AI Methods in Public Health (Part 1) Townhall-Meeting Recording

October 26, 2026, 4:00PM

1h 32m 7s



Belair, Marc-Andre (PHAC/ASPC) 0:08

Good morning, everyone. Welcome to our session this morning. I will give folks about a minute or so to filter in before we get started. So please just take a moment and get yourself settled in, and we'll be there in just a moment. Thank you so much.
Good morning for those who are just joining us. We'll get the session started in just a quick moment, so please take a moment and get yourself settled in.
Okay, good morning, everyone, and welcome. Good morning here in Vancouver. Good afternoon, or good evening, depending on where you're joining us from today. Welcome to the second session of the AI for PH [Public Health] Seminar Series: From Pilot to Practice, AI Methods in Public Health, Part 1. My name is Marc André Bélair, and I'm an epidemiologist and trainer at the Training and Development Unit at the Regulatory, Operations, and Emergency Management Branch at the Public Health Agency of Canada. And I'm based here in Vancouver, BC, and I'm honoured to be your host today. Une note pour mes collègues francophone, la session d'aujourdhui se déroulera uniquement en anglais, mais la vidéo de la session sera disponible en ligne avec des sous-titres et transcrit en français sous peu. Par contre, je vous invite à poser vos questions en français et j'aurais le plaisir de les traduire pour vous.
Just before we get started, I want to take a moment to acknowledge that I have the honor and privilege of living, working, and playing in the lands and waters in the unceded traditional territories of the xʷməθkʷəy̓əm (Musqueam), Sḵwx̱wú7mesh Úxwumixw (Squamish Nation), and səlilwətaɬ (Tsleil-Waututh) Nations and the Coast Salish people. I'm grateful as a settler and colonizer for their stewardship that they've had over this wonderful region. As we explore this topic of AI methods in public health, I encourage you to consider how the methods we choose, especially when working with Indigenous communities, can impact outcomes, perceptions, and stigma, and always seek to preserve and elevate Indigenous dignity and data sovereignty.

I'll be starting a poll here just shortly. As part of our work, we want to better understand who you are and what your experience in using AI in public health is. So just starting a poll here shortly. Perfect. That should be popping up on your screen momentarily. Based on our registration statistics, we have attendees joining us from all over the country and the world. Epidemiologists are the most represented here today, with data analysts, program and policy analysts, and medical officers of health and physicians close behind, and of course, the ever-present catch-all "other" category. Interestingly, there's a nearly even split among functional roles between researchers, scientists, academics, frontline practitioners,
or technical specialists and senior advisors/managers, followed up by a strong contingent of students and trainees, and a sprinkling of senior management and executives and decision makers. Approximately 40% of our registrants identified as early career researchers, and 70% of our registrants identified as working primarily in Canada.
Which means we have a large international contingent today. Bonjour and welcome. And finally, what's particularly interesting is that we have an almost even split between those working at the local, provincial and territorial, federal and academic domains. So I'd say we have some pretty good representation from across the jurisdictional landscape, which is always a pleasure to see.
As you're answering the poll, we'd also like to mention that we're hoping to help you make connections with each other. To that end, we've created a form that you can fill out where you can share a project you're working on and seek out partners or to share your expertise with others. You'll find a link to this form in the pinned comments in the Q&A section. So please take a moment to fill that out at some point, and we'll be sharing that information with participants
after the session in the hopes of enabling you to make some new connections. For attendees, we'll also be sending the link out as well as part of a post-session e-mail.
As in our last session, sorry, let's just take a quick second here to view the results of the poll. We have, let's see here, oh.
I see. Hold on a second. Has it not refreshed?
Okay, poll answers are starting to flow in.
We'll give that a little bit of time here. Question.
Great, okay. So poll questions are starting to come in. We have a fairly good even split between those who regularly use AI to improve their work and using AI occasionally for specific tasks. We have some good spread in the confidence level

from those who are slightly confident over to very confident
on their understanding of AI and its use of public health. And the biggest barrier to using AI in your work seems to be trust in AI outputs, followed by organizational readiness and policy, privacy, and governance concerns. Okay, so in our last session, Intro to AI and Public Health, we learned about some of the fundamentals of artificial intelligence
and learned some of the benefits and cautions about integrating AI into public health practice. If you missed that session and would like to watch it, session recordings and transcripts are available at the Public Health Agency of Canada's Training Portal and the AI4PH website. One thing we heard coming out of the session was a desire to learn more about how AI can be used with examples of methods applied to real-life projects. So this is part one of “Moving from fundamentals to applications”. Our 3 speakers will be sharing some projects and give us a deeper understanding of the methods used. First, today, we'll hear from Dr. Lisa Lix. Dr. Lix is a professor of biostatistics in the College of Community and Global Health at the University of Manitoba, Canada, a Tier 1 Canada Research Chair in Methods for Electronic Health Quality,
and Director of Data Science in the George and Fay Yee Centre for Healthcare Innovation at the University of Manitoba. Her research expertise lies in statistical and machine learning methods for complex healthcare data and patient reported outcomes. Dr. Lix will talk to us about how data are the foundation for applications of AI. They'll explore different types of data,
issues that may affect their usability and usefulness for AI applications, and we'll touch on the differences between statistical methods and machine learning algorithms for mining insights from public health data. Then we'll hear from Gemma Postil. Gemma Postill is a MD PhD candidate at the University of Toronto. Her research focuses on the application of machine learning and predictive modeling to population health.
Her doctoral thesis investigates the use of machine learning to predict long-term outcomes among older adults following traumatic injury to enable improved anticipation of patient outcomes. Gemma's thesis work has received support from the Vanier Canada Graduate Scholarship and the AI4PH Graduate Scholarship. Gemma is also the recipient of the 2026
Adel S. Sedra Distinguished Graduate Award, which recognizes her leadership beyond the graduate program. Specifically, Gemma also serves as the student

education colleague at the Temerty Centre for Artificial Intelligence Research Education in Medicine, where she contributes to initiatives advancing AI literacy and education among future and current healthcare professionals
at the University of Toronto and beyond. Gemma will talk to us about how risk and prediction modeling are increasingly important in applications of AI in public health. She'll explore how statistical and machine learning approaches can be used to predict health outcomes and identify individuals at risk, increased risk. She'll discuss key considerations in developing and evaluating prediction models, including model performance,
interpretability, and generalizability.
And then we'll hear from Dr. Affan Shoukat. Doctor Shoukat is an SGI Assistant Professor in Data Science in the Mathematics Department and an Adjunct Professor in Computer Science at the University of Manitoba. He leads a multinational team developing the Hybrid Epidemiological Networks Platform, an innovative framework that blends advanced AI techniques.
with long-established epidemiological modeling supported by NSERC, CIHR, and the French National Research Agency. Affan will discuss how AI is being used to forecast infectious diseases, provide a brief introduction to methods and data, the data used in forecasting, and discuss how public health decision makers can interpret forecasts in terms of its practical implications,
accuracy, and uncertainty. And then following our presentations, I'll invite Dr. Dan Lizotte to host our Q&A session. Dr. Lizotte is an associate professor in the Department of Computer Science and the Department of Epidemiology and Biostatistics at Western [University]. His research aims to support health decision making by developing and applying machine learning and statistical tools to new sources of data.
including electronic health records and social media, to better support patients and health professionals, particularly in public health and primary health care. His methodological research combines machine learning, optimal sequential decision-making, and multiple objective optimization. So just a note, we will be using the Q&A function of
Teams today. As you listen to our presenters speak, I encourage you to add your questions using the Q&A function and to upvote questions already posted that you'd like to see answered. But it's enough about from me. I think we're all interested in waiting to hear from our presenters. So I'd like to invite our first presenter, Dr. Lix

to share the screen and so that we may start the presentation. So we're just getting that pulling up now. I'll give you a second. Good morning, Dr. Lix. How are you today?

