CODA-19: a Collaborative Data Analysis Platform to Improve Clinical Care in Patients with COVID-19

Principal Investigator

Michaël Chassé, MD, PhD, FRCPC, CHUM, University of Montreal

Co-principal Investigators

David Buckeridge, MD, PhD, McGill University Health Centre Research Institute, McGill
Jonathan Afilalo, MD, MSc, Jewish General Hospital, McGill University
Han Ting Wang MD, MSc, FRCPC, Maisonneuve-Rosemont Hospital, University of Montreal
Yiorgos A. Cavayas, MD, MSc, FRCPC, Hôpital Sacré-Coeur de Montréal, University of Montreal
Alexis Turgeon MD, MSc, FRCPC, CHU de Québec-Université Laval Research Center, Université Laval
Patrick Archambault, MD, MSc, FRCPC, CISSS Chaudière-Appalaches, Université Laval
Joelle Pineau, PhD, Mila, McGill University

Co-Investigators

Marc Afilalo, MD, Jewish General Hospital, McGill University
François Martin Carrier, MD, MSc, FRCPC, CHUM, University of Montreal
Emmanuel Charbonney, MD, PhD, University of Montreal
Carl Chartrand-Lefebvre, MD, MSc, FRCPC, CHUM, University of Montreal
Joseph Paul Cohen, PhD, Mila, University of Montreal
Audrey Durand, PhD, Mila, Université Laval
Madeleine Durand, MD, MSc, FRCPC, University of Montreal
Shane W. English, MD, MSc, The Ottawa Hospital, University of Ottawa
Philippe Jouvet, MD, PhD, MBA, Hôpital Sainte-Justine, University of Montreal
Louis-Antoine Mullie, MD, FRCPC, CHUM, University of Montreal
Esli Osmanlliu MD, MSc (cand), FRCPC, McGill University Health Centre Research Institute, McGill
Guillaume Plourde, MD, PhD, CHUM, University of Montreal
Brent Richards, MD, MSc, Jewish General Hospital, McGill University
Antony Robert, MD, MASc, FRCPC, McGill University Health Center, McGill University
Michaël Sauthier, MD, MBI, Hôpital Sainte-Justine, University of Montreal
Nicolas Sauthier, MD, MSc (cand), CHUM, University of Montreal
An Tang, MD, MSc, FRCPC, CHUM, University of Montreal

Participating Sites

Centre Hospitalier de l'Université de Montréal (CHUM)
Hôpital Maisonneuve-Rosemont (CIUSSS de l'Est-de-l'Île-de-Montréal)
Hôpital Général Juif (CIUSSS du Centre-Ouest-de-l'Île-de-Montréal)
Centre Universitaire Santé McGill (CUSM/MUHC/Montreal Children Hospital)
Hôpital Sacré-Cœur de Montréal (CIUSSS du Nord-de-l'Île-de-Montréal)
Centre Hospitalier Universitaire Sainte-Justine (CHU Sainte-Justine)
Centre Hospitalier Universitaire de Québec - Université Laval (CHU de Québec)
CISSS de Chaudière-Appalaches

Submitted to

Canadian Institutes of Health Research
May 12th, 2020

Lay Abstract

COVID-19 is a highly contagious acute respiratory illness that has undergone rapid global spread in the beginning of 2020. There is a pressing need to develop tools that can help physicians diagnose COVID-19 rapidly, determine if different disease presentations warrant different types of treatment, flag patients at high risk of deteriorating, and ensure healthcare resources are attributed efficiently and equitably. Through an established partnership with 9 hospitals, a collaborative analysis platform has been developed to pool data from multiple sites while minimizing the exchange of patient-level information. This collaboration is building on a large database of biological data from patients tested for COVID-19 that is being collected in these hospitals. Risk prediction models will be developed to identify patients at high risk of COVID prior to the availability of definitive testing, characterize distinct disease trajectories, intervene pre-emptively in patients at high risk of clinical deterioration, and make forecasts to plan hospital resources and staffing. The accuracy of predictions will be continuously verified using new cases, which will be identified from different hospital sites in real time. These predictive models will be used to build tools that can help physicians better treat patients with COVID-19, and provide actionable recommendations to support Canada's response to COVID-19.

Priority Announcement: List of Relevant Research Areas

Clinical management and health system interventions

The COVID-19 pandemic has led to **unprecedented demand for health system resources**, sometimes resulting in catastrophic situations in locations where demands exceeded available capacity.

In order to support **evidence-based public health policies**, there is a pressing need to develop **predictive models that forecast resource allocation**, including ward and critical care beds, ventilators, and personal protective equipment.

The development of these predictive models is needed urgently to **inform clinical and administrative decision-making** during the pandemic response, and make the most efficient and equitable use of hospital resources while **mitigating the impact on COVID-negative patients** that require care.

