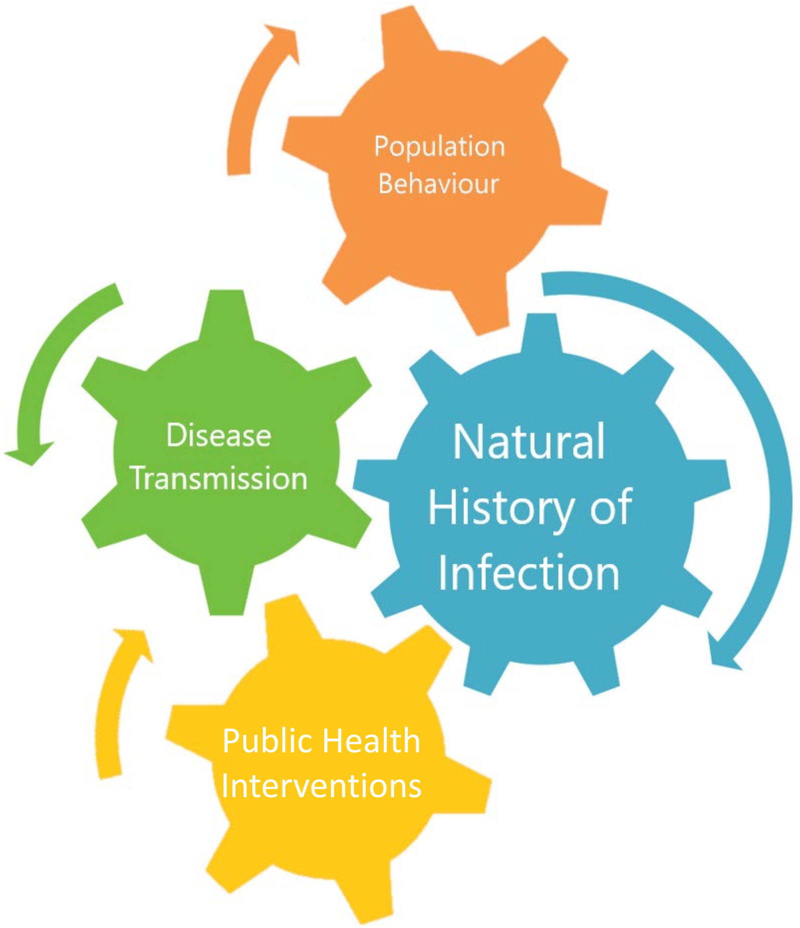


Figure adapted from: Brooks-Pollock, Ellen, et al. "Modelling that shaped the early COVID-19 pandemic response in the UK." *Philosophical Transactions of the Royal Society B* 376.1829 (2021): 20210001.

Thank you

BCCDC Data Science & Innovation | Data and Analytics Services

- Hind Sbihi - hind.sbihi@bccdc.ca
- Mike Irvine – mike.irvine@bccdc.ca

SUPPLEMENTAL