Lisa Lix 10:21
I'm doing well, thanks. Okay.

Belair, Marc-Andre (PHAC/ASPC) 10:22
Excellent, wonderful. I see your screen is coming up here. Perfect.

Lisa Lix 10:28

Let me just...

Belair, Marc-Andre (PHAC/ASPC) 10:35
Excellent.

Lisa Lix 10:36
Okay, it's in full screen mode then.

Belair, Marc-Andre (PHAC/ASPC) 10:37
It’s in full screen, yeah.

Lisa Lix 10:38
All right. Thank you very much, Marc-André. And at the outset, I'd like to acknowledge that where I'm based, at the University of Manitoba, are the original lands of the Anishinaabeg, the Anisininew, the Dakota Oyate, Dene, and Inuit peoples, and it is also the national homeland of the Red River Metis.
And we recognize that the reconciliation process involves thinking about data and ownership and responsibility for data, which is the focus of my presentation today.

I'd like to thank everyone for joining this second session in AI methods for public health. where you will find complementary perspectives provided by myself, Gemma, and Affan as we cover topics related to big data and modeling.
Data is the foundation, which is why I'm leading off today's session. So when we think about the value of clean and well-organized data, it can improve the reliability and accuracy of predictions.
Noisy data or errors in the data can affect the performance of machine learning models. And removing potentially anomalous or erroneous data points can affect the accuracy of a supervised machine learning model, such as one that's used for prediction, as Gemma will talk about today.
And finally, AI models are data hungry, so they require lots of data, and diversity of data can help with that. But where that data comes from is something that we need to really consider, and that can affect really the role of data in decision making as well. Marc-André,
just let me know that my slides are advancing?



Belair, Marc-Andre (PHAC/ASPC) 12:24
You are advancing just perfectly.

LL

Lisa Lix 12:26
Okay, that's great. Glad to hear that. All right, so that's hence my topic beyond the numbers, looking at health data and also statistical methods and how they relate to health data will be the focus today, an introductory session. Those of you that had joined the first presentation that was given by Laura Rosella,
will have seen this framework, our AI for Public Health National Training platform, which offers a plethora of opportunities for individuals to learn about the use and applications of AI in the public health domain. This framework identifies data in two different locations. First of all, as a foundational component,
and as we're teaching trainees around AI methods. We also touch on the role and importance of data and the tools by which we apply different kinds of AI methods to data. The second is in the area of health equity, a concept that permeates our discussions.
Having access to measures that enable data to be disaggregated, for example, by

attributes such as those for equity deserving groups is important. Being able to conduct stratified analyses is a component of that. Secondly, data are important, equity
in data are important from the perspective that we need to know which groups may be missing from data or underrepresented in the data. Being critical consumers in this respect and critical users of the data is important when we think about data and AI applications in the public health sphere.
So let's start off with an example that may help to dive into the importance of data. The US National Institutes of Health has a really diverse data set called All of Us, research data that are used to study population health and health service use. The data can be linked to a variety of sources. A recent study examined
unplanned readmissions to hospital for patients who had sepsis, which we know is a complex condition. And up to 1/4 of all patients with sepsis are likely to be readmitted to hospital.
The investigators of this particular study were interested in using machine learning models. And what we know about kind of traditional models for looking at sepsis readmission was that they focused primarily on demographics and past medical history in order to build risk prediction models.
Well, this particular study was focused on adding the social determinants of health into machine learning models. So linking together a variety of data and found that in fact, using these rich data resulted in better predictions. So including such variables as insurance coverage, which we know is really important in the US,
and also the stability of insurance coverage.
Other social determinants of health, such as educational attainment, the stability of housing, and also employment status, were important. So bringing together a diversity of data,
and using machine learning models resulted in improved prediction. This study also highlights the power of data linkage. So often we arrive at rich data for use with AI applications by linking together a variety of diverse data sources.
And we know that in Canada, that's possible, often at a deterministic level. In other words, we know individuals are in different data sources and are able to accurately link these data sources together. It's a really cost-effective and efficient way to be able to gather information
and from a variety of domains that can be important in terms of decision making that goes on in the public health sphere.

Bringing together many different kinds of data can assist in building accurate machine learning models. And in Canada, we have a diversity of data to be able to use. Examples include survey data, such as what comes from the Canadian Community Health Survey or the Canadian Health Measures Survey;
Social media data are used heavily in the public health sphere, such as data from Twitter, Facebook, and Reddit that can be used to look at behaviours or perceptions or interactions with different kinds of content in the social media domain.
Administrative data, which I use extensively for my research, which is all about managing the health care system, and hospital records and prescription medication records exist in data enclaves in different provinces, such as ICES in Ontario and the Manitoba Centre for Health
Policy in Manitoba.
As well, there are registry data focused on individuals with specific health conditions, such as cancer, or who require specific procedures, such as hip replacement. And then electronic medical records, which are increasingly being explored because of the rich variety of structured data
as well as narrative or unstructured data that can be used to explore the social determinants of health and also health behaviours.
But many of these data sources are secondary data sources other than for survey data. So they're used, when we think about public health applications for purposes other than that for which they were originally intended. So if we think about administrative health data, which are used for managing the health care system, though…
They can give us a population-wide perspective on population health and also health services use.
So being a critical consumer of these data, because we can't control how they are collected, is important.
And being a critical consumer means recognizing that there can be biases in secondary data sources, which are so valuable for public health surveillance and investigation. And these data can take a variety of forms. For example, the data itself can have missing values in it. It can incompletely represent particular populations. Or there can be errors in entry, because those that are entering the data aren't necessarily thinking about the quality of the data for surveillance or for research purposes.
When we think about healthcare data and how it's used in the public health domain

with AI applications, we recognize that individuals may implicitly have biases that influence how data are captured. Those individuals who are represented in healthcare data
are those that contact the health care system. So there may be referral biases. Not all individuals may end up getting admitted to a facility, for example. And so there can be issues of representation of individuals who may be medically underserved. Diagnostic information that's reflected in electronic medical records may not always be accurate. And there's the potential that different groups who may have differences in risk may not be well represented in the data.
Then we think about biases that can come from equipment. For example, with clinical registries that may capture clinical measurements for individuals. If equipment malfunctions, for example, we may have errors in the data. So being able to spot and recognize these errors is a critical part before even thinking about using AI with these data.
So let's look at some examples. If we think about social media data, it only represents particular populations that engage with social media. So particular demographics are represented. If we think about using social media data to explore public health topics, there can be bias in terms of the methodologies that are used.
For example, a lot of these data are in unstructured text format. So the choice of keywords that we might explore with the data and the hashtags that we might use will have an influence on what we may extract or retrieve from these data. And that could influence the results of our analyses. Natural language processing techniques are widely used with social media data. The choices that we make in terms of pre-processing these data, for example, how we remove stop words such as and, or, or the;
how we apply other kinds of processing methods, removal of punctuation, for example, can result in different interpretations of the data.
And there's a really nice review paper here from the International Journal of Medical Informatics that talks about using social media data for public health surveillance and the challenges that can exist, particularly in terms of population bias and methods bias.
Thinking about clinical registry data, there again can be biases in terms of who's represented in these data. It may be individuals who are most ill, for example, or have the most severe forms of disease. Therefore, completeness of [data] capture
of particular populations. And because many registries exist over time and