Through an established partnership with **9 hospital sites at the Canadian epicenter of the pandemic**, we have built a large repository of anonymized biological data from patients with confirmed or suspected COVID-19. Using a multicenter collaborative platform to support **real-time data development** and **prospective clinical validation** of the prediction models, we will develop point-of-care decision support tools. Online simulation tools that anticipate activity peaks and visualize bed occupancy projections will be created and disseminated to frontline healthcare workers with the aim of ensuring the maximal provision of scheduled clinical activities while planning for surges.

Diagnostics

Polymerase chain reaction (PCR) testing for COVID-19 has imperfect sensitivity, variable turnaround time, and is vulnerable to delays during peaks of disease activity due to reagent shortages. Moreover, it is not available in several geographically remote settings.

Predicting COVID-19 status from readily available clinical information - symptoms, comorbidities and prior expositions, demographics, laboratory and imaging studies - would **assist clinicians in early diagnosis** prior to the availability of a PCR test result and help with the appropriate management of these patients.

We will develop machine learning and epidemiological models to **predict COVID-19 status** among patients who seek medical attention with acute respiratory symptoms, whether in an academic urban institution or in a rural firstline facility. They will have the potential to **enable better patient triage, isolation and orientation through the healthcare system**. A user-friendly online interface will be developed in order to enable the use of our predictive models by a non-technical audience of practicing clinicians.

Therapeutics

Based mainly on bedside clinical observations, different phenotypes of COVID-19 disease have been proposed. Some authors proposed a classification based on respiratory mechanics and radiological findings, with possible implications in terms of supportive therapy selection. A hypercoagulable phenotype has also been observed in a subset of patients, prompting recommendations for intensified antithrombotic therapy. It is not yet known if the existence of significant phenotypes is borne out by clinical data; **the identification of such entities could help guide patient management**.

Using machine learning techniques, disease phenotypes will be derived among patients hospitalized for COVID-19. Once they will be identified, we will assess their clinical correlates and evaluate their impact on specific treatments. We will also generate **hypotheses that can inform future interventional trials** in patients with COVID-19.

Objective: To build a multi-centre data infrastructure enabling the rapid development and prospective validation of predictive models to aid the clinical management of Coronavirus Disease 2019 (COVID-19) and optimize healthcare resource utilization in response to the COVID-19 pandemic.

Rationale: An evidence-based Canadian response to the ongoing COVID-19 pandemic requires the collection and analysis of high-quality, structured data on patients in whom COVID-19 is suspected or confirmed. Data infrastructures are urgently needed in order to develop and prospectively validate predictive models aimed at facilitating early diagnosis of COVID-19, adapting management to distinct disease presentations, identifying patients at risk of adverse outcomes, and forecasting resource usage.

Preliminary Results: We have developed **CODA-19**, a data repository of all patients with suspected or confirmed COVID-19 at 8 hospital sites in Québec and 1 in Ontario. In addition to standard clinical characteristics and outcomes, this repository contains **multi-modality biological signals**, including **clinical parameter trends** (e.g. laboratory tests, vital signs, ventilator settings over time), **2D and 3D chest imaging data**, and other sensor recordings (e.g. electrocardiograms). We have obtained multi-centric IRB approval for this data repository, built a prototype infrastructure, and imported a first batch of data **(*n* = 5,637 positive / 37,315 tested).** Data extraction is funded and ongoing.

Aims: We will develop a collaborative analysis platform to conduct **real-time, prospective clinical validation** of epidemiological and machine learning models developed using **CODA-19** under 4 domains: **1) Diagnostic risk stratification**: To rapidly estimate the probability of COVID-19 in patients presenting with compatible symptomatology; **2) Clinical phenotypes**: To identify distinct clinical phenotypes in patients with confirmed COVID-19, and assess whether phenotypes influence the response to supportive treatments; **3) Early warning system**: To identify early warning signs that predict the time to an adverse outcome among inpatients with confirmed COVID-19; and, **4) Health system resource use**: To forecast the need for beds, materials and staff, identify at-risk thresholds for equipment shortages, and optimize the delivery of care for patients with and without COVID-19.

Team: We have established partnerships with **9 hospital sites** – including 6 sites in Montréal, the COVID-19 epicenter in Canada – to recruit patients into **CODA-19** and conduct prospective clinical validation of the models developed. Our multi-disciplinary team combines **leading expertise** in the fields of machine learning, data science, epidemiology, and biostatistics, with **strong clinical health data research experience** in radiology, internal medicine, emergency medicine, and critical care.

Knowledge Translation: We will develop and distribute a web-based, interactive **diagnostic risk stratification tool** as well as a **resource-planning forecasting and simulation tool**. We will implement and validate an **early warning system** to identify patients at risk of deterioration in real-time, as well as a **COVID risk dashboard** to simplify bed management at each site. We will distribute an **open-source library** to facilitate access to CODA-19 data by authorized researchers.

Relevance: Results will help mount an evidence-based response to the COVID-19 pandemic by providing tools to inform triage at the frontlines, to select the optimal care setting for each patient, and to ensure equitable delivery of health care services for patients with and without COVID-19. Our data analysis platform will enhance the readiness of Canadian hospitals to future pandemics, and help catalyze further multi-centric research efforts between participating sites.