accumulate data, there may be different policies or practices that can result in incompleteness [in the] capture of the data.
Differences in recording of clinical measurements, for example, biological measurements, blood work, for example, and machine-related biases can also impact on the quality of the data. And for these reasons, it's important for us to take a variety of considerations -
a healthy dose of them - into our use of secondary health information. So I've touched on some of these around linkage characteristics and the quality of the data, but others that can come into play in terms of gaining access to the data include who's responsible, who has ownership for the data.
And what are the policies around access that may exist? Has consent been provided and how does that influence our ability to access certain types of data? And then overriding all of this is a consideration of the security and privacy of health data and how that influences access. This again can result in
biases, if particular groups of individuals are excluded, or if particular types of data are not made available due to concerns about privacy and security. So how can we address these issues or have our eyes wide open as we explore them?
Well, an example that comes from my own province and the Manitoba Centre for Health Policy, which houses vast repositories of administrative health data, clinical registry data, and electronic medical records, is that there's a systematic process available for us to explore data quality and understand how it might impact on our ability to do public health,
surveillance, and research.
We recognize that data quality is a multi-dimensional concept. While we might initially think of concepts such as accuracy and validity of the data, both internal and external, timeliness and our ability to interpret the data are also important. The pandemic highlighted for us
the importance of timely information for decision making. The time to be able to provide information from these data is clearly a data quality dimension. And also interpretability, having available documents, information, metadata about the data. can influence the choices that we make in terms of measures that are included in AI-focused models and how we interpret the outcomes. So we can drill down into each of those data components. But equally important is that we have tools. So in this particular case, there are macros,
or algorithms, that allow us to explore the data, looking at various aspects of data

quality, including trends over time, how we can link the data together, and who's missing from these linkages, and how valid the data are when comparing across sources.

Another topic that I'd like to touch on with respect to these considerations of using secondary data is around privacy and access to data. And with respect to this, AI is increasingly being explored to produce synthetic data. In other words, data that have been generated using some sort of purpose-built model or algorithm.

and purpose-built for a particular solution. For example, and I'll go through a few examples in my next slide. And synthetic data are increasingly being used in the public health sphere because a variety of kinds of data can be generated and have value in terms of our being able to explore applications in AI.

So we could generate synthetic image data or time series data or even text data that might look very realistic in terms of what comes from social media sources. But the benefits of using synthetic data are that with these purpose-built

AI algorithms that can be used to generate the data. We can provide access to data that are in fact not real, preserving data privacy and assurances to data custodians that real data may not be disclosed to individuals.

Thinking about AI models, synthetic data can be particularly beneficial as well for developing and testing out and then training these models, which are very data hungry. With this in mind, some of my application areas have been around how we can augment existing data. In other words,

add new data to existing data sets that better represent groups that may be underrepresented in real data. And this is particularly valuable in terms of training the next generation of individuals in how to use secondary health data for a variety of

purposes.

So the risks of exposing personal health information through use of these data in public health applications means that synthetic data are inherently appealing. They can be freely shared without privacy restrictions, and that can speed up access to data.

Also, helping to build out a diversity of large-scale databases that reflect diverse populations, potentially rare health conditions, or other specific scenarios where we may have a scarcity of real-world data can be a strong selling point for using synthetic data in AI applications.

And related to that is that they can help us to train machine learning models, for

example, and also validate these models, looking at a variety of what-if scenarios where we want to build prediction models. They can help address biases that may exist in the data so that we have more equitable
machine learning algorithms that will help with policy decisions to diverse populations. And all of this really is about accelerated innovation, being able to access data more quickly, ensuring the privacy and confidentiality of these data. And allow us to explore new applications of AI algorithms.
All right, so I'll switch gears just a little bit as I come to a close in my presentation. I've talked a lot about the different kinds of data that may be used in public health surveillance or in applied research applications.
Equally important, just thinking about how secondary or primary data are used is managing those resources, making sure that they're available for future groups of individuals. And so having a very defined and deliberate framework for thinking about how data are collected or how they might be extracted from existing repositories.
How their quality can be explored, how metadata can be maintained, and how the data can be documented for historical purposes. It's critically important to ensure reproducibility of results and transparency of the methodologies that are applied to the data.
For example, having well-documented processes about any pre-processing applications that are applied to text data that comes from social media ensures the reproducibility and the ability of new users to try out different what-if scenarios.
Now, being able to use secondary health data in a variety of applications. When we think about, should we be using newer machine learning models? What about, can I not use existing statistical models with these data? And I think it's inherently important to consider various characteristics
of the data, the complexity, for example. I talked early on in my presentation about the ability to link many kinds of data, which can produce really high dimensional data, in other words, a lot of variables. And when that's the case, it may be beneficial to be working with machine learning models. And we're working with unstructured text data,
Natural language processing techniques are much more efficient than trying to wade through the data using rule-based algorithms.
The data size and relatedness will affect the choice of models that we might apply to the data. Where there are a lot of variables that may be included in a data set, and

many that may potentially be redundant, look at applications like a lasso regression model or a tree-based recursive partitioning machine learning model.
as being able to help us sort through or sift through a large set of potentially redundant variables to hone in or be able to mine through the data and identify the important variables. When we think about using statistical models, we're trained early on to recognize that these models
often rely on data that have specific characteristics, such as a normal distribution, that the data are independent, for example, that they have a constant variance across different groups, and that there are linear relationships represented in the data. This might mean that thinking about messy data would lead us to say that machine learning models are
superior because they make minimal data distribution assumptions. But remember that these data are, that these models are hungry. They require large amounts of data to be able to learn data patterns and relationships. They're not always efficient. We can end up with very small cell sizes, for example, even in very large databases. And that can stymie machine learning methods and means that they may not be the optimal choice. So with that in mind, we have to think about the question that we're exploring, as well as what our interests are in being able to describe or interpret what's going on in the data when thinking about the data requirements. And then also we think [about]
the tools that we might be most familiar with or have skills with may influence the decisions that we make as well. All right, so in closing, what I've talked about today is primarily about secondary health data and being critical consumers of these data, recognizing that biases exist in them that may influence the accuracy
of machine learning models that are applied to the data and also generalizability from the perspective of representing entire populations. And ultimately, the value of these models and the data for decision-making purposes. While machine learning models do not necessarily make stringent assumptions about the data that
that they may be working, that we may be working with, they may not always be the optimal or the needed choice. So being critical consumers of AI models, I think is important. We have to consider the research question, as well as the complexity of the data and its other characteristics of it. So with that, I'll stop. Thank you very much for
attending, and I'll turn it over to our next speaker.

Belair, Marc-Andre (PHAC/ASPC) 34:33
Thank you so much, Dr. Lix. I appreciate your time today. Just as we get our next presentation, Gemma Postill ready, just a reminder to folks that you can ask questions using the Q&A feature, which is available there in Teams. And we're seeing some questions starting to come in, which I'd love to see. Additionally, there's some questions that have come up about whether slides will be shared. The slides will be made available on the PHAC Training Portal, and attendees will receive an e-mail roughly about two weeks with details on how to access that, as well as the recordings for the session. Okay, Gemma, I see your slides are coming up here on the screen. I will
put you up. Are you ready to go?

Gemma Postill 35:16
I'm ready to go. Thank you, MarcAndré.

Belair, Marc-Andre (PHAC/ASPC) 35:18
Excellent, I hear you loud and clear and your slides are up, so over to you.

Gemma Postill 35:23
All right, that's the two first criteria for a good presentation. And thank you to Professor Lix for starting us off. This presentation will pick up on a lot of threads on that. So we're going to move from the data towards an application and specifically using AI for prediction and risk modeling in public health. And I want to first start by acknowledging the Population Health Analytics Lab. So this is where I'm currently doing my PhD.
The lab is headed by Professor Laura Rosella. And the work I'll be talking about today is both my own as well as some of the broader applications of AI and public health that the lab is working on.
All right, so today we're going to talk about an introduction to prediction and risk modeling in public health. We'll talk about statistics versus machine learning for prediction. We'll talk about how to build and evaluate a machine learning prediction model in public health, and then end with some general principles on the responsible

use of prediction in public health.
All right, so let's begin with what are we trying to predict and why does it matter? So to contextualize our topic in epidemiology and in public health, there are three main domains of questions that we ask, descriptive, predictive, and also causation. So we're looking at what does a population look like or what does a disease look like? We're also looking at who is likely to experience the outcome, and that's what we mean by prediction. And then causation, what would happen if we intervene, for example, or what are the factors that truly cause this phenomenon. And today we're going to focus on prediction.
And why do we ask prediction questions in public health? Well, we might want to know who is going to develop a disease. We might want to know how that risk will vary across groups in the future. We might want to know how many cases we're anticipating to have, where could service demand rise, and also how should capacity be planned.
And prediction models can help us answer those questions. So one example, and throughout the presentation, I've also included some QR codes in case you want to reference the material. And as well, later on, I'll point out where those QR codes are, but we've included some coding resources that are publicly available that you're welcome to
access to help contextualize some of the things we're going to talk about today. All right, so first we'll talk about DPoRT. So this is a validated population model that predicts diabetes risk among a population. And with DPoRT, we can ask several different questions related to public health.
We can ask about the individual level, so who is likely to develop diabetes in the next 10 years. We can look at population subgroups, so which groups have the highest burden or predicted burden of diabetes. We can look at it geographically, which communities, for example, are expected to experience the greatest rise in the burden of a specific disease.
And we can also look at it temporally. So how will the future burden of diabetes vary over a given period of time?
Now, today we're here to talk about AI. So how do these statistical models like DPoRT, which is a classic statistical prediction model, vary and evolve in the context of AI? And does a more complex algorithm, as Dr. Lix mentioned, even improve the decision? And how do we know that?
So I'm going to throw around a few different terms throughout the rest of the

presentation, but I really want to ground us in a similar understanding. So AI is a broad field of studies in which we use computers to do human-like things, so learning, vision, language, et cetera.
And when we say machine learning, this is a domain within AI, and it refers to algorithms where the computer learns patterns or identifies patterns within a data without their explicit programming. So AI is a broad field. I often think of it like an equivalent to biology, for example. And then
Machine learning is a field within AI, and I kind of think about that as like microbiology. So the key here is that machine learning is not just one thing, and it's not just one model. When we talk about using machine learning for risk prediction, we're talking about many different types of models. So there's regression-based models, there's
tree-based methods. There's boosting methods, which are an extension of these tree methods. We can also use support vector machines for classification, neural networks, and deep learning. And there's also machine learning applications of our typical survival models, survival analysis models. And the appropriate method within this field of machine learning depends on the data you have and also what you are trying to predict.
So one example, just to highlight this point further, that ML is not just one model. So we're all familiar with decision trees. These have been used for a long time in public health and also in clinical medicine as well. And so, for example, you would have a decision that you're trying to make and you would go through many different steps or nodes to get to that decision.
So for example, let's say we're trying to classify somebody at risk of diabetes or heart disease, for example, you would pick the most important variable, and that is sort of the primary node. And then subsequently, you would make the other nodes from there. And we would use either data or clinical knowledge to construct these decision trees.
Machine learning takes that principle and extends that even further. So random forest, which is a machine learning prediction model that some of you may be familiar with, takes these decision trees and then you as the researcher get to set up the parameters of the forest. So for example, maybe you'll enable up to 1000 trees or 10,000 trees within the model.
And what happens is that these answers are bagged together. So the model, when it is given the data, learns the relationships between the outcome, let's say we're using

diabetes, and then the input variables. And it learns that relationship by constructing many different types of trees. Some of the trees might have
3 variables, for example, age, sex, and BMI. And another tree might have a totally different set of three variables. Maybe it uses location, maybe it uses ethnicity, et cetera. And then what all of these trees do in the model is that when given an input, they all make a guess or a calculation as to what the output is. So for example, having diabetes or not. And then those answers are combined. So if we have 1000 trees and each is voting based on the variables that it uses within its tree, those answers are added together. So we can see that
as compared to a regular decision tree, this is a much more complicated model, but it also has the potential to customize or maybe be better fit to the various different individuals that exist within a population. Boosting is a similar but slightly different application of this. So the same idea applies with the different decision trees. However, the answer moves through the trees within the model, and then it uses that to come to the final answer. So one way that when we're teaching AI in a classroom setting that we think about it is, imagine you have many different desks set up within a classroom. Group one is given a problem sheet and tries to solve the answers. They write down their steps as they go.
And then they pass that on to the next group who double checks their work and adds in maybe their own perspectives or other notes. So that's how a boosting tree works, for example. So as you can see, these are very different structures to these models. And they have different applications. Boosting trees have shown to work quite well in public health
and also clinical medicine prediction related problems. And so they're quite commonly used. And we'll show an example shortly of that.
And by shortly, I mean on the next slide. That was perfect. So, DPoRT, as I mentioned, is a diabetes population risk tool created by the Population Health Analytics Lab. DPoRT ML is an extension on that. You're welcome to scan the QR code if you want to go to the paper to read more. But as Professor Lix mentioned,
the value of machine learning here is being able to parse through complex data. We created DPoRT ML at the Population Health Analytics Lab, not because DPoRT didn't work well, DPoRT works fantastically, but because we wanted to ask a slightly different prediction question with slightly different data.
So DPoRT uses CCHS data, whereas DPoRT ML takes administrative health data. And we do a process where the different healthcare, the points of which individuals