Background and Rationale: COVID-19 is a highly contagious acute respiratory illness that has undergone rapid global spread in the beginning of 2020. It is a major national public health threat that has disproportionately affected vulnerable Canadians, such as immunocompromised patients, older patients, and those living in long-term care facilities.[1,2] There is a pressing need to develop predictive models in large patient cohorts to enable early diagnosis of COVID-19, adapt management strategies to individual risk profiles, identify early warning signs for clinical deterioration, forecast resource usage, and optimize health system organization. Scalable data sharing and distributed analysis infrastructures are urgently needed to enable the development and clinical validation of robust predictive models across multiple sites, and translate models into impactful decision support tools.

In collaboration with **9 hospital sites** at the Canadian epicenter of the pandemic, we have built a **large repository** of anonymized, **multi-modality data** from patients with suspected or confirmed COVID-19. Our team of **leading experts** in **clinical research** and **health data science** is uniquely positioned to tackle the challenge of putting big data at the service of clinicians and administrators managing COVID-19. Using a **scalable collaborative analysis platform** to conduct multicenter **prospective model validation**, we will develop **point-of-care decision support tools** and **provide actionable insights** to support Canada's response to COVID-19.

Preliminary Results: We have developed CODA-19, which likely represents the largest acute care COVID-19 data repository in Canada. CODA-19 will contain data for a cohort of all patients with suspected or confirmed COVID-19 tested at **8 hospital sites in Québec and 1 in Ontario** (as of writing, ready for analysis, n = 5,637 positive / 37,315 tested). In addition to standard clinical characteristics and outcomes, this repository includes **multi-modality biological data**, including clinical parameter trends (e.g. laboratory tests, vital signs, ventilator settings over time), 2D and 3D imaging data (e.g. chest X-rays, chest CT scans, lung ultrasound), and other sensor recordings (e.g. electrocardiograms). To facilitate data collection and analysis, we have established a **common data model and variable dictionary** (available at www.coda19.com), which aligns with the WHO/ISARIC reporting guidelines for severe acute respiratory infection.[3] We have also developed and implemented a **scalable secure network infrastructure** (**Figure 1**) for multi-site data analysis.

Clinical Domains of Research: CODA-19 will be used to develop and **prospectively validate** epidemiological and machine learning prediction models, both supervised and unsupervised, answering **4 key clinical needs**: **1)** achieving early diagnostic risk stratification, in order to improve early patient triage, and reduce nosocomial transmission; **2)** understanding how distinct **clinical**

Figure 1. Model development via secure data sharing infrastructure

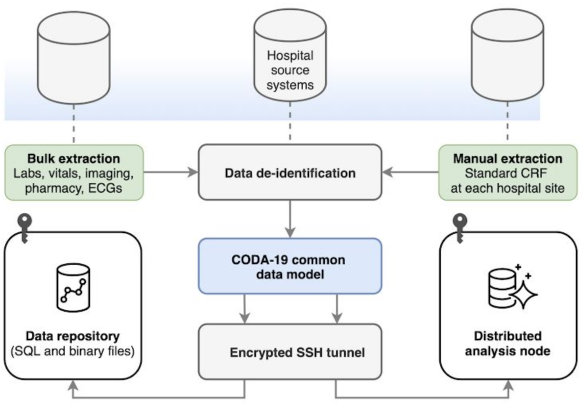
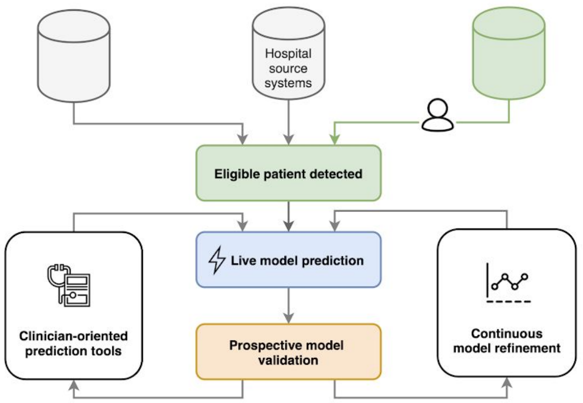


Figure 2. Continuous prospective clinical validation and model refinement


Hospital source systems
Eligible patient detected
Live model prediction
Clinician-oriented prediction tools
Prospective model validation
Continuous model refinement

phenotypes respond to alternative treatments, to provide individualized risk assessment and management; **3)** identifying **early warning signs** that herald clinical deterioration, to enable proactive monitoring and treatment; and **4)** forecasting **resource usage**, to optimize health system resource use.

Prospective Clinical Validation: We will conduct **ongoing, prospective clinical validation** of the epidemiological and machine learning models developed in the 4 clinical domains. A **scalable data analysis platform** has been developed and will be deployed across sites to continuously **aggregate, monitor and analyze** the results of model validation (**Figure 2**). Patients meeting the inclusion criteria for each model will be identified on an **automated, real-time basis** and will be used to prospectively validate model performance. New patient data will be incorporated on an ongoing basis to refine existing models. State-of-the-art **distributed and federated learning** strategies will be used to train statistical models across sites, while minimizing the amount of patient-level data sharing.[3–5]

Team Expertise and Roles: **Dr. Michaël Chassé** will lead the team. He is an intensivist, health data scientist at the Centre Hospitalier de l'Université de Montréal (CHUM), and Associate Professor at University of Montreal. He is also an IVADO.ca professor and the scientific director of the CHUM Center for Integration and Analysis of Medical Data (CITADEL). He has led several national clinical trials, including multi-centre data science projects. He brings together a group of scientists and professionals specialized in health data science (**Pineau, Buckeridge, Cohen, Osmanlliu, Mullie, Sauthier**), biostatistics (**Buckeridge, Carrier, Turgeon**), bioinformatics and machine learning (**JAfilalo, ADurand, Pineau, Mullie, Sauthier**), epidemiology (**JAfilalo, Carrier, English, MDurand, Plourde, Turgeon**), knowledge transfer (**Archambault, MAfilalo**), clinical informatics (**Robert, Sauthier)**, as well as physicians with expertise in radiology (**Chartrand-Lefebvre, Tang**), adult/pediatric critical care (**Wang, Cavayas, English, Jouvet, Turgeon, Archambault, Charbonney, Carrier, Plourde, Mullie, Jouvet, Sauthier**) and emergency medicine (**Archambault, Esli, Robert, MAfilalo**). Our collaborative effort brings together the networks of University of Montreal, Université Laval, University of Ottawa and McGill University, as well as the Quebec Institute of Artificial Intelligence (Mila) (**Cohen, Pineau, Buckeridge, ADurand**) Reasoning and Learning Lab co-lead (**Pineau**). The team will be managed by an executive committee (PIs, Mullie, Plourde), and for daily management, by a steering committee (PIs and co-Is, domain-specific coordination, clinical validation).

Domain 1: Diagnostic Risk Stratification

Aim: To develop and validate methods to estimate the probability of COVID-19 in patients presenting to an acute care setting for symptoms or complications of COVID-19.

Overview: Polymerase chain reaction (PCR) testing for COVID-19 has a 6-19% false negative rate, a variable turnaround time, and is vulnerable to delays when the number of tests performed increases.[6,7] Predicting COVID-19 status from readily available clinical information would improve early patient orientation, prior to the availability of a PCR test result. Automated prediction systems can also assist in identifying missed cases and nosocomial transmission on an ongoing basis.

Methods: Epidemiological and machine learning models will be developed to identify determinants and predictors of COVID-19 status among patients presenting to the ER or transferred from another hospital for symptomatology compatible with COVID-19, in whom a COVID-19 PCR test was sampled within ± 5 days of presenting to care, and in whom laboratory tests and chest X-ray were performed within 24h of arrival. Patients with any positive test in the window period will be considered positive. **A. Epidemiological model.** In order to control for site-specific differences in patient populations, and variations in COVID-19 treatment designation status over time, COVID-positive patients will be matched in a 1:4 ratio to controls for age, sex, calendar week, hospital site.[8] Conditional logistic regression will be performed in the matched patient cohort, with predefined independent variables identified via a review of the literature.[9] The most abnormal result in the first 24 hours will be sampled for discrete-time data. Missing data will be imputed using multiple imputation by chained equations (MICE) with random forest (RF) regressors.[10,11] All analyses will be stratified for significant comorbidities that may affect COVID-19 outcomes, such as diabetes and immunocompromising conditions at baseline.[12] **B. Machine learning model.** Gradient boosting techniques will be used to identify features of interest among numerical and categorical input variables in the matched patient cohort.[13,14] Synthetic Minority Oversampling Technique (SMOTE) will be used to balance classes in the training data set.[15–17] Supervised classifiers will be trained to estimate the probability of a COVID-19 diagnosis, based on imaging features and other selected input variables. Convolutional autoencoders will be used in order to learn a feature map representation for chest X-rays, using our data and publicly available chest X-ray data sets.[18,19] Model performance will be assessed using k-fold validation, with careful separation of training, validation and test sets. Predicted probabilities will be calibrated using Platt's method.[20–22] Prediction explanation techniques will be applied to gain insights into model reasoning.[23]

Expected output: We will develop and validate clinically an **interactive diagnostic risk stratification tool** that clinicians can use to estimate the probability of a COVID-19 diagnosis at the point-of-care. Using our predictive model, we will deploy a **COVID-19 risk dashboard** providing a color-coded overview of risk categories for patients in the ER. Based on epidemiological modeling, we will develop and validate a **simple clinical decision rule** to facilitate patient triage by first responders.

Domain 2: Clinical Phenotypes

Aim: To identify clinical phenotypes in patients with COVID-19, assess their interaction with the response to specific supportive management strategies, and evaluate their association with outcomes.