access healthcare are reconstructed into patient-specific timelines. And that information is used and run through a
gradient boosting decision tree models, so what I talked about at the end, those boosting models, to predict complications within an individual as opposed to just risk of diabetes. And with this more complex prediction challenge, ML added quite a bit of value. And in most papers that construct ML models, there's often a benchmark of using a traditional statistic model such as logistic regression
and comparing whether or not ML actually adds value in the prediction. And by adding value, I mean potentially improved discrimination or improved ability to predict.
And so the key difference here is that ML is being used on more complex data set, and there's a more complex relationship as well between the input variables, so prior health care utilization, and the outcome, which is specific complications from diabetes.
So complexity should never be the goal. It's really a choice. And this we're increasingly seeing within the AI literature. I think initially it was very exciting that AI models existed. And so a common goal of research was to try and see whether you could apply AI here. Could it work within this?
data ecosystem, for example, do we have the compute power to do this model or run this model? Versus now there's increasing understanding that we can do all of that, but really when should we? And so as I mentioned, it's about complexity and whether the relationship between the input and output features is complex.
And also whether the data itself and the structure of it is complex.
And so machine learning models as compared to traditional models, just in summary, are more complex in their architecture. And they're also more flexible for that reason. That complexity, as Professor Lix mentioned, really enables them to learn more nuanced patterns.
As opposed to logistic regression models, for example, which require some additional assumptions about the data and require the data to almost conform to the model. Whereas with AI and with machine learning, we're being more flexible and the model is conforming to the data.
All right, so let's build a model then.
And up here, I have two additional QR codes, so you're welcome to scan those and access these resources. I have both the personal GitHub repository and the Population Health Analytics Lab also has a GitHub repository. And as I'm going to

show in the next few slides, there are a lot of choices we make when we're building a machine learning model.
And this really demands increasing transparency in how these models are reported and shared.
Okay, so as I mentioned, you first begin by specifying a use case, which should not be, I just want to build a machine learning model, but rather I have a complex prediction challenge. We then construct our cohort and select the predictors, as we do with logistic regression or typical statistical prediction models.
Predictors must be available at the time point at which the prediction is made, so ensuring there's no label leakage between the variables that go in and the answer that you're predicting after. The data is split, and then we train the model. It's very important that half of the data or part of the data -
usually 20 to 30% of the data - is held out so that we can test how well the model works on that separate data set. And then we evaluate whether or not it has an action and whether we can do something with this model.
So when we're training the data, so that's that middle step that I was talking about, we've taken 70 to 80% of the data, we've given the machine learning model the input and the output, and we said, learn the relationship. There's a few other decisions we make as a researcher, and this is also what makes machine learning models flexible, as we call it. So there are the model parameters or how the model is set up. So for example, when I described a random forest, the forest, the structure, the fact that there is trees within a forest, those are the model parameter. The hyper parameters are things we control and things that we can also fine tune
to improve our prediction. So these, for examples, would be how complicated your trees are, how many trees are in the forest, how many variables are needed, or how many instances are needed with a given variable for it to form a node or a branch within a tree.
And there's various approaches within the ML space for selecting the hyperparameters. So often, for example, we would set up a range, and we would say, okay, well, I'm going to look at different levels of complexity. I'm going to do 400 trees, 600 trees, 1,000 trees, and 2,000 trees, and see how good my model is at predicting with this different amount of trees.
And grid search would systematically search parameters within that range. So for example, if you are also varying not just the number of trees, but also the depth of the trees, it would test every combination of those parameters, which is a more

rigorous way to ensure you're truly optimizing your model.
But it takes a lot of compute. And so another good way, which is slightly less rigorous, but still can produce excellent models, is random search where you try various combinations and optimize from there.
So, well, wouldn't you just want to add as many trees as possible and wouldn't you want to make them as complex as possible? But no, unfortunately not. So underfitting, and this is sometimes the criticism of traditional statistical models, although not always the case,
some can be fit quite well, but the worry that you haven't truly captured nuance within the data and nuance within your population in terms of the relationship between the input and output variable.
Overfitting, by contrast, is when we train a model, let's say we added 10,000 trees to our random forest model, and we said, learn the relationship, but maybe we only have 8,000 patients within our data set. Well, if we have 10,000 trees, that means effectively the model could probably basically memorize the answer within the data. There would almost be like more than one tree per person in terms of predicting perfect input-output relationships. So ideally, we want something that's more balanced. So how do we know that our model is not overfit and is rather balanced? This is where the testing data comes in. And testing data is not a new phenomenon. Traditional prediction models have been validated on external holdouts within the same data set and even on external data sets. And we can tell whether our model is overfit in the ML space if we're comparing
the relationship between the performance of the model on the testing data versus the training data. So the training data gives us our bench line. And if, for example, the performance drops drastically when we give it new data, that means it's probably overfit or it's essentially memorized patterns between input and output combinations.
So what do we do then if we've got a model that's overfit? Does that mean ML is not the right tool? Well, not quite. We can fix overfitting. And so ML as opposed to like our traditional approaches within epidemiology and public health is a bit more iterative.
So if we see that the model is overfit, that doesn't necessarily mean ML is not the right approach. It typically means we go back to those hyperparameters and change it, or even refine the specific model architecture that we're using.
All right. When I'm saying evaluating and seeing whether or not the performance

drops, what exactly do I mean? The previous example used an area under the receiver operating curve, which is a very common model for evaluating machine learning, a common metric for evaluating machine learning models.
But when we're applying ML to public health, a ROC or AUROC isn't always the best metric and certainly should never be the only metric. It evaluates how well the model has learned patterns, but when we're applying these models to public health decisions,
we need to think about the specific sensitivity and specificity of these models as they determine the false positives and false negatives that we're going to use. Now, whether we want to maximize or minimize false positives versus false negatives very much depends on the public health challenge that we are speaking about.
And so that's where domain expertise really adds quite a bit of value to this typical machine learning workflow. And another thing that I should mention is when we're training the models and selecting the hyperparameters, we have the option to, the default setting of these models is usually to optimize
AUROC. However, we can also, as researchers, specify whether the model should be optimizing sensitivity or specificity.
All right. And the threshold that we will select very much determines on the public health action in mind. I'm moving quickly through these last three slides just to stay on time. But calibration and discrimination are very much not the same thing. I'll share an example from my own research. So
here, this is part of my thesis work, which is still ongoing, but we are predicting one year recovery after injury among older adults. And in this instance, I've developed a logistic regression model, model A, and a machine learning model, model B. Now, what's really interesting is that the machine learning model has a higher AUROC, However, you can tell it's slightly less calibrated. So it very much determines what your goal is with the model. If it's just to classify, potentially model B is slightly better than model A. However, it does either overestimate or underestimate risk at the extremes.
Which if you're using machine learning prediction models to risk adjust, would be problematic.
Now, as sort of a final example, Amazon had an AI model very early on in 2018 that was helping to triage reference or application letters. And they wanted to spend more time on the good quality application and less time on the bad quality applications.

And what they found out very quickly was that their model displayed kind of domains of, we would say, bias. A woman apparently had to be much more established for the resume to get through the first round of triage by this AI algorithm.

And why did that happen? Well, the model actually penalized words like female because it hadn't seen those in its training data. It was given data that had resumes of current employees at Amazon and said, this is examples of good references or good resumes. And apparently the word female wasn't really seen in there very frequently. And so

the model learned that that was the sign that the resume probably wasn't going to make it. Now, this example is older now. AI has moved much further. But why do I point this out? Because it's a bias not driven by the model. It's not a bias added by the researcher. It's something present in the data.

So with ML models, what we're increasingly encouraging is subgroup evaluation. It's not enough to say that the model has an overall accuracy rate of 0.89. We want to think about what it looks like on each of the different subgroups. And the difference that we're going to accept or tolerate as being okay

between those different subgroups will obviously vary by the public health decision that the model is designed to support. So for example, is a 0.90 performance on females versus a 0.88 performance on males acceptable to the population? That very much depends on what the model is predicting.