Overview: At least two clinical phenotypes of COVID-19 pneumonia have been described, on the basis of distinctive lung mechanics and radiological findings, and implications for clinical management have been inferred.[24,25] Multiple disease phases have also been described, according to whether viral pathogenicity or host inflammatory response is predominant.[26] In a subset of patients with COVID-19, a hypercoagulable phenotype has been observed, prompting recommendations for intensified antithrombotic therapy.[27–29] Although it is not yet known if the existence of distinct disease phenotypes is borne out by clinical data, the identification of such entities may have important implications for routine patient management and trial design.[30]

Methods: Clinical phenotypes will be derived in an unsupervised fashion among patients with an inpatient admission at one of the participating sites for ≥ 72 hours for an acute respiratory illness due to COVID-19, in whom a chest X-ray is available within ± 72 hours of admission. Patients meeting inclusion criteria will be split into a derivation and a validation cohort (75%/25%).[31] Inputs for phenotype derivation will include cross-sectional data (e.g. age, sex), time-varying data (e.g. laboratory tests, vital signs), and chest X-ray data. Missing data will be imputed using MICE-RF.[10] Clustering will be performed using deep embedded clustering.[32] Silhouette coefficients will be assessed, and the reproducibility of phenotypes will be evaluated.[33] Chord diagrams will be used to visualize the relation between phenotypes and organ dysfunction, as assessed by Sequential Organ Failure Assessment (SOFA) sub-scores (pulmonary, cardiovascular, renal, hepatic, coagulation).[34,35] The association between phenotypes and biomarkers of infection and inflammation will be assessed. The effects of phenotype assignment on patient-centered outcomes (e.g. all-cause mortality, ICU-free days), as well as organ failure (as assessed by SOFA score), will be assessed using linear mixed effects models, with participating centers as random effects and phenotypes as fixed effects. Phenotype × treatment interactions will be evaluated to determine if phenotypes are associated with differential responses to specific supportive therapies. The association between phenotype assignment and the response to candidate treatments (e.g. optimal ventilation strategies) will be evaluated using a generalized mixed effects model.[36]

Expected output: We will develop and validate clinically the first large-scale, multi-dimensional phenotypic analysis of patients with COVID-19. The identification and validation of candidate disease phenotypes will help our understanding of the pleomorphic clinical expression patterns of COVID-19. Deriving distinct associated organ dysfunction profiles will identify opportunities for intervention through adjustments in supportive care, and drive hypotheses for future interventional trials.

Domain 3: Early Warning System

Aim: To identify factors that predict and determine the time to an adverse clinical outcome in patients with COVID-19, and find early warning signs that predict deterioration in the acute care setting.

Overview: Epidemiological data suggests that older adults are disproportionately affected by complications of COVID-19.[37] In addition to age, other risk factors may influence patient evolution and predict clinical deterioration. Recent evidence suggests that patients who at first look stable may suddenly deteriorate, the so-called "*Happy hypoxemic patients*".[38] Accordingly, clinicians may preemptively admit patients to higher acuity units, leading to suboptimal use of scarce ICU resources. Risk stratification tools are urgently needed to determine which patients are likely to die or have an unfavourable outcome while hospitalized with COVID-19.

Methods: Predictors and determinants of adverse clinical outcome, defined as the need for invasive mechanical ventilation or death, will be analyzed among patients admitted for an acute respiratory illness due to COVID-19 to an inpatient unit at one of the participating sites. **A. Epidemiological model.** A marginal Cox regression model will be fitted to determine the association between predefined independent variables and time to occurrence of an adverse clinical outcome, using robust variances to take into account center effect and clustering.[39] **B. Machine learning model.** Machine learning models will be developed to assess the incremental value of using multiple sampling time points and incorporating imaging data into predictions. A semi-supervised anomaly detection model, using long-short term memory networks (LSTMs) and deep autoencoders, will be used to predict the occurrence of an adverse clinical outcome within the next 24 hours, based on data sampled in the preceding 72 hours.[40,41] Nonparametric regression models (recurrent neural networks, random forests) will be evaluated to numerically estimate the time to adverse clinical outcome. Mean absolute percentage errors will be determined using stratified k-fold validation, with stratification by site.[42]

Expected output: We will implement and validate clinically an early warning system for inpatients with COVID-19, using continuous predictions to identify patients at risk of clinical deterioration in real time. Epidemiological models will identify factors that determine an unfavourable clinical evolution early in the course of illness, and help clinicians adapt treatment to goals of care.

Domain 4: Health System Resource Use

Aim: To develop and validate models to predict resource demand by patients with suspected or confirmed COVID-19, and to optimize allocation of resources when the demand exceeds capacity.

Overview: The COVID-19 crisis has led to unprecedented demand for health system resources, resulting in at times catastrophic situations in locations where demands exceeded available capacity.