And to conclude, model explainability. So this is something that's a bit different with machine learning prediction models and worth calling out. This is a different model that we developed to predict recovery after injury among older adults. Now, the features that are determined important by machine learning models, you can see with a SHAP analysis.

So it looks at the features that are weighted most important. Now, while these SHAP analyses are great at kind of showing us that our model has learned probably the right relationships in the data and adding a bit of transparency and interpretability to the machine learning model, because with machine learning, we add so many different predictor variables.

The importance is often weighted across these many different variables. And so the interpretation is not quite the same as coefficients within a logistic regression model. The wrong take away would be like, oh, the most important predictor of being alive and at home is days with congestive heart failure.

Because we construct the models differently, the coefficients and their importance represents something different with machine learning models.
All right. So I will conclude there with just a final note. That is to say, our responsible use of AI and prediction models very much depends on you as excellent researchers in terms of applying the right models to the right data, ensuring transparent reporting of the steps used in the models.
as well as rigorous evaluation.



Belair, Marc-Andre (PHAC/ASPC) 59:09
Excellent. Thank you so much, Gemma, for your presentation. We really appreciate it. Now I'd like to invite our next presenter, Affan Shoukat, to join us and share his screen. And we will get that sorted. And while we wait for that to get set up, just a reminder to folks that if you'd like to ask questions, the Q&A is active.
And Affan, are you there with us? I see slides coming up. Great. Perfect. No worries. Thank you. And over to you.

Affan Shoukat 59:36
And...
Yeah, thank you.
Thanks, Marc-André, and thanks, Gemma and Professor Lix, for the talks.
Following on Gemma's talk on prediction, you know,
what I would like to talk about is a little bit more about forecasting.
In other words, given what we know today, what is likely to happen over the next coming weeks or months if the current conditions continue? And I'll also show how we can determine whether those forecasts are correct, how we score them, and more importantly, and Gemma touched on this at the end, how do we interpret the results?
So, so let's get started.
The value of forecasting is not really predicting the future. You know, that's not where the value is. The value is supporting make better decisions and allowing public health agencies to be proactive rather than reactive. And so forecasts can help determine, for example, how many beds and staff is needed in a hospital,
when to launch vaccination campaigns, whether medication supplies are adequate,

whether to implement vector control, and so on and so forth. And actually, over the last decade, there have been several forecasting hubs or competitions created by public health agencies, not just in Canada, but worldwide.
And in these hubs or these platforms, individual groups or teams can submit their forecasts in real time using whatever methodology. It's not really confined to machine learning, but I mean, you know, your traditional statistical or mathematical models are perfectly fine. A fairly popular one that you may be familiar with familiar with is the CDC FluSight platform, which has been running since 2013. And in Canada, we have an AI forecasting hub, which is a relatively small hub. It's based out of the University of Guelph, run by my colleagues Monica and Thommes.
During COVID, there was a few platforms that popped up. This was the COVID-19 forecasting hubs in the US, the European hub, and as well as the German hub. And I think the last one I'm familiar with is the Brazil dengue forecasting competition.

There, you know, we should be a little bit careful with what we mean by forecasting. So I just want to quickly go over this, some of the different flavors of forecasting. The first one is what we call nowcasting, which asks what is happening right now to essentially fill or correct today's missing information.
So, you know, for example, you might see influenza cases reported today on a public health agency website, but today's reported cases may actually reflect infections from one or two weeks ago. And so now casting corrects for the reporting delays, it can correct for incomplete data.
And aims to, you know, really, it's statistically similar to forecasting, but it's also slightly different. Forecasting, on the other hand, predicts what's likely to happen in the near future, so days to weeks or even months in advance. And it is really this middle column of forecasting that helps with operational planning, preparedness, general decision making
and so on. And the last one is a projection or scenario analysis. Gemma called, you know, touched upon prediction. And so these are answered questions like, what would happen if this or if that, for example, a vaccination uptake reaches a certain target?
Or what if a new variant appears for a particular pathogen? And so projections are typically used for comparing options. They are judged on whether, on what, on the what if scenario, not whether they came true.

And most of this talk here is on the middle column. Like I mentioned, now casting is essentially, or I should say statistically, the same as forecasting. And in the context of this presentation or this session, I want to talk about how
Machine learning methods are increasingly being used for forecasting.

You know, the basics of AI have been covered before, so I don't want to go into details on how these methods work, but here's a couple, here's an overview, a simplified map of the modelling methodologies used, right? So if you look at from left to right, from left to right, if you read these methods,
each, sorry, let me just rephrase that. Let me just, yeah. So from left to right, it's basically how much the method knows about the world it's modeling. So how much of the complex system that it's trying to model, how much can we implement that through equations and mathematics. As it moves to the right,
Models typically learn more from the data rather than the assumptions about the system that is trying to model from bottom to top.
Are so from the models that are at the bottom, like mechanistic models and hybrid models, they typically require the least amount of data, and as you go up the ladder, they're more based on data, or they require enormous amounts of data.
Bottom left are where your traditional models sort of lie, the mechanistic models. Some of you may be familiar with SIR, SEIR model, agent-based models. And these are the type of models that encode the biology. They encode the transmission dynamics, even the statistical models to some degree.
Extrapolate trend and seasonality. And both of these require little data, I would say. And again, moving to the right are your machine learning methods, right? The tree-based methods, deep learning, the graph networks, and they need more data to learn from. Something interesting over the last couple of years are these hybrid models. So
these are neural differential equations or something called physics inform neural networks. And what they do is they encode the biology in their equations, but also they learn from the data. And so it's sort of, you know, the best of both worlds. But there's also significant problems with their interpretability.
And then last thing is I just want to talk a little bit about the foundation models. These are the odd ones out, right? So these are hugely data hungry. You need enormous amount of data to train them. The data could actually come from other fields. And so, but once they're trained, you don't need to train them anymore.

They're used over and over again.
And of course, the biggest ones, I think everyone's familiar with, with ChatGPT or Gemini, these are your typical foundation models.

And I kind of put these two boxes there, the ensembles, which are and early warning, which are, I think, important to touch upon. So ensemble models are not a single model, but they combine any one of these and, you know, kind of look at their average. And for early warning systems, these are systems, again, that use these models to flag
whether a change is coming or not, without predicting absolute numbers. Typically, early warning systems comes from a theory called critical slowing down theory.

So I mean, this slide is sort of a repetition from the previous slide. So I'm going to go over this quickly. I just want to touch upon this last box here, which is that no family of models really wins. What you need is the real question is, how do we...
combine them and how do we score these models and how do we compare them against some sort of baseline?
So let's talk about that. How do we actually score a model? Because I think it's good that we are all on the same page in terms of what makes a forecast good or not. Generally speaking, if a forecast just produces a single number, that's not enough. It doesn't capture the uncertainty that the decision makers are looking for.
Or, you know, the uncertainty that's in the data that's producing the forecast to begin with. And what we want is something like, you know, 90% chance that there's going to be 900 to 1600 hospital admissions over the next few months. And when you have something like that, when you have a statement or forecast like that, then you can plan surge capacity
around that, you can have operational preparedness around that statement.
There's two ways to score a forecast. The first is, actually there's a few ways, but the two I want to talk about are the coverage of a model. And so if a model provides a 95% range, then the observed value, the truth, should fall
inside that range about 95% of the time. This is very similar to what a confidence interval is in statistics, but it's not exactly a confidence interval. To work it out, what you do is you collect many forecasts across different times, different horizons, and then you count how many,
how often, the observed value fell inside the range, and then compare that rate with

the range's nominal level. So we'll go over that. I have an example of this in the next slide, so you know, I'll provide more context.
The other main measure that we use to score forecasts is the weighted interval score. And this is essentially a number from zero to, could be a very large number. And how it's calculated is that for each forecast, you look at
how far away you are from the range that you have provided. And that's, and then so you use that information to calculate a ratio, which is what the model is producing divided by what the baseline is producing. And I still haven't talked about what a baseline is.
That slide is coming up. And so if a number less than one typically suggests that your model is doing better than the baseline model.
So let's just go over this example. I think it's important to provide examples. Here I have...
On the left-hand side, we have 3 weeks, and we can see that we have different, we have the baseline model, we have the green, that's an ensemble model, probably made-up of a few models, few individual models, and then you have the truth value, which is the black.
And so we can see, for example, here we are outside the entire range of our predicted forecast, but inside the baseline range.
And the way to calculate the weighted interval score here is, for example, for the baseline on week one, week two, week three, these are the numbers it's producing. Let's say the ensemble is producing these numbers, and then the weighted interval score just ends up being the division of 6.9 and 11.3.
So if you divide these numbers, you get a relative score of 0.961 based on this toy example. And this is not saying that the model is 61% accurate. I think there is a common misconception there that a lot of folks think that the model is 61% accurate, whatever that means.
This is a relative score, so relative to the baseline, you have about 40% less error. The coverage sort of goes hand in hand with this. So in the baseline, you can see that the range for the baseline, the ranges for the baseline are actually quite large, and you can capture the truth value.
So all three ranges here had the truth value in them. But this also means that your range is too cautious. I mean, I can provide you a range that's saying the next, that says the next, you know, hospital admissions in the next six months are going to range from zero to 10 million. And

my range is most likely going to be accurate, but not very helpful. On the other hand, the ensemble got two of the ranges correct and then missed one. And in this case, the coverage was around 67%. And we might say that the model is quite overconfident because the ranges it's producing are
are tiny or small relative to the baseline, so...
What we want to ask is how we can balance the weighted interval score and the coverage.
And well, and the summary here is that the WIS tells you how good the forecast is, and the coverage tells you whether you can trust its ranges. You sort of need both, and they go hand in hand.
I've been talking a lot about the baseline, the weighted interval score, for example, is relative to the baseline. And there's really two different types of baseline values that we can compute. One is the naive baseline, and this is just saying...
It's just taking the latest value. It's basically saying the next week will look like this week, or the next month will look like this month. And it takes a forecast number, it takes the latest number and adds some uncertainty around it, and that becomes your baseline naive model.
There's another way to set up a baseline, which is a historical baseline. This says that.... this looks at forecasts.
It measures forecasts and what usually happens at a particular time of the year based on past seasons. So it ignores the current season's data in many cases. And it's only looking at historical data. And sometimes this is, you know, we see this in weather reports, right? So we sometimes hear
weather report saying it's unusually warm for November in Ontario. And that's because at this time in the last 50 or 100 years, this was what the average temperature was.
So a model that beats the baseline really only means something once you know which baseline you're comparing against. I think this is important to keep in mind as we go on to the interpretability and the explainability of our models.

I'm going to skip the slide. This is just saying the ensemble methods typically do much better than most individual models. And a lot of forecasting hubs or competitions, when teams submit forecasts, what they're doing is, well, the team may be using a very particular specific model,
and then the hub, the platform that's accepting these forecasts, then either takes the

average or the median and then develops that level of uncertainty around each of the forecasts that are coming in.
Where models struggle, you know, one quick note before we move on to explainability is that every model will struggle with turning points or critical events. In Brazil, I worked on a few projects for Dengue in Brazil, and
they had a record year in 2024 and doesn't matter what method you're using, the data is not there to predict an extreme event like this.
You also have things like early warning systems. So as a system approaches a tipping point, it tends to wobble more. Sort of like, you know, you can calculate that the mean and the variance and the autocorrelation, all of these things are changing. This comes from the critical slowing down theory. Like I said,
but these early warning systems often come with a high rate of false positives just simply because what you're trying to predict is an extreme event,
which is often very difficult to do.
Okay, so we have the ability to forecast, we have a few models, we take those models, we put them into an ensemble, we score the ensemble models using either weighted interval and coverage, and then we pass that information on to decision makers. For decision makers,
they need to be able to explain the forecast, right? Why did the model forecast what it did? And there's three reasons why we need to talk about this. One is accountability. If a decision maker is going to act on a forecast, they need to be able to explain it. The second is to catch issues in the training
process or the learning process of a model, which can lead to a good model, but for the wrong reasons. I'll give you an example. You know, in the fall or in the winter months, search engine trends may have a higher number of
searches for, for example, like indoor recreational sports. So if you're training a model using search engine data, it might have this correlation that indoor recreational sports have a, is positively correlated with
with flu, for example. And so that is a model that is predicting something correctly, but for the wrong reasons. And the third reason to have explainability and to study the explainability of a model is to learn about drivers and datasets that actually drive the model results. We look at inside the model to see which input it relies on the most and how much each one matters in the forecast.
And so by doing this, it can either teach you something about the disease you're

studying or the pathogen you're studying, or show you which data are worth collecting. There's two ways.



Belair, Marc-Andre (PHAC/ASPC) 1:19:12
Sorry to interrupt, but we're just coming up to time just for...

AS **Affan Shoukat** 1:19:15
I see. Okay. Yeah, I only have three more slides. So how do we get these explanations? There's two ways to get it. So there's some models that are interpretable by design. Statistical models, for example, regression models have coefficients, which have some sort of interpretation to them. The parameters of a mechanistic or hybrid model, like the reproduction number, the reporting rate,



AS **Affan Shoukat** 1:19:40
have self-explanatory interpretation to them. You can read them directly. Some models are black boxes, right? We explain them after the fact.

So the first thing I want to talk about is accuracy is not an explanation. In this, two models can be equally accurate, but for completely different reasons. And for example, in this figure, you can see that I have two models, the linear model in the dark green and the gradient boosting model, the tree model in the light green, and they're both
producing relatively the same forecast. But if you ask them what's causing dengue in a particular city, then the linear model says that, you know, it's temperature, the tree model says it's rainfall, and you might have another model that says it's humidity. And
why? It's because temperature, humidity, rainfall all rise in the summer. And a model can lean on any one of them to get the same accuracy.
So again, I'll give you an example of this.
Well, I'll hold off the example until the next slide.
The identifiability of variables is also very important. Just because the variable is

interpretable doesn't mean it's the same as being identifiable. So these are two very common terms that are used in forecasting. Interpretable means that
the parameter stands for something real. So we may have a parameter in our model, let's say it's the reporting rate, that parameter is interpretable. It means it's the share of infections that get reported, and we all know what that is. Identifiable means whether the data contains enough information to work out its value.
and only one value, or a narrow range that might fit the data well. Mechanistic models, your typical SIR-SEIR models, are interpretable because their parameters mean something.
On the other hand, neural networks, the weights and the biases of a neural network have no interpretation to them.