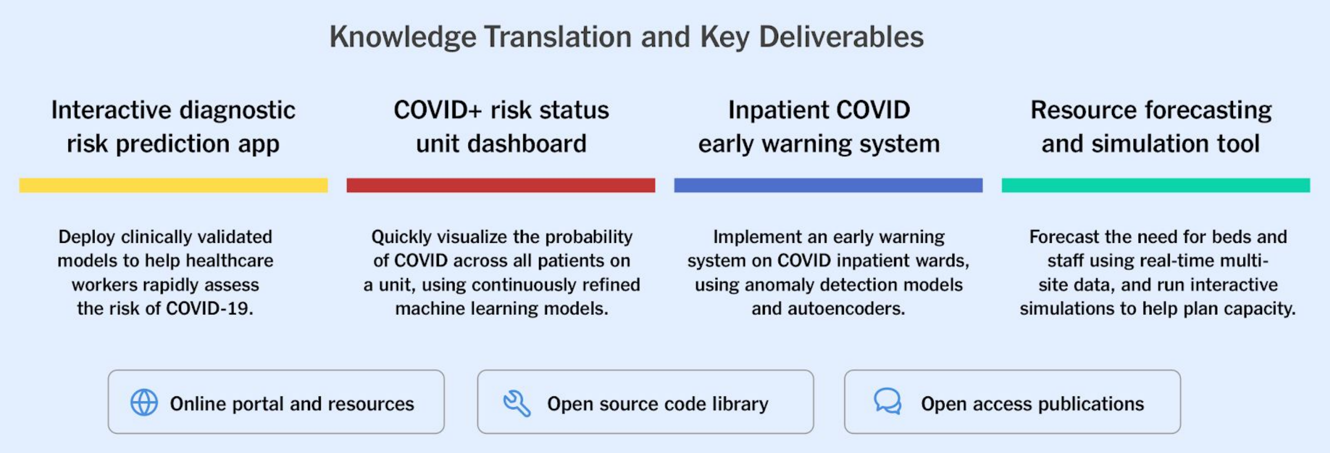
Models that forecast demand for different types of resources (e.g. ward beds, ICU beds, ventilators, protective equipment) are needed urgently to inform clinical and administrative decision-making during the pandemic response, and make the most efficient and equitable use of hospital resources.

Methods: Under a mandate from the Québec government, we have already developed a Markov state-transition model, trained using data on all hospital admissions for COVID-19 in Quebec, to project demand for health system resources (i.e., ward beds, ICU beds, ventilators) at the provincial and regional level (https://covid.mchi.mcgill.ca).[43] The model takes as input the projected rate of outside arrivals and estimates state-transition parameters from the data for movement through the ED, ward beds, ICU beds, ventilators, and discharge or death. The first extension to this model will be to stratify the patient states by features that influence patient trajectories (from Domain 3) and to expand the resources measured, including human and physical resources. The second extension will incorporate capacity constraints into the more richly-stratified Markov model and estimate parameters for decline in clinical status when patients cannot receive the required level of care due to capacity constraints. In a third step, we will explore machine learning methods such as LSTMs for predicting demand and optimizing resource allocation.[44] To simulate patient arrivals, we will sample from historical admissions, accounting for hospital closures and policies to allocate patients across hospitals.[45] Monte Carlo simulation models will be used to identify key thresholds that predict resource shortages.

Expected output: Models that anticipate disease activity peaks could be used to predict when demand is likely to exceed capacity, triggering actionable responses to balance load across sites. An online forecasting dashboard will be created and disseminated. A web-based resource planning simulation tool will be created, allowing clinicians and administrators to visualize bed occupancy projections under adjustable constraints, thus facilitating the optimization of bed capacity in anticipation of surges, informing public health policies, such as confinement, and optimizing scheduled clinical activities.

Sex and Gender Considerations: It is unclear how COVID-19 outcomes are affected by sex; some reports have suggested a higher mortality among men, while others have found no difference.[46,47] All the algorithms and statistics obtained in this project will consider biological sex as either a covariate or effect modifier, and as a key clustering variable for unsupervised machine learning models aiming to identify distinct clinical phenotypes.

Ethics and Privacy Considerations: Projects will be overseen by a Governance Committee, with one senior representative from each partner institution, and each subproject will obtain specific IRB authorization. Data sharing agreements have been established between participating sites. A data security framework has been adopted to stipulate best practices for protection of patient confidentiality. Given the observational nature of the study, individual patient consent will not be required.



Knowledge Translation and Key Deliverables

Interactive diagnostic risk prediction app

Deploy clinically validated models to help healthcare workers rapidly assess the risk of COVID-19.

COVID+ risk status unit dashboard

Quickly visualize the probability of COVID across all patients on a unit, using continuously refined machine learning models.

Inpatient COVID early warning system

Implement an early warning system on COVID inpatient wards, using anomaly detection models and autoencoders.

Resource forecasting and simulation tool

Forecast the need for beds and staff using real-time multi-site data, and run interactive simulations to help plan capacity.

Online portal and resources

Open source code library

Open access publications

References

1. Wang Y, Wang Y, Chen Y, Qin Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. J Med Virol. 2020 Mar 5;
2. COVID-19 Daily Epidemiology Update [Internet]. [cited 2020 May 11]. Available from: https://www.canada.ca/content/dam/phac-aspc/documents/services/diseases/2019-novel-coronavirus-infection/surv-covid19-epi-update-eng.pdf
3. Lu C-L, Wang S, Ji Z, Wu Y, Xiong L, Jiang X, et al. WebDISCO: a web service for distributed cox model learning without patient-level data sharing. J Am Med Inform Assoc JAMIA. 2015 Nov;22(6):1212–9.
4. Choudhury O, Park Y, Salonidis T, Gkoulalas-Divanis A, Sylla I, Das AK. Predicting Adverse Drug Reactions on Distributed Health Data using Federated Learning. AMIA Annu Symp Proc AMIA Symp. 2019;2019:313–22.
5. Ma J, Zhang Q, Lou J, Ho JC, Xiong L, Jiang X. Privacy-Preserving Tensor Factorization for Collaborative Health Data Analysis. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management [Internet]. Beijing China: ACM; 2019 [cited 2020 May 12]. p. 1291–300. Available from: https://dl.acm.org/doi/10.1145/3357384.3357878
6. He J-L, Luo L, Luo Z-D, Lyu J-X, Ng M-Y, Shen X-P, et al. Diagnostic performance between CT and initial real-time RT-PCR for clinically suspected 2019 coronavirus disease (COVID-19) patients outside Wuhan, China. Respir Med. 2020 Apr 21;168:105980.
7. Kim H, Hong H, Yoon SH. Diagnostic Performance of CT and Reverse Transcriptase-Polymerase Chain Reaction for Coronavirus Disease 2019: A Meta-Analysis. Radiology. 2020 Apr 17;201343.
8. Rasmy L, Wu Y, Wang N, Geng X, Zheng WJ, Wang F, et al. A study of generalizability of recurrent neural network-based predictive models for heart failure onset risk using a large and heterogeneous EHR data set. J Biomed Inform. 2018;84:11–6.
9. Kuo C-L, Duan Y, Grady J. Unconditional or Conditional Logistic Regression Model for Age-Matched Case-Control Data? Front Public Health. 2018;6:57.
10. Tang F, Ishwaran H. Random Forest Missing Data Algorithms. Stat Anal Data Min. 2017 Dec;10(6):363–77.
11. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. Stat Methods Med Res. 2007 Jun;16(3):219–42.
12. Guo W, Li M, Dong Y, Zhou H, Zhang Z, Tian C, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. Diabetes Metab Res Rev. 2020 Mar 31;e3319.
13. Kilic A, Goyal A, Miller JK, Gjekmarkaj E, Tam WL, Gleason TG, et al. Predictive Utility of a Machine Learning Algorithm in Estimating Mortality Risk in Cardiac Surgery. Ann Thorac Surg. 2019 Nov 7;
14. Xu Y, Yang X, Huang H, Peng C, Ge Y, Wu H, et al. Extreme Gradient Boosting Model Has a Better Performance in Predicting the Risk of 90-Day Readmissions in Patients with Ischaemic Stroke. J Stroke Cerebrovasc Dis Off J Natl Stroke Assoc. 2019 Dec;28(12):104441.
15. He H, Garcia EA. Learning from Imbalanced Data. IEEE Trans Knowl Data Eng. 2009 Sep;21(9):1263–84.
16. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res. 2002 Jun 1;16:321–57.
17. Blagus R, Lusa L. Class prediction for high-dimensional class-imbalanced data. BMC

Bioinformatics. 2010 Dec;11(1):523.