I'll give you an example that I wanted to give you last slide, but if you think about reported cases, if your data set contains reported cases, then mathematically reported cases are essentially the number of infections multiplied by the reporting rate.
And so if you only see reported cases, then it might come from lots of infections and a few reported cases or very little infections and most of them being reported. So if we, for example, look at this figure, you know, if I fit a model to these data points, using either 10% reporting rate, 25% reporting rate, or 60% reporting rate, I can certainly fit the model to data. That's not the issue. The issue is when I forecast using these three different values, I get completely different results. And so again, this is something to keep in mind
when forecasting.
And the last thing that Gemma also touched upon is that an explanation is not a causal effect. So you may have a model, you may have data that suggests, or you may have data for hospital admissions, for ICU admissions, for folks with asthma. And you know,
what happens is that a model that predicts death amongst pneumonia patients can learn that asthma lowers risk. Why? Because in a hospital, patients with asthma can come in with pneumonia, they are sent to intensive care, the intensive care work, they recover, and so the model learns that this particular
pathway that, you know, if you have asthma, then it lowers that rate, which is a perfectly valid output, but biologically inconsistent.
And so, you know, it is a perfectly interpretable model, I should say, but biologically

inconsistent. This is very true for mobility data that we saw in COVID-19, where, you know, we had mobility data explained the transmission very well across a lot of countries, but as controls were relaxed, transmission sort of decoupled from mobility. In almost every country, and...
The median correlation value fell significantly, but...
It was basically, so basically the point is that mobility does not cause or does not prevent COVID, not mobility by itself.
And so this is the last slide. I just want to have a few take-home messages as modelers, as you know, as the folks that are providing the forecast. We want to use multiple models, we want to use ensembles, we want to clearly state our baseline score and evaluate our forecast against it.
If there are other models that produce similar results, why? We want to be able to answer that question. And then even if we do provide the explanation, just remember the explanations describe the model, but not particularly the epidemic or the complex system or the policy that you might be studying.
For decision makers, there's five questions that you should ask, right? Was the output tested in real time on data that had never seen? Does it beat simple baselines? Does it work with turning points or critical events? Are there other accurate models that give the same results?
With the same explanation or a different explanation, and are all the parameters identified in the in the model? So, sorry, I went over time, but...



Belair, Marc-Andre (PHAC/ASPC) 1:25:55
Great. Excellent. Thank you so much. Yeah, I appreciate it. I'd like to invite our presenters to turn on their cameras. And we have a couple of few minutes for questions. So I will turn it over to you, Dan, to go ask a couple of questions. You've got about 5 or 6 minutes.

DL

Dan Lizotte 1:26:17
Thanks so much and thanks everybody for the fantastic talks. Really appreciate it. Let's jump right in. Lisa, I'm putting you on the spot. We had a couple of questions, you may have seen them in the chat, about synthetic data. So to set the stage, what I want to ask you is how you think about what purpose synthetic data serve. Do you

think of it as
for method development or for answering like substantive questions about the world, about like incidents and prevalence and stuff, or is it both? So let's start there and then let's talk a little bit more.

Lisa Lix 1:26:51
I saw the question about how real can you make synthetic data. And there is a trade-off here between how close it is to real world data or how private it is. And that can't be answered as a one-size-fits-all. It really applies to
what the application is and how comfortable the data custodians are with sharing synthetic data that may be close to being real. So there has to be a conversation and it can have many different applications, but there first has to be a conversation around
what tool is used and how real it will look.

Dan Lizotte 1:27:37
Totally. And do you think, are you thinking about using synthetic data to answer actual like epidemiologic questions? Or do you think of it more as like, we're considering using a different kind of method. We want some assurances that it actually works. We're going to test it in synthetic data. Do you see what I mean?

Lisa Lix 1:27:54
I use it primarily in methods development, but for complex studies that involve multiple sites, synthetic data are an alternative to answer substantive questions, where data from individual sites cannot all be shared into a common repository.

Dan Lizotte 1:28:13
That's great. And so it can be sort of, you can think about it as an alternative or complementary idea to federated learning if people have heard of that. Do you agree? Do people talk about synthetic data and federated learning as sort of, maybe we might do one, maybe we might do the other?

Lisa Lix 1:28:20
Yes.
Correct.
Correct, they can, they really can be complementary.

Dan Lizotte 1:28:28
Yeah.
That's great. Thank you. I'd like to ask, let me just see here, sorry. Maybe we'll, I'd like to ask Gemma a question about shop values and those sort of explainability components to two models. Can you say a little bit about
what you feel, so if you know, if I do a logistic regression or whatever, and I'm trying to sort of explain what's going on with the model, then I refer to the coefficients, most likely, in some capacity, and Affan kind of talked about this a little bit.
And that serves a particular kind of purpose to understand how changes to the inputs change outputs in the model. In your view, what's the, is there a relationship between doing that sort of old school, you know, classical statistical approach compared with what you get out of SHAP values? And maybe part of what I want to know is,
is there a danger in making that comparison because they're not really giving you the same thing?

Gemma Postill 1:29:34
Yeah, and I'll also open it up to Affan after, because I'd be curious to know what you think. I think for me, one of the key differences between the SHAP values and sort of your traditional coefficients is that we, before we create the logistic regression model, we make sure there's no collinearity. So each variable
in theory represents a distinct construct. That's not a requirement for these machine learning models, which is a blessing and a curse, because then when we go to look at the SHAP values, I think often we have our, you know, epidemiological hats on and we think, okay, well, this is a distinct construct and this represents this, but
we don't really have that same inference because often they're very correlated. And so, you know, Affan mentioned like the person with asthma might be more likely to

be admitted to the ICU and that's why that variable is important. So I look at it to sound check that there's no crazy variables the model is learning from like "Do I have a pet at home?" and that's predicting my risk of mortality, but I don't look at it in the same way that I do with like a logistic regression model.

Dan Lizotte 1:30:38
Right.
OK, thanks. I think Marc is online, so we may we need to wrap it up, right?



Belair, Marc-Andre (PHAC/ASPC) 1:30:45
Yes. I'm so sorry to be the bearer of bad news, but we are at time. I wanted to take a quick moment to say a huge thank you to all of you, our presenters, for taking the time to be with us today. I will be sharing on the screen momentarily a link to the Slido where you can create our evaluation.
to fill out our evaluation. Very important for us, we use it to help improve future sessions. So that should be coming up on your screen right now. So you can go to slido.com slash and use the code AI4PH2. And we're also excited that our next session will be on Friday, November 13th. So please watch your emails and the AI for PHAC website for information on how to register for the upcoming session. Again, thank you.
Dan and Lisa and Affan and Gemma for your time today. Really appreciate it and also like to thank the folks at AI4PH, the team here at the Training and Development Unit and AV Services at PHAC. So on behalf of Public Health Agency of Canada and AI4PHI thank you for choosing to spend some time with us today and I hope to see you at future sessions and I wish you all a great rest of your day. Thank you.

Lisa Lix 1:31:54
Great. Thank you, Marc Andre. Thank you to the other speakers, and thank you, Dan, for moderating.

Dan Lizotte 1:31:56
Bye.

◉ stopped transcription