18. Abiyev RH, Ma'aitah MKS. Deep Convolutional Neural Networks for Chest Diseases Detection. J Healthc Eng. 2018 Aug 1;2018:1–11.
19. Irvin J, Rajpurkar P, Ko M, Yu Y, Ciurea-Ilcus S, Chute C, et al. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. ArXiv190107031 Cs Eess [Internet]. 2019 Jan 21 [cited 2020 May 12]; Available from: http://arxiv.org/abs/1901.07031
20. Chen W, Sahiner B, Samuelson F, Pezeshk A, Petrick N. Calibration of medical diagnostic classifier scores to the probability of disease. Stat Methods Med Res. 2018;27(5):1394–409.
21. Jiang X, Osl M, Kim J, Ohno-Machado L. Calibrating predictive model estimates to support personalized medicine. J Am Med Inform Assoc JAMIA. 2012 Apr;19(2):263–74.
22. Jiang X, Osl M, Kim J, Ohno-Machado L. Smooth Isotonic Regression: A New Method to Calibrate Predictive Models. AMIA Summits Transl Sci Proc. 2011 Mar 7;2011:16–20.
23. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. ArXiv160204938 Cs Stat [Internet]. 2016 Aug 9 [cited 2020 May 11]; Available from: http://arxiv.org/abs/1602.04938
24. Gattinoni L, Coppola S, Cressoni M, Busana M, Rossi S, Chiumello D. Covid-19 Does Not Lead to a "Typical" Acute Respiratory Distress Syndrome. Am J Respir Crit Care Med. 2020 Mar 30;
25. Gattinoni L, Chiumello D, Caironi P, Busana M, Romitti F, Brazzi L, et al. COVID-19 pneumonia: different respiratory treatments for different phenotypes? Intensive Care Med [Internet]. 2020 Apr 14 [cited 2020 May 11]; Available from: https://doi.org/10.1007/s00134-020-06033-2
26. Siddiqi HK, Mehra MR. COVID-19 Illness in Native and Immunosuppressed States: A Clinical-Therapeutic Staging Proposal. J Heart Lung Transplant [Internet]. 2020 Mar 20 [cited 2020 May 11]; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7118652/
27. Kollias A, Kyriakoulis KG, Dimakakos E, Poulakou G, Stergiou GS, Syrigos K. Thromboembolic risk and anticoagulant therapy in COVID-19 patients: emerging evidence and call for action. Br J Haematol. 2020 Apr 18;
28. Spiezia L, Boscolo A, Poletto F, Cerruti L, Tiberio I, Campello E, et al. COVID-19-Related Severe Hypercoagulability in Patients Admitted to Intensive Care Unit for Acute Respiratory Failure. Thromb Haemost. 2020 Apr 21;
29. Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost JTH. 2020;18(4):844–7.
30. Prescott HC, Calfee CS, Thompson BT, Angus DC, Liu VX. Toward Smarter Lumping and Smarter Splitting: Rethinking Strategies for Sepsis and Acute Respiratory Distress Syndrome Clinical Trial Design. Am J Respir Crit Care Med. 2016 15;194(2):147–55.
31. Dobbin KK, Simon RM. Optimally splitting cases for training and testing high dimensional classifiers. BMC Med Genomics. 2011 Apr 8;4:31.
32. Enguehard J, O'Halloran P, Gholipour A. Semi-Supervised Learning With Deep Embedded Clustering for Image Classification and Segmentation. IEEE Access. 2019;7:11093–104.
33. Pourahmad S, Pourhashemi S, Mohammadianpanah M. Colorectal Cancer Staging Using Three Clustering Methods Based on Preoperative Clinical Findings. Asian Pac J Cancer Prev APJCP. 2016;17(2):823–7.
34. Seymour CW, Kennedy JN, Wang S, Chang C-CH, Elliott CF, Xu Z, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. JAMA. 2019 28;321(20):2003–17.
35. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA

(Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med. 1996 Jul;22(7):707–10.
36. Diaz FJ. Measuring the individual benefit of a medical or behavioral treatment using generalized linear mixed-effects models. Stat Med. 2016 15;35(23):4077–92.
37. McMichael TM, Currie DW, Clark S, Pogosjans S, Kay M, Schwartz NG, et al. Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington. N Engl J Med. 2020 Mar 27;
38. Archer SL, Sharp WW, Weir EK. Differentiating COVID-19 Pneumonia from Acute Respiratory Distress Syndrome (ARDS) and High Altitude Pulmonary Edema (HAPE): Therapeutic Implications. Circulation. 2020 May 5;
39. Chen Y, Chen K, Ying Z. ANALYSIS OF MULTIVARIATE FAILURE TIME DATA USING MARGINAL PROPORTIONAL HAZARDS MODEL. Stat Sin. 2010;20(33):1025–41.
40. Laptev N, Yosinski J, Li LE, Smyl S. Time-series Extreme Event Forecasting with Neural Networks at Uber. :5.
41. Maya S, Ueno K, Nishikawa T. dLSTM: a new approach for anomaly detection using deep learning with delayed prediction. Int J Data Sci Anal. 2019 Sep;8(2):137–64.
42. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2. Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc.; 1995. p. 1137–1143. (IJCAI'95).
43. Iskandar R. A theoretical foundation for state-transition cohort models in health decision analysis. Gontis V, editor. PLOS ONE. 2018 Dec 11;13(12):e0205543.
44. Zhu X, Fu B, Yang Y, Ma Y, Hao J, Chen S, et al. Attention-based recurrent neural network for influenza epidemic prediction. BMC Bioinformatics. 2019 Nov;20(S18):575.
45. Merad M, Martin JC. Pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat Rev Immunol. 2020 May 6;1–8.
46. Fagone P, Ciurleo R, Lombardo SD, Iacobello C, Palermo CI, Shoenfeld Y, et al. Transcriptional landscape of SARS-CoV-2 infection dismantles pathogenic pathways activated by the virus, proposes unique sex-specific differences and predicts tailored therapeutic strategies. Autoimmun Rev. 2020 May 3;102571.
47. Aggarwal S, Garcia-Telles N, Aggarwal G, Lavie C, Lippi G, Henry BM. Clinical features, laboratory characteristics, and outcomes of patients hospitalized with coronavirus disease 2019 (COVID-19): Early report from the United States. Diagn Berl Ger. 2020 26;7(2):91–